United States Patent Application Publication 20250249084
Kind Code A1
Publication Date August 07, 2025
Inventor(s) Blair; Wade et al.

ALPHAVIRUS ANTIGEN VECTORS

Abstract

Disclosed herein are alphavirus vectors that include neoantigen-encoding nucleic acid sequences derived from a tumor of a subject. Also disclosed are nucleotides, cells, and methods associated with the vectors including their use as vaccines.

Inventors: **Blair; Wade (Gaithersburg, MD), Jooss; Karin (San Diego, CA), Rappaport; Amy Rachel (Daly City, CA), Scallan; Ciaran Daniel (San Francisco, CA), Gitlin; Leonid (Foster City, CA)**

Applicant: **Gritstone Bio, Inc.** (Emeryville, CA)

Family ID: **64105513**

Appl. No.: **18/830813**

Filed: **September 11, 2024**

Related U.S. Application Data

parent US division 18045812 20221011 parent-grant-document US 12109257 child US 18830813
parent US division 16612352 20191108 parent-grant-document US 11504421 US division PCT/US2018/031696 20180508 child US 18045812
us-provisional-application US 62590163 20171122
us-provisional-application US 62523201 20170621
us-provisional-application US 62503283 20170508

Publication Classification

Int. Cl.: **A61K39/00** (20060101); **A61K39/12** (20060101); **A61P31/12** (20060101); **C07K14/47** (20060101); **C07K14/74** (20060101); **C12N15/86** (20060101)

U.S. Cl.:

CPC **A61K39/001191** (20180801); **A61K39/001188** (20180801); **A61K39/12** (20130101); **A61P31/12** (20180101); **C07K14/4748** (20130101); **C07K14/70539** (20130101); **C12N15/86** (20130101); A61K2039/53 (20130101); A61K2039/545 (20130101); A61K2039/585 (20130101); A61K2039/6037 (20130101); A61K2039/605 (20130101); A61K2039/70 (20130101)

Background/Summary

PRIORITY CLAIMS AND RELATED APPLICATIONS [0001] This application is a divisional of U.S. application Ser. No. 18/045,812, filed Oct. 11, 2022, which is a divisional of U.S. application Ser. No. 16/612,352, filed Nov. 8, 2019, now issued as U.S. Pat. No. 11,504,421, which is the National Stage of International Application No. PCT/US2018/031696, filed Aug. 5, 2018, which claims the benefit of U.S. Provisional Application No. 62/590,163, filed Nov. 22, 2017, U.S. Provisional Application No. 62/523,201, filed Jun. 21, 2017, and U.S. Provisional Application No. 62/503,283, filed May 8, 2017, the entire contents of each is incorporated herein by reference for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted via Patent Center and is hereby incorporated herein by reference in its entirety. The accompanying sequence listing .XML file name GSO-006D2, was created on 7 Nov. 2022, and is 538 kb in size.

BACKGROUND

[0003] Therapeutic vaccines based on tumor-specific neoantigens hold great promise as a next-generation of personalized cancer immunotherapy..sup.1-3 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..sup.4-5 Early evidence shows that neoantigen-based vaccination can elicit T-cell responses.sup.6 and that neoantigen targeted cell-therapy can cause tumor regression under certain circumstances in selected patients..sup.7

[0004] One question for neoantigen vaccine design is which of the many coding mutations present in subject tumors can generate the “best” therapeutic neoantigens, e.g., antigens that can elicit anti-tumor immunity and cause tumor regression.

[0005] Initial methods have been proposed incorporating mutation-based analysis using next-generation sequencing, RNA gene expression, and prediction of MHC binding affinity of candidate neoantigen peptides.sup.8. However, these proposed methods can fail to model the entirety of the epitope generation process, which contains many steps (e.g., TAP transport, proteasomal cleavage, and/or TCR recognition) in addition to gene expression and MHC binding.sup.9. Consequently, existing methods are likely to suffer from reduced low positive predictive value (PPV). (FIG. **1**A)

[0006] Indeed, analyses of peptides presented by tumor cells performed by multiple groups have shown that <5% of peptides that are predicted to be presented using gene expression and MHC binding affinity can be found on the tumor surface MHC.sup.10,11 (FIG. **1**B). This low correlation between binding prediction and MHC presentation was further reinforced by recent observations of the lack of predictive accuracy improvement of binding-restricted neoantigens for checkpoint inhibitor response over the number of mutations alone..sup.12

[0007] This low positive predictive value (PPV) of existing methods for predicting presentation presents a problem for neoantigen-based vaccine design. If vaccines are designed using predictions with a low PPV, most patients are unlikely to receive a therapeutic neoantigen and fewer still are likely to receive more than one (even assuming all presented peptides are immunogenic). Thus, neoantigen vaccination with current methods is unlikely to succeed in a substantial number of subjects having tumors. (FIG. **1**C)

[0008] Additionally, previous approaches generated candidate neoantigens using only cis-acting mutations, and largely neglected to consider additional sources of neo-ORFs, including mutations in splicing factors, which occur in multiple tumor types and lead to aberrant splicing of many genes.sup.13, and mutations that create or remove protease cleavage sites.

[0009] Finally, standard approaches to tumor genome and transcriptome analysis can miss somatic mutations that give rise to candidate neoantigens due to suboptimal conditions in library construction, exome and transcriptome capture, sequencing, or data analysis. Likewise, standard tumor analysis approaches can inadvertently promote sequence artifacts or germline polymorphisms as neoantigens, leading to inefficient use of vaccine capacity or auto-immunity risk, respectively.

[0010] In addition to the challenges of current neoantigen prediction methods certain challenges also exist with the available vector systems that can be used for neoantigen 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 neoantigen delivery for cancer treatment.

SUMMARY

[0011] Disclosed herein is a composition for delivery of a neoantigen expression system, comprising: the neoantigen expression system, wherein the neoantigen expression system comprises one or more vectors, the one or more vectors comprising: (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises: (i) at least one promoter nucleotide sequence, and (ii) at least one polyadenylation (poly(A)) sequence; and (b) a

neoantigen cassette, wherein the neoantigen cassette comprises: (i) at least one neoantigen-encoding nucleic acid sequence derived from a tumor present within a subject, comprising: (I) at least one tumor-specific and subject-specific MHC class I neoantigen-encoding nucleic acid sequence derived from the tumor, and comprising: (A) a MHC class I epitope encoding nucleic acid sequence with at least one alteration that makes the encoded peptide sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence, and (B) optionally, a 5′ linker sequence, and (C) optionally, a 3′ linker sequence; (ii) optionally, a second promoter nucleotide sequence operably linked to the neoantigen-encoding nucleic acid sequence; and (iii) optionally, at least one MHC class II antigen-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 alphavirus.

[0012] Also disclosed herein is a composition for delivery of a neoantigen expression system, comprising: the neoantigen expression system, wherein the neoantigen expression system comprises one or more vectors, the one or more vectors comprising: (a) an RNA alphavirus backbone, wherein the RNA alphavirus backbone comprises the nucleic acid sequence set forth in SEQ ID NO:6, wherein the RNA alphavirus backbone sequence comprises a 26S promoter nucleotide sequence and a poly(A) sequence, wherein the 26S promoter sequence is endogenous to the RNA alphavirus backbone, and wherein the poly(A) sequence is endogenous to the RNA alphavirus backbone; and (b) a neoantigen cassette integrated between the 26S promoter nucleotide sequence and the poly(A) sequence, wherein the neoantigen cassette comprises: (i) at least one neoantigen-encoding nucleic acid sequence derived from a tumor present within a subject, comprising: (I) at least 10 tumor-specific and subject-specific MHC class I neoantigen-encoding nucleic acid sequences linearly linked to each other and each comprising: (A) a MHC class I epitope encoding nucleic acid sequence with at least one alteration that makes the encoded peptide sequence distinct from the corresponding peptide sequence encoded by a wild-type nucleic acid sequence, wherein the MHC 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 I epitope, and wherein the 5′ linker sequence encodes a peptide that is at least 3 amino acids in length, (C) a 3′ linker sequence, wherein the 3′ linker sequence encodes a native N-terminal acid sequence of the MHC I epitope, and wherein the 3′ linker sequence encodes a peptide that is at least 3 amino acids in length, and wherein the neoantigen cassette is operably linked to the 26S promoter nucleotide sequence, wherein each of the MHC class I neoantigen-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length, and wherein each 3′ end of each MHC class I neoantigen-encoding nucleic acid sequence is linked to the 5′ end of the following MHC class I neoantigen-encoding nucleic acid sequence with the exception of the final MHC class I neoantigen-encoding nucleic acid sequence in the neoantigen cassette; and (ii) at least two MHC class II antigen-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 (SEQ ID NO: 56) 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 (SEQ ID NO: 56) linking the 5′ end of the at least two MHC class II antigen-encoding nucleic acid sequences to the at least 20 tumor-specific and subject-specific MHC class I neoantigen-encoding nucleic acid sequences, (V) optionally, a third nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO: 56) at the 3′ end of the at least two MHC class II antigen-encoding nucleic acid sequences.

[0013] In some aspects, an ordered sequence of each element of the neoantigen cassette is described in the formula, from 5′ to 3′, comprising:

Pa-(L5b-Nc-L3d)X-(G5e-Uf)Y-G3g

[0014] wherein P comprises the second promoter nucleotide sequence, where a=0 or 1, N comprises one of the MHC class I 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 (SEQ ID NO: 56), where e=0 or 1, G3 comprises one of the at least one nucleic acid sequences encoding a GPGPG amino acid linker (SEQ ID NO: 56), where g=0 or 1, U comprises one of the at least one MHC class II antigen-encoding nucleic acid sequence, where f=1, X=1 to 400, where for each X the corresponding Nc is a epitope encoding nucleic acid sequence, and Y=0, 1, or 2, where for each Y the corresponding Uf is an antigen-encoding nucleic acid sequence. In some aspects, for each X the corresponding Nc is a distinct MHC class I epitope encoding nucleic acid sequence. In some aspects, for each Y the corresponding Uf is a distinct MHC class II antigen-encoding nucleic acid sequence.

[0015] In some aspects, a=0, b=1, d=1, e=1, g=1, h=1, X=20, Y=2, the at least one promoter nucleotide sequence

is a single 26S promoter nucleotide sequence provided by the RNA alphavirus backbone, the at least one polyadenylation poly(A) sequence is a poly(A) sequence of at least 100 consecutive A nucleotides provided by the RNA alphavirus backbone, each N encodes a MHC class I epitope 7-15 amino acids in length, L5 is a native 5′ linker sequence that encodes a native N-terminal amino acid sequence of the MHC I epitope, and wherein the 5′ linker sequence encodes a peptide that is at least 3 amino acids in length, L3 is a native 3′ linker sequence that encodes a native nucleic-terminal acid sequence of the MHC I epitope, and wherein the 3′ linker sequence encodes a peptide that is at least 3 amino acids in length, U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the RNA alphavirus backbone is the sequence set forth in SEQ ID NO:6, and each of the MHC class I neoantigen-encoding nucleic acid sequences encodes a polypeptide that is between 13 and 25 amino acids in length.

[0016] 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 neoantigen expression system.

[0017] In some aspects, any of the above compositions further comprise a plurality of LNPs, wherein the LNPs comprise: the neoantigen 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.

[0018] In some aspects, the non-cationic lipid is a mixture of (1) a phospholipid and (2) cholesterol or a cholesterol derivative.

[0019] 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 dialkyloxypropyl (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 (C.sub.10) conjugate, a PEG-dilauryloxypropyl (C.sub.12) conjugate, a PEG-dimyristyloxypropyl (C.sub.14) conjugate, a PEG-dipalmityloxypropyl (C.sub.16) conjugate, a PEG-distearyloxypropyl (Cis) conjugate, and a mixture thereof.

[0020] In some aspects, the neoantigen expression system is fully encapsulated in the LNPs.

[0021] In some aspects, the non-lamellar morphology of the LNPs comprises an inverse hexagonal (H.sub.II) or cubic phase structure.

[0022] 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.

[0023] 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.

[0024] 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.

[0025] 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.

[0026] 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.

[0027] 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.

[0028] In some aspects, the phospholipid comprises dipalmitoylphosphatidylcholine (DPPC), distearoylphosphatidylcholine (DSPC), or a mixture thereof.

[0029] 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-dialkyloxypropyl (PEG-DAA) conjugate, or a mixture thereof. In some aspects, the PEG-DAA conjugate comprises a PEG-dimyristyloxypropyl (PEG-DMA) conjugate, a PEG-distearyloxypropyl (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.

[0030] In some aspects, the conjugated lipid comprises from 1 mol % to 2 mol % of the total lipid present in the LNPs.

[0031] In some aspects, the LNP comprises a compound having a structure of Formula I:

##STR00001##

or a pharmaceutically acceptable salt, tautomer, prodrug or stereoisomer thereof, wherein: L.sup.1 and L.sup.2 are each independently -0(C=0)-, —(C=0)0-, —C(═0)-, —0-, —S(0).sub.x-, —S—S—, —C(═0)S—, —SC(═0)-, —R.sup.aC(=0)-, —C(═0) R.sup.a—, —R.sup.aC(=0) R.sup.a—, —OC(═0) R.sup.a—, —R.sup.aC(═0)0- or a direct bond; G.sup.1 is Ci-C.sub.2 alkylene, —(C=0)-, -0(C=0)-, —SC(═0)-, —R.sup.aC(=0)- or a direct bond: —C(═O)—, —(C=0)0-, —C(═0)S—, —C(═0) R.sup.a— or a direct bond; G is Ci-C.sub.6 alkylene; R.sup.a is H or C1-C12 alkyl; R.sup.1a and R.sup.1b are, at each occurrence, independently either: (a) H or C.sub.1-C.sub.12 alkyl; or (b) R.sup.1a is H or C.sub.1-C.sub.12 alkyl, and R.sup.1b together with the carbon atom to which it is bound is taken together with an adjacent R.sup.1b and the carbon atom to which it is bound to form a carbon-carbon double bond; R.sup.2a and R.sup.2b are, at each occurrence, independently either: (a) H or C.sub.1-C.sub.12 alkyl; or (b) R.sup.2a is H or C.sub.1-C.sub.12 alkyl, and R.sup.2b together with the carbon atom to which it is bound is taken together with an adjacent R.sup.2b and the carbon atom to which it is bound to form a carbon-carbon double bond; R.sup.3a and R.sup.3b are, at each occurrence, independently either (a): H or C.sub.1-C.sub.12 alkyl; or (b) R.sup.3a is H or C.sub.1-C.sub.12 alkyl, and R.sup.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.sup.4a and R.sup.4b are, at each occurrence, independently either: (a) H or C1-C12 alkyl; or (b) R.sup.4a is H or C1-C12 alkyl, and R.sup.4b together with the carbon atom to which it is bound is taken together with an adjacent R.sup.4b and the carbon atom to which it is bound to form a carbon-carbon double bond; R.sup.5 and R.sup.6 are each independently H or methyl; R.sup.7 is C4-C20 alkyl; R.sup.8 and R.sup.9 are each independently C1-C12 alkyl; or R.sup.8 and R.sup.9, 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.

[0032] In some aspects, the LNP comprises a compound having a structure of Formula II:

##STR00002##

or a pharmaceutically acceptable salt, tautomer, prodrug or stereoisomer thereof, wherein: L.sup.1 and L.sup.2 are each independently -0(C=0)-, —(C=0)0- or a carbon-carbon double bond; R.sup.1a and R.sup.1b are, at each occurrence, independently either (a) H or C.sub.1-C.sub.12 alkyl, or (b) R.sup.1a is H or C.sub.1-C.sub.12 alkyl, and R.sup.1b together with the carbon atom to which it is bound is taken together with an adjacent R.sup.1b and the carbon atom to which it is bound to form a carbon-carbon double bond; R.sup.2a and R.sup.21 are, at each occurrence, independently either (a) H or C.sub.1-C.sub.12 alkyl, or (b) R.sup.2a is H or C.sub.1-C.sub.12 alkyl, and R.sup.21 together with the carbon atom to which it is bound is taken together with an adjacent R.sup.2b and the carbon atom to which it is bound to form a carbon-carbon double bond; R.sup.3a and R.sup.3b are, at each occurrence, independently either (a) H or C.sub.1-C.sub.12 alkyl, or (b) R.sup.3a is H or C.sub.1-C.sub.12 alkyl, and R.sup.3b together with the carbon atom to which it is bound is taken together with an adjacent R.sup.3b and the carbon atom to which it is bound to form a carbon-carbon double bond; R.sup.4a and R.sup.4b are, at each occurrence, independently either (a) H or C.sub.1-C.sub.12 alkyl, or (b) R.sup.4a is H or C.sub.1-C.sub.12 alkyl, and R.sup.4b together with the carbon atom to which it is bound is taken together with an adjacent R.sup.4b and the carbon atom to which it is bound to form a carbon-carbon double bond; R.sup.5 and R.sup.6 are each independently methyl or cycloalkyl; R.sup.7 is, at each occurrence, independently H or C.sub.1-C.sub.12 alkyl; R.sup.8 and R.sup.9 are each independently unsubstituted C.sub.1-C.sub.12 alkyl; or R.sup.8 and R.sup.9, 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.sup.1a, R.sup.2a, R.sup.3a or R.sup.4a is C1-C12 alkyl, or at least one of L.sup.1 or L.sup.2 is -0(C=0)- or —(C=0)0-; and R.sup.1a and R.sup.1b are not isopropyl when a is 6 or n-butyl when a is 8.

[0033] 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.

[0034] In some aspects, the molar ratio of the compound to the neutral lipid ranges from about 2:1 to about 8:1.

[0035] 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.

[0036] 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 dialkyoxypropylcarbamate. In some aspects, the pegylated lipid has the following structure III:

##STR00003## [0037] or a pharmaceutically acceptable salt, tautomer or stereoisomer thereof, wherein: R.sup.10 and R.sup.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.sup.10 and R.sup.11 are each independently straight, saturated alkyl chains having 12 to 16 carbon atoms. In some aspects, the average z is about 45. [0038] start here

[0039] 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 60 nm and 120 nm. In some aspects, the non-bilayer structures have a diameter of about 70 nm, about 80 nm, about 90 nm, or about 100 nm. In some aspects, wherein the nanoparticulate delivery vehicle has a diameter of about 100 nm.

[0040] In some aspects, the neoantigen cassette is integrated between the at least one promoter nucleotide sequence and the at least one poly(A) sequence. In some aspects, the at least one promoter nucleotide sequence is operably linked to the neoantigen-encoding nucleic acid sequence.

[0041] 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.

[0042] In some aspects, the RNA alphavirus backbone comprises at least one nucleotide sequence of an Aura virus, a Fort Morgan virus, a Venezuelan equine encephalitis virus, a Ross River virus, a Semliki Forest virus, a Sindbis virus, or a Mayaro virus. In some aspects, the RNA alphavirus backbone comprises at least one nucleotide sequence of a Venezuelan equine encephalitis virus. In some aspects, the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, a poly(A) sequence, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, and a nsP4 gene encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, the RNA alphavirus backbone comprises at least sequences for nonstructural protein-mediated amplification, a 26S promoter sequence, and a poly(A) sequence encoded by the nucleotide sequence of the Aura virus, the Fort Morgan virus, the Venezuelan equine encephalitis virus, the Ross River virus, the Semliki Forest virus, the Sindbis virus, or the Mayaro virus. In some aspects, sequences for nonstructural protein-mediated amplification are selected from the group consisting of: an alphavirus 5′ UTR, a 51-nt CSE, a 24-nt CSE, a 26S subgenomic promoter sequence, a 19-nt CSE, an alphavirus 3′ UTR, or combinations thereof.

[0043] In some aspects, the RNA alphavirus backbone does not encode structural virion proteins capsid, E2 and E1. In some aspects, the neoantigen cassette is inserted in place of the 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.

[0044] In some aspects, the Venezuelan equine encephalitis virus (VEE) comprises the strain TC-83. In some aspects, the Venezuelan equine encephalitis virus comprises the sequence set forth in SEQ ID NO:3 or SEQ ID NO:5 In some aspects, the Venezuelan equine encephalitis virus comprises the sequence of SEQ ID NO:3 or SEQ ID NO:5 further comprising a deletion between base pair 7544 and 11175. In some aspects, the RNA alphavirus backbone is the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7. In some aspects, the neoantigen cassette is inserted to replace the deletion between base pair 7544 and 11175 set forth in the sequence of SEQ ID NO:3 or SEQ ID NO:5.

[0045] In some aspects, the insertion of the neoantigen cassette provides for transcription of a polycistronic RNA comprising the nsP1-4 genes and the at least one of antigen-encoding nucleic acid sequences, wherein the nsP1-4 genes and the at least one of antigen-encoding nucleic acid sequences are in separate open reading frames.
[0046] In some aspects, the at least one promoter nucleotide sequence is the native 26S promoter nucleotide sequence encoded by the RNA alphavirus backbone. In some aspects, the at least one promoter nucleotide sequence is an exogenous RNA promoter. In some aspects, the second promoter nucleotide sequence is a 26S promoter nucleotide sequence. In some aspects, the second promoter nucleotide sequence comprises multiple 26S promoter nucleotide sequences, wherein each 26S promoter nucleotide sequence provides for transcription of one or more of the separate open reading frames.
[0047] In some aspects, the one or more neoantigen expression vectors are each at least 300 nt in size. In some aspects, the one or more neoantigen expression vectors are each at least 1 kb in size. In some aspects, the one or more neoantigen expression vectors are each 2 kb in size. In some aspects, the one or more neoantigen expression vectors are each less than 5 kb in size.
[0048] In some aspects, at least one of the at least one neoantigen-encoding nucleic acid sequences encodes a polypeptide sequence or portion thereof that is presented by MHC class I on the tumor cell. In some aspects, each antigen-encoding nucleic acid sequence is linked directly to one another. In some aspects, at least one of the at least one antigen-encoding nucleic acid sequences is linked to a distinct antigen-encoding nucleic acid sequence with a nucleic acid sequence encoding a linker. In some aspects, the linker links two MHC class I 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 (SEQ ID NO: 56).
[0049] In some aspects, at least one sequence of the at least one antigen-encoding nucleic acid sequences is linked, operably or directly, to a separate or contiguous sequence that enhances the expression, stability, cell trafficking, processing and presentation, and/or immunogenicity of the at least one antigen-encoding nucleic acid sequences. 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.
[0050] In some aspects, at least one of the at least one neoantigen-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 at least one neoantigen-encoding nucleic acid sequences in the plurality 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 at least one neoantigen-encoding nucleic acid sequences in the plurality 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.
[0051] In some aspects, at least one mutation 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.
[0052] 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.
[0053] In some aspects, the at least one neoantigen-encoding nucleic acid sequence comprises at least 2-10, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleic acid sequences. In some aspects, the at least one neoantigen-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 nucleic acid sequences.

[0054] In some aspects, the at least one neoantigen-encoding nucleic acid sequence comprises at least 2-400 nucleic acid sequences and wherein at least two of the neoantigen-encoding nucleic acid sequences encode polypeptide sequences or portions thereof that are presented by MHC class I on the tumor cell surface. In some aspects, at least two of the neoantigen-encoding nucleic acid sequences encode polypeptide sequences or portions thereof that are presented by MHC class I on the tumor cell surface. In some aspects, when administered to the subject and translated, at least one of the neoantigens encoded by the at least one neoantigen-encoding nucleic acid sequence are presented on antigen presenting cells resulting in an immune response targeting at least one of the neoantigens on the tumor cell surface. In some aspects, the at least one neoantigen-encoding nucleic acid sequences when administered to the subject and translated, at least one of the MHC class I or class II neoantigens are presented on antigen presenting cells resulting in an immune response targeting at least one of the neoantigens on the tumor cell surface, and optionally wherein the expression of each of the at least one neoantigen-encoding nucleic acid sequences is driven by the at least one promoter nucleotide sequence.
[0055] In some aspects, each MHC class I neoantigen-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.
[0056] In some aspects, at least one MHC class II antigen-encoding nucleic acid sequence is present. In some aspects, at least one MHC class II antigen-encoding nucleic acid sequence is present and comprises at least one MHC class II neoantigen-encoding nucleic acid sequence that comprises at least one mutation that makes it distinct from the corresponding wild-type, parental nucleic acid sequence. In some aspects, the at least one MHC class II antigen-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 antigen-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.
[0057] 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.
[0058] In some aspects, the at least one poly(A) sequence comprises a poly(A) sequence native to the alphavirus. In some aspects, the at least one poly(A) sequence comprises a poly(A) sequence exogenous to the alphavirus. In some aspects, the at least one poly(A) sequence is operably linked to at least one of the at least one 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, or at least 90 consecutive A nucleotides. In some aspects, the at least one poly(A) sequence is at least 100 consecutive A nucleotides.
[0059] In some aspects, the neoantigen 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.
[0060] In some aspects, the neoantigen 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.
[0061] In some aspects, the one or more vectors further comprise 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.
[0062] 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.
[0063] Also, disclosed herein is an adenovirus vector comprising a neoantigen cassette, the neoantigen cassette comprising: a plurality of antigen-encoding nucleic acid sequences derived from a tumor present within a subject,

the plurality comprising: at least two MHC class I neoantigen-encoding nucleic acid sequences each comprising at least one alteration that makes it distinct from the corresponding wild-type, parental nucleic acid sequence, and optionally, at least one MHC class II antigen-encoding nucleic acid sequence; and at least one promoter sequence operably linked to at least one sequence of the plurality.

[0064] In some aspects, the adenovirus vector is a chimpanzee adenovirus (ChAd) vector, optionally a C68 vector. In some aspects, the adenovirus vector comprises the sequence set forth in SEQ ID NO:1. In some aspects, the adenovirus vector comprises the sequence set forth in SEQ ID NO:1, except that the sequence is fully deleted or functionally deleted in at least one gene selected from the group consisting of the chimpanzee adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO: 1, optionally wherein the sequence is fully deleted or functionally deleted in: (1) E1A and E1B; (2) E1A, E1B, and E3; or (3) E1A, E1B, E3, and E4 of the sequence set forth in SEQ ID NO: 1. In some aspects, the adenovirus vector comprises a gene or regulatory sequence obtained from the sequence of SEQ ID NO: 1, optionally wherein the gene is selected from the group consisting of the chimpanzee adenovirus inverted terminal repeat (ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO: 1.

[0065] In some aspects, the neoantigen cassette is inserted in the adenovirus vector at the E1 region, E3 region, and/or any deleted AdV region that allows incorporation of the neoantigen cassette.

[0066] In some aspects, the at least one promoter sequence of the adenovirus vector is inducible. In some aspects, the at least one promoter sequence of the adenovirus vector is non-inducible. In some aspects, the at least one promoter sequence of the adenovirus vector is a CMV, SV40, EF-1, RSV, PGK, or EBV promoter sequence.

[0067] In some aspects, the neoantigen cassette of the adenovirus vector further comprises at least one polyA sequence operably linked to at least one of the sequences in the plurality, optionally wherein the polyA sequence is located 3′ of the at least one sequence in the plurality.

[0068] In some aspects, the adenovirus vector is generated from one of a first generation, a second generation, or a helper-dependent adenoviral vector.

[0069] In some aspects, the adenovirus vector comprises one or more deletions between base pair number 577 and 3407 and optionally wherein the adenovirus vector further comprises one or more deletions between base pair 27,141 and 32,022 or between base pair 27,816 and 31,332 of the sequence set forth in SEQ ID NO:1. In some aspects, the adenovirus vector further comprises one or more deletions between base pair number 3957 and 10346, base pair number 21787 and 23370, and base pair number 33486 and 36193 of the sequence set forth in SEQ ID NO:1.

[0070] In some aspects, the at least one MHC class I neoantigen-encoding nucleic acid sequence is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome tumor nucleotide sequencing data from the tumor, wherein the tumor nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of neoantigens; (b) inputting the peptide sequence of each neoantigen into a presentation model to generate a set of numerical likelihoods that each of the neoantigens is presented by one or more of the MHC alleles on the tumor cell surface of the tumor, 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 neoantigens based on the set of numerical likelihoods to generate a set of selected neoantigens which are used to generate the at least one MHC class I neoantigen-encoding nucleic acid sequence.

[0071] In some aspects, each of the at least one MHC class I neoantigen-encoding nucleic acid sequence is selected by performing the steps of: (a) obtaining at least one of exome, transcriptome, or whole genome tumor nucleotide sequencing data from the tumor, wherein the tumor nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of neoantigens; (b) inputting the peptide sequence of each neoantigen into a presentation model to generate a set of numerical likelihoods that each of the neoantigens is presented by one or more of the MHC alleles on the tumor cell surface of the tumor, 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 neoantigens based on the set of numerical likelihoods to generate a set of selected neoantigens which are used to generate the at least one MHC class I neoantigen-encoding nucleic acid sequence.

[0072] In some aspects, a number of the set of selected neoantigens is 2-20.

[0073] In some aspects, the presentation model represents dependence between: presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and likelihood of presentation on the tumor 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.

[0074] In some aspects, selecting the set of selected neoantigens comprises selecting neoantigens that have an increased likelihood of being presented on the tumor cell surface relative to unselected neoantigens based on the presentation model. In some aspects, selecting the set of selected neoantigens comprises selecting neoantigens that have an increased likelihood of being capable of inducing a tumor-specific immune response in the subject

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

[0075] In some aspects, the neoantigen cassette comprises junctional epitope sequences formed by adjacent sequences in the neoantigen cassette. In some aspects, at least one or each junctional epitope sequence has an affinity of greater than 500 nM for MHC. In some aspects, each junctional epitope sequence is non-self. In some aspects, the neoantigen 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 neoantigen cassette. In some aspects, the prediction is based on presentation likelihoods generated by inputting sequences of the non-therapeutic epitopes into a presentation model. In some aspects, an order of the at least one antigen-encoding nucleic acid sequences in the neoantigen cassette is determined by a series of steps comprising: (a) generating a set of candidate neoantigen cassette sequences corresponding to different orders of the at least one antigen-encoding nucleic acid sequences; (b) determining, for each candidate neoantigen cassette sequence, a presentation score based on presentation of non-therapeutic epitopes in the candidate neoantigen cassette sequence; and (c) selecting a candidate cassette sequence associated with a presentation score below a predetermined threshold as the neoantigen cassette sequence for a neoantigen vaccine.

[0076] 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.

[0077] Also disclosed herein is an isolated nucleotide sequence or set of isolated nucleotide sequences comprising the neoantigen 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. In some aspects, the sequence or set of isolated nucleotide sequences comprises a neoantigen cassette disclosed herein inserted at position 7544 of the sequence set forth in SEQ ID NO:6 or SEQ ID NO:7. In some aspects, the isolated nucleotide sequence further comprises 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 optionally one or more restriction sites 3′ of the poly(A) sequence. In some aspects, the the neoantigen cassette disclosed herein is inserted at position 7563 of SEQ ID NO:8 or SEQ ID NO:9. In another aspect, the sequences set forth in SEQ ID NO:8 or SEQ ID NO:9 further comprise an additional adenine nucleotide inserted at position 17.

[0078] Also disclosed herein is an isolated nucleotide sequence comprising a neoantigen cassette disclosed herein and at least one promoter disclosed herein. In some aspects, the isolated nucleotide sequence further comprises a ChAd-based gene. In some aspects, the ChAd-based gene is obtained from the sequence of SEQ ID NO: 1, optionally wherein the gene is selected from the group consisting of the chimpanzee adenovirus ITR, E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4, and L5 genes of the sequence set forth in SEQ ID NO: 1, and optionally wherein the nucleotide sequence is cDNA.

[0079] Also disclosed herein is an isolated cell comprising an isolated nucleotide sequence disclosed 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.

[0080] Also disclosed herein is a vector comprising an isolated nucleotide sequence disclosed herein.

[0081] Also disclosed herein is a kit comprising a vector or a composition disclosed herein and instructions for

use.
[0082] Also disclosed herein is a method for treating a subject with cancer, the method comprising administering to the subject a vector disclosed herein or a pharmaceutical composition disclosed herein. In some aspects, the at least one MHC class I neoantigen-encoding nucleic acid sequence derived from a tumor are derived from the tumor of the subject with cancer. In some aspects, the at least one MHC class I neoantigen-encoding nucleic acid sequence are not derived from the tumor of the subject with cancer.
[0083] 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.
[0084] In some aspects, the vector or composition is administered intramuscularly (IM), intradermally (ID), or subcutaneously (SC), or intravenously (IV).
[0085] In some aspects, the methods described herein further comprise 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.
[0086] In some aspects, the methods described herein further comprise administering to the subject a second vaccine composition. In some aspects, the second vaccine composition is administered prior to the administration of the composition or the pharmaceutical composition described above. In some aspects, the second vaccine composition is administered subsequent to the administration of the composition or the pharmaceutical compositions described above. In some aspects, the second vaccine composition is the same as the composition or the pharmaceutical compositions described above. In some aspects, the second vaccine composition is different from the composition or the pharmaceutical compositions described above. 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 compositions or vectors.
[0087] 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 RNA alphavirus backbone and the neoantigen 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.
[0088] 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 neoantigen expression system; and providing conditions sufficient for the nanoparticulate delivery vehicle and the neoantigen expression system to produce the composition for delivery of the neoantigen expression system. In some aspects, the conditions are provided by microfluidic mixing.
[0089] 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 neoantigen cassette; transfecting the plasmid sequence into one or more host cells; and isolating the adenovirus vector from the one or more host cells.
[0090] 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.
[0091] 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.

Description

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0092] 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:

[0093] FIG. **1**A shows current clinical approaches to neoantigen identification.

[0094] FIG. **1**B shows that <5% of predicted bound peptides are presented on tumor cells.

[0095] FIG. **1**C shows the impact of the neoantigen prediction specificity problem.

[0096] FIG. **1**D shows that binding prediction is not sufficient for neoantigen identification.

[0097] FIG. **1**E shows probability of MHC-I presentation as a function of peptide length.

[0098] FIG. **1**F shows an example peptide spectrum generated from Promega's dynamic range standard. Figure discloses SEQ ID NO: 59.

[0099] FIG. **1**G shows how the addition of features increases the model positive predictive value.

[0100] FIG. **2**A is an overview of an environment for identifying likelihoods of peptide presentation in patients, in accordance with an embodiment.

[0101] FIG. **2**B, FIG. **2**C, FIG. **2**D, and FIG. **2**E illustrate a method of obtaining presentation information, in accordance with an embodiment. FIG. **2**B discloses SEQ ID NO: 62. FIG. **2**C discloses SEQ ID NOS 62-67, respectively, in order of appearance. FIG. **2**D discloses SEQ ID NO: 157. FIG. **2**E discloses SEQ ID NOS 62-65, 68, and 67, respectively in order of appearance.

[0102] FIG. **3** is a high-level block diagram illustrating the computer logic components of the presentation identification system, according to one embodiment.

[0103] FIG. **4**A illustrates an example set of training data, according to one embodiment related to MHC class I alleles. FIG. **4**A discloses Peptide Sequences as SEQ ID NOS 70-73 and C-Flanking Sequences as SEQ ID NOS 74, 158, 159, and 159, respectively, in order of appearance. FIG. **4**B illustrates an example set of training data, according to one embodiment related to an MHC class II allele. FIG. **4**B discloses SEQ ID NO: 75.

[0104] FIG. **5** illustrates an example network model in association with an MHC allele.

[0105] FIG. **6**A illustrates an example network model NNH(·) shared by MHC alleles, according to one embodiment. FIG. **6**B illustrates an example network model NN.sub.H(·) shared by MHC alleles, according to another embodiment.

[0106] FIG. **7** illustrates generating a presentation likelihood for a peptide in association with an MHC allele using an example network model.

[0107] FIG. **8** illustrates generating a presentation likelihood for a peptide in association with a MHC allele using example network models.

[0108] FIG. **9** illustrates generating a presentation likelihood for a peptide in association with MHC alleles using example network models.

[0109] FIG. **10** illustrates generating a presentation likelihood for a peptide in association with MHC alleles using example network models.

[0110] FIG. **11** illustrates generating a presentation likelihood for a peptide in association with MHC alleles using example network models.

[0111] FIG. **12** illustrates generating a presentation likelihood for a peptide in association with MHC alleles using example network models.

[0112] FIG. **13**A shows performance results for peptide presentation determined by mass-spectrometry comparing various peptide presentation models. Shown are result for the maximum of per-alleles presentation model shown in equation (12) using the affine dependency function g.sub.h(·) and the expit function $f(\cdot)$ and trained on a subset of mass spectrometry data for HLA-A*02:01 and HLA-B*07:02 (“MS”). Also shown are state-of-the-art models based on affinity predictions NETMHCpan “Affinity” and based on stability predictions NETMHCstab “Stability.” The data shows the positive predictive value (PPV) at a 10% recall rate, and error bars (as indicated in solid lines) show 95% confidence intervals.

[0113] FIG. **13**B shows performance results for peptide presentation determined by T-cell epitopes comparing various peptide presentation models. Shown are results for the maximum of per-alleles presentation model shown in equation (12) using the affine dependency function g.sub.h(·) and the expit function $f(\cdot)$ and trained on a subset of mass spectrometry data for HLA-A*02:01. Also shown are state-of-the-art models based on affinity predictions NETMHCpan “Affinity” and based on stability predictions NETMHCstab “Stability.” The data shows the positive predictive value (PPV) at a 10% recall rate, and error bars (as indicated in solid lines) show 95%

confidence intervals.
[0114] FIG. **13**C shows performance results for peptide presentation determined by mass-spectrometry for an example function-of-sums model (equation (13)), an example sum-of-functions model (equation (19)), and an example second order model (equation (23)) for predicting peptide presentation. The first column refers to the the area-under-curve (AUC) of the receiver operating characteristic (ROC) when each presentation model was applied to the test set, the second column refers to the value of the negative log likelihood loss, and the third column refers to the the positive predictive value (PPV) at a 10% recall rate.
[0115] FIG. **13**D shows performance results for peptide presentation determined by mass-spectrometry for two example presentation models that are trained with and without single-allele mass spectrometry data. The first column refers to the the area-under-curve (AUC) of the receiver operating characteristic (ROC) when each presentation model was applied to the test set, the second column refers to the value of the negative log likelihood loss, and the third column refers to the the positive predictive value (PPV) at a 10% recall rate.
[0116] FIG. **13**E shows performance results for peptide presentation determined by mass-spectrometry for two example presentation models that are trained with and without single-allele mass spectrometry data. "Correlation" refers to the correlation between the actual labels that indicate whether the peptide was presented on the corresponding allele in the test data, and the label for prediction.
[0117] FIG. **13**F shows the frequency of common anchor residues at positions 2 (P2) and 9 (P9) among nonamers predicted by a presentation model trained without single-allele mass spectrometry data.
[0118] FIG. **13**G shows performance results for peptide presentation determined by mass-spectrometry for an example presentation model that incorporated C- and N-terminal flanking sequences as allele-interacting variables, and an example presentation model that incorporated C- and N-terminal flanking sequences as allele-noninteracting variables. The first column refers to the the area-under-curve (AUC) of the receiver operating characteristic (ROC) when each presentation model was applied to the test set, the second column refers to the value of the negative log likelihood loss, and the third column refers to the the positive predictive value (PPV) at a 10% recall rate.
[0119] FIG. **13**H shows the dependency between mRNA abundance and the frequency of peptides presented on a tumor cell as determined by mass-spectrometry. The horizontal axis indicates mRNA expression in terms of transcripts per million (TPM) quartiles. The vertical axis indicates fraction of presented epitopes from genes in corresponding mRNA expression quartiles. Each solid line is a plot relating the two measurements from a tumor sample that is associated with corresponding mass spectrometry data and mRNA expression measurements.
[0120] FIG. **13**I shows performance performance results for peptide presentation determined by mass-spectrometry for example presentation models "MHCflurry+RNA filter" is a model similar to the current state-of-the-art model that predicts peptide presentation based on affinity predictions with a standard gene expression filter that removed all peptides from proteins with mRNA quantification measurements that were less than 3.2 FPKM. The "Example Model, no RNA" model is the "sum-of-sigmoids" example presentation model shown in equation (21). The "Example Model, with RNA" model is the "sum-of-sigmoids" presentation model shown in equation (19) incorporating mRNA quantification data through a log function. The data shows the positive predictive value (PPV) at a 20% recall rate.
[0121] FIG. **13**J shows the probability of peptide presentation for different peptide lengths for presentation models that take peptide length into account and state-of-the-art models that do not account for peptide length when predicting peptide presentation. The plot "Truth (Blind Test Data)" showed the proportion of presented peptides according to the length of the peptide in a sample test data set. The plot "Models Ignoring Length" indicated predicted measurements if state-of-the-art models that ignore peptide length applied to the same test data set for presentation prediction. The "Example Model, with RNA" model is the "sum-of-sigmoids" presentation model shown in equation (19) incorporating mRNA quantification data through a log function.
[0122] FIG. **13**K is a histogram of lengths of peptides eluted from class II MHC alleles on human tumor cells and tumor infiltrating lymphocytes (TIL) using mass spectrometry.
[0123] FIG. **13**L illustrates the dependency between mRNA quantification and presented peptides per residue for two example datasets.
[0124] FIG. **13**M compares performance results for example presentation models trained and tested using two example datasets.
[0125] FIG. **13**N is a histogram that depicts the quantity of peptides sequenced using mass spectrometry for each sample of a total of 39 samples comprising HLA class II molecules.
[0126] FIG. **13**O is a histogram that depicts the quantity of samples in which a particular MHC class II molecule allele was identified.
[0127] FIG. **13**P is a histogram that depicts the proportion of peptides presented by the MHC class II molecules in the 39 total samples, for each peptide length of a range of peptide lengths.

[0128] FIG. **13**Q is a line graph that depicts the relationship between gene expression and prevalence of presentation of the gene expression product by a MHC class II molecule, for genes present in the 39 samples.
[0129] FIG. **13**R is a line graph that compares the performance of identical models with varying inputs, at predicting the likelihood that peptides in a testing dataset of peptides will be presented by a MHC class II molecule.
[0130] FIG. **13**S is a line graph that compares the performance of four different models at predicting the likelihood that peptides in a testing dataset of peptides will be presented by a MHC class II molecule.
[0131] FIG. **13**T is a line graph that compares the performance of a best-in-class prior art model using two different criteria and the presentation model disclosed herein with two different inputs, at predicting the likelihood that peptides in a testing dataset of peptides will be presented by a MHC class II molecule.
[0132] FIG. **14** illustrates an example computer for implementing the entities shown in FIGS. **1** and **3**.
[0133] FIG. **15** illustrates development of an in vitro T cell activation assay. Schematic of the assay in which the delivery of a vaccine cassette to antigen presenting cells, leads to expression, processing and MHC-restricted presentation of distinct peptide antigens. Reporter T cells engineered with T cell receptors that match the specific peptide-MHC combination become activated resulting in luciferase expression.
[0134] FIG. **16**A illustrates evaluation of linker sequences in short cassettes and shows five class I MHC restricted epitopes (epitopes 1 through 5) concatenated in the same position relative to each other followed by two universal class II MHC epitopes (MHC-II). Various iterations were generated using different linkers. In some cases the T cell epitopes are directly linked to each other. In others, the T cell epitopes are flanked on one or both sides by its natural sequence. In other iterations, the T cell epitopes are linked by the non-natural sequences AAY, RR, and DPP.
[0135] FIG. **16**B illustrates evaluation of linker sequences in short cassettes and shows sequence information on the T cell epitopes embedded in the short cassettes. Figure discloses SEQ ID NOS 132, 133, 136, 135, 134, 160, and 161, respectively, in order of appearance.
[0136] FIG. **17** illustrates evaluation of cellular targeting sequences added to model vaccine cassettes. The targeting cassettes extend the short cassette designs with ubiquitin (Ub), signal peptides (SP) and/or transmembrane (TM) domains, feature next to the five marker human T cell epitopes (epitopes 1 through 5) also two mouse T cell epitopes SIINFEKL (SII) (SEQ ID NO: 57) and SPSYAYHQF (A5) (SEQ ID NO: 58), and use either the non natural linker AAY- or natural linkers flanking the T cell epitopes on both sides (25mer).
[0137] FIG. **18** illustrates in vivo evaluation of linker sequences in short cassettes. A) Experimental design of the in vivo evaluation of vaccine cassettes using HLA-A2 transgenic mice.
[0138] FIG. **19**A illustrates in vivo evaluation of the impact of epitope position in long 21-mer cassettes and shows the design of long cassettes entails five marker class I epitopes (epitopes 1 through 5) contained in their 25-mer natural sequence (linker=natural flanking sequences), spaced with additional well-known T cell class I epitopes (epitopes 6 through 21) contained in their 25-mer natural sequence, and two universal class II epitopes (MHC-II0, with only the relative position of the class I epitopes varied.
[0139] FIG. **19**B illustrates in vivo evaluation of the impact of epitope position in long 21-mer cassettes and shows the sequence information on the T cell epitopes used. Figure discloses SEQ ID NOS 132, 133, 136, 135, 134, 162-164, 137, and 165-176, respectively, in order of appearance.
[0140] FIG. **20**A illustrates final cassette design for preclinical IND-enabling studies and shows the design of the final cassettes comprises 20 MHC I epitopes contained in their 25-mer natural sequence (linker=natural flanking sequences), composed of 6 non-human primate (NHP) epitopes, 5 human epitopes, 9 murine epitopes, as well as 2 universal MHC class II epitopes.
[0141] FIG. **20**B illustrates final cassette design for preclinical IND-enabling studies and shows the sequence information for the T cell epitopes used that are presented on class I MHC of non-human primate (SEQ ID NOS 177-182, respectively, in order of appearance), mouse (SEQ ID NOS 57, 58 and 183-189, respectively, in order of appearance) and human origin (SEQ ID NOS 134-136, 132, and 133, respectively, in order of appearance), as well as sequences of 2 universal MHC class II epitopes PADRE and Tetanus toxoid (SEQ ID NOS 160 and 190, respectively, in order of appearance).
[0142] FIG. **21**A illustrates ChAdV68.4WTnt.GFP virus production after transfection. HEK293A cells were transfected with ChAdV68.4WTnt.GFP DNA using the calcium phosphate protocol. Viral replication was observed 10 days after transfection and ChAdV68.4WTnt.GFP viral plaques were visualized using light microscopy (40× magnification).
[0143] FIG. **21**B illustrates ChAdV68.4WTnt.GFP virus production after transfection. HEK293A cells were transfected with ChAdV68.4WTnt.GFP DNA using the calcium phosphate protocol. Viral replication was observed 10 days after transfection and ChAdV68.4WTnt.GFP viral plaques were visualized using fluorescent microscopy at 40× magnification.

[0144] FIG. **21**C illustrates ChAdV68.4WTnt.GFP virus production after transfection. HEK293A cells were transfected with ChAdV68.4WTnt.GFP DNA using the calcium phosphate protocol. Viral replication was observed 10 days after transfection and ChAdV68.4WTnt.GFP viral plaques were visualized using fluorescent microscopy at 100× magnification.
[0145] FIG. **22**A illustrates ChAdV68.5WTnt.GFP virus production after transfection. HEK293A cells were transfected with ChAdV68.5WTnt.GFP DNA using the lipofectamine protocol. Viral replication (plaques) was observed 10 days after transfection. A lysate was made and used to reinfect a T25 flask of 293A cells. ChAdV68.5WTnt.GFP viral plaques were visualized and photographed 3 days later using light microscopy (40× magnification)
[0146] FIG. **22**B illustrates ChAdV68.5WTnt.GFP virus production after transfection. HEK293A cells were transfected with ChAdV68.5WTnt.GFP DNA using the lipofectamine protocol. Viral replication (plaques) was observed 10 days after transfection. A lysate was made and used to reinfect a T25 flask of 293A cells. ChAdV68.5WTnt.GFP viral plaques were visualized and photographed 3 days later using fluorescent microscopy at 40× magnification.
[0147] FIG. **22**C illustrates ChAdV68.5WTnt.GFP virus production after transfection. HEK293A cells were transfected with ChAdV68.5WTnt.GFP DNA using the lipofectamine protocol. Viral replication (plaques) was observed 10 days after transfection. A lysate was made and used to reinfect a T25 flask of 293A cells. ChAdV68.5WTnt.GFP viral plaques were visualized and photographed 3 days later using fluorescent microscopy at 100× magnification.
[0148] FIG. **23** illustrates the viral particle production scheme.
[0149] FIG. **24** illustrates the alphavirus derived VEE self-replicating RNA (srRNA) vector.
[0150] FIG. **25** illustrates in vivo reporter expression after inoculation of C57BL/6J mice with VEE-Luciferase srRNA. Shown are representative images of luciferase signal following immunization of C57BL/6J mice with VEE-Luciferase srRNA (10 ug per mouse, bilateral intramuscular injection, MC3 encapsulated) at various timepoints.
[0151] FIG. **26**A illustrates T-cell responses measured 14 days after immunization with VEE srRNA formulated with MC3 LNP in B16-OVA tumor bearing mice. B16-OVA tumor bearing C57BL/6J mice were injected with 10 μg of VEE-Luciferase srRNA (control), VEE-UbAAY srRNA (Vax), VEE-Luciferase srRNA and anti-CTLA-4 (aCTLA-4) or VEE-UbAAY srRNA and anti-CTLA-4 (Vax+aCTLA-4). In addition, all mice were treated with anti-PD1 mAb starting at day 7. Each group consisted of 8 mice. Mice were sacrificed and spleens and lymph nodes were collected 14 days after immunization. SIINFEKL-specific T-cell responses ("SIINFEKL" disclosed as SEQ ID NO: 57) were assessed by IFN-gamma ELISPOT and are reported as spot-forming cells (SFC) per 106 splenocytes. Lines represent medians.
[0152] FIG. **26**B illustrates T-cell responses measured 14 days after immunization with VEE srRNA formulated with MC3 LNP in B16-OVA tumor bearing mice. B16-OVA tumor bearing C57BL/6J mice were injected with 10 μg of VEE-Luciferase srRNA (control), VEE-UbAAY srRNA (Vax), VEE-Luciferase srRNA and anti-CTLA-4 (aCTLA-4) or VEE-UbAAY srRNA and anti-CTLA-4 (Vax+aCTLA-4). In addition, all mice were treated with anti-PD1 mAb starting at day 7. Each group consisted of 8 mice. Mice were sacrificed and spleens and lymph nodes were collected 14 days after immunization. SIINFEKL-specific T-cell responses ("SIINFEKL" disclosed as SEQ ID NO: 57) were assessed by MHCI-pentamer staining, reported as pentamer positive cells as a percent of CD8 positive cells. Lines represent medians.
[0153] FIG. **27**A illustrates antigen-specific T-cell responses following heterologous prime/boost in B16-OVA tumor bearing mice. B16-OVA tumor bearing C57BL/6J mice were injected with adenovirus expressing GFP (Ad5-GFP) and boosted with VEE-Luciferase srRNA formulated with MC3 LNP (Control) or Ad5-UbAAY and boosted with VEE-UbAAY srRNA (Vax). Both the Control and Vax groups were also treated with an IgG control mAb. A third group was treated with the Ad5-GFP prime/VEE-Luciferase srRNA boost in combination with anti-CTLA-4 (aCTLA-4), while the fourth group was treated with the Ad5-UbAAY prime/VEE-UbAAY boost in combination with anti-CTLA-4 (Vax+aCTLA-4). In addition, all mice were treated with anti-PD-1 mAb starting at day 21. T-cell responses were measured by IFN-gamma ELISPOT. Mice were sacrificed and spleens and lymph nodes collected at 14 days post immunization with adenovirus.
[0154] FIG. **27**B illustrates antigen-specific T-cell responses following heterologous prime/boost in B16-OVA tumor bearing mice. B16-OVA tumor bearing C57BL/6J mice were injected with adenovirus expressing GFP (Ad5-GFP) and boosted with VEE-Luciferase srRNA formulated with MC3 LNP (Control) or Ad5-UbAAY and boosted with VEE-UbAAY srRNA (Vax). Both the Control and Vax groups were also treated with an IgG control mAb. A third group was treated with the Ad5-GFP prime/VEE-Luciferase srRNA boost in combination with anti-CTLA-4 (aCTLA-4), while the fourth group was treated with the Ad5-UbAAY prime/VEE-UbAAY boost in combination with anti-CTLA-4 (Vax+aCTLA-4). In addition, all mice were treated with anti-PD-1 mAb starting

at day 21. T-cell responses were measured by IFN-gamma ELISPOT. Mice were sacrificed and spleens and lymph nodes collected at 14 days post immunization with adenovirus and 14 days post boost with srRNA (day 28 after prime).

[0155] FIG. **27**C illustrates antigen-specific T-cell responses following heterologous prime/boost in B16-OVA tumor bearing mice. B16-OVA tumor bearing C57BL/6J mice were injected with adenovirus expressing GFP (Ad5-GFP) and boosted with VEE-Luciferase srRNA formulated with MC3 LNP (Control) or Ad5-UbAAY and boosted with VEE-UbAAY srRNA (Vax). Both the Control and Vax groups were also treated with an IgG control mAb. A third group was treated with the Ad5-GFP prime/VEE-Luciferase srRNA boost in combination with anti-CTLA-4 (aCTLA-4), while the fourth group was treated with the Ad5-UbAAY prime/VEE-UbAAY boost in combination with anti-CTLA-4 (Vax+aCTLA-4). In addition, all mice were treated with anti-PD-1 mAb starting at day 21. T-cell responses were measured by MHC class I pentamer staining. Mice were sacrificed and spleens and lymph nodes collected at 14 days post immunization with adenovirus.

[0156] FIG. **27**D illustrates antigen-specific T-cell responses following heterologous prime/boost in B16-OVA tumor bearing mice. B16-OVA tumor bearing C57BL/6J mice were injected with adenovirus expressing GFP (Ad5-GFP) and boosted with VEE-Luciferase srRNA formulated with MC3 LNP (Control) or Ad5-UbAAY and boosted with VEE-UbAAY srRNA (Vax). Both the Control and Vax groups were also treated with an IgG control mAb. A third group was treated with the Ad5-GFP prime/VEE-Luciferase srRNA boost in combination with anti-CTLA-4 (aCTLA-4), while the fourth group was treated with the Ad5-UbAAY prime/VEE-UbAAY boost in combination with anti-CTLA-4 (Vax+aCTLA-4). In addition, all mice were treated with anti-PD-1 mAb starting at day 21. T-cell responses were measured by MHC class I pentamer staining. Mice were sacrificed and spleens and lymph nodes collected at 14 days post immunization with adenovirus and 14 days post boost with srRNA (day 28 after prime).

[0157] FIG. **28**A illustrates antigen-specific T-cell responses following heterologous prime/boost in CT26 (Balb/c) tumor bearing mice. Mice were immunized with Ad5-GFP and boosted 15 days after the adenovirus prime with VEE-Luciferase srRNA formulated with MC3 LNP (Control) or primed with Ad5-UbAAY and boosted with VEE-UbAAY srRNA (Vax). Both the Control and Vax groups were also treated with an IgG control mAb. A separate group was administered the Ad5-GFP/VEE-Luciferase srRNA prime/boost in combination with anti-PD-1 (aPD1), while a fourth group received the Ad5-UbAAY/VEE-UbAAY srRNA prime/boost in combination with an anti-PD-1 mAb (Vax+aPD1). T-cell responses to the AH1 peptide were measured using IFN-gamma ELISPOT. Mice were sacrificed and spleens and lymph nodes collected at 12 days post immunization with adenovirus.

[0158] FIG. **28**B illustrates antigen-specific T-cell responses following heterologous prime/boost in CT26 (Balb/c) tumor bearing mice. Mice were immunized with Ad5-GFP and boosted 15 days after the adenovirus prime with VEE-Luciferase srRNA formulated with MC3 LNP (Control) or primed with Ad5-UbAAY and boosted with VEE-UbAAY srRNA (Vax). Both the Control and Vax groups were also treated with an IgG control mAb. A separate group was administered the Ad5-GFP/VEE-Luciferase srRNA prime/boost in combination with anti-PD-1 (aPD1), while a fourth group received the Ad5-UbAAY/VEE-UbAAY srRNA prime/boost in combination with an anti-PD-1 mAb (Vax+aPD1). T-cell responses to the AH1 peptide were measured using IFN-gamma ELISPOT. Mice were sacrificed and spleens and lymph nodes collected at 12 days post immunization with adenovirus and 6 days post boost with srRNA (day 21 after prime).

[0159] FIG. **29** illustrates ChAdV68 eliciting T-Cell responses to mouse tumor antigens in mice. Mice were immunized with ChAdV68.5WTnt.MAG25mer, and T-cell responses to the MHC class I epitope SIINFEKL (OVA) (SEQ ID NO: 57) were measured in C57BL/6J female mice and the MHC class I epitope AH1-A5 measured in Balb/c mice. Mean spot forming cells (SFCs) per 10.sup.6 splenocytes measured in ELISpot assays presented. Error bars represent standard deviation.

[0160] FIG. **30** illustrates cellular immune responses in a CT26 tumor model following a single immunization with either ChAdV6, ChAdV+anti-PD-1, srRNA, srRNA+anti-PD-1, or anti-PD-1 alone. Antigen-specific IFN-gamma production was measured in splenocytes for 6 mice from each group using ELISpot. Results are presented as spot forming cells (SFC) per 10.sup.6 splenocytes. Median for each group indicated by horizontal line. P values determined using the Dunnett's multiple comparison test; *** P<0.0001, **P<0.001, *P<0.05. ChAdV=ChAdV68.5WTnt.MAG25mer; srRNA=VEE-MAG25mer srRNA.

[0161] FIG. **31** illustrates CD8 T-Cell responses in a CT26 tumor model following a single immunization with either ChAdV6, ChAdV+anti-PD-1, srRNA, srRNA+anti-PD-1, or anti-PD-1 alone. Antigen-specific IFN-gamma production in CD8 T cells measured using ICS and results presented as antigen-specific CD8 T cells as a percentage of total CD8 T cells. Median for each group indicated by horizontal line. P values determined using the Dunnett's multiple comparison test; *** P<0.0001, **P<0.001, *P<0.05. ChAdV=ChAdV68.5WTnt.MAG25mer; srRNA=VEE-MAG25mer srRNA.

[0162] FIG. **32** illustrates tumor growth in a CT26 tumor model following immunization with a ChAdV/srRNA heterologous prime/boost, a srRNA/ChAdV heterologous prime/boost, or a srRNA/srRNA homologous primer/boost. Also illustrated in a comparison of the prime/boost immunizations with or without administration of anti-PD1 during prime and boost. Tumor volumes measured twice per week and mean tumor volumes presented for the first 21 days of the study. 22-28 mice per group at study initiation. Error bars represent standard error of the mean (SEM). P values determined using the Dunnett's test; *** P<0.0001, **P<0.001, *P<0.05. ChAdV=ChAdV68.5WTnt.MAG25mer; srRNA=VEE-MAG25mer srRNA.
[0163] FIG. **33** illustrates survival in a CT26 tumor model following immunization with a ChAdV/srRNA heterologous prime/boost, a srRNA/ChAdV heterologous prime/boost, or a srRNA/srRNA homologous primer/boost. Also illustrated in a comparison of the prime/boost immunizations with or without administration of anti-PD1 during prime and boost. P values determined using the log-rank test; *** P<0.0001, **P<0.001, *P<0.01. ChAdV=ChAdV68.5WTnt.MAG25mer; srRNA=VEE-MAG25mer srRNA.
[0164] FIG. **34**A, FIG. **34**B, FIG. **34**C, and FIG. **34**D illustrates antigen-specific cellular immune responses measured using ELISpot. Antigen-specific IFN-gamma production to six different mamu A01 restricted epitopes was measured in PBMCs for the VEE-MAG25mer srRNA-LNP1 (30 μg) (FIG. **34**A), VEE-MAG25mer srRNA-LNP1 (100 μg) (FIG. **34**B), or VEE-MAG25mer srRNA-LNP2 (100 μg) (FIG. **34**C) homologous prime/boost or the ChAdV68.5WTnt.MAG25mer/VEE-MAG25mer srRNA heterologous prime/boost group (FIG. **34**D) using ELISpot 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after the first boost immunization (6 rhesus macaques per group). Results are presented as mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope in a stacked bar graph format. Values for each animal were normalized to the levels at pre-bleed (week 0).
[0165] FIG. **35** shows antigen-specific cellular immune response measured using ELISpot. Antigen-specific IFN-gamma production to six different mamu A01 restricted epitopes was measured in PBMCs after immunization with the ChAdV68.5WTnt.MAG25mer/VEE-MAG25mer srRNA heterologous prime/boost regimen using ELISpot prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after the initial immunization. Results are presented as mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope (6 rhesus macaques per group) in a stacked bar graph format.
[0166] FIG. **36** shows antigen-specific cellular immune response measured using ELISpot. Antigen-specific IFN-gamma production to six different mamu A01 restricted epitopes was measured in PBMCs after immunization with the VEE-MAG25mer srRNA LNP2 homologous prime/boost regimen using ELISpot prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, or 15 weeks after the initial immunization. Results are presented as mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope (6 rhesus macaques per group) in a stacked bar graph format.
[0167] FIG. **37** shows antigen-specific cellular immune response measured using ELISpot. Antigen-specific IFN-gamma production to six different mamu A01 restricted epitopes was measured in PBMCs after immunization with the VEE-MAG25mer srRNA LNP1 homologous prime/boost regimen using ELISpot prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, or 15 weeks after the initial immunization. Results are presented as mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope (6 rhesus macaques per group) in a stacked bar graph format.
[0168] FIG. **38** illustrates determining distance metrics for two example cassette sequences. Figure discloses SEQ ID NOs 191 and 192, respectively, in order of appearance.

DETAILED DESCRIPTION

I. Definitions

[0169] 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.
[0170] As used herein the term “antigen” is a substance that induces an immune response.
[0171] 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 nonframeshift 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.
[0172] As used herein the term “tumor neoantigen” is a neoantigen present in a subject's tumor cell or tissue but

not in the subject's corresponding normal cell or tissue.
[0173] As used herein the term “neoantigen-based vaccine” is a vaccine construct based on one or more neoantigens, e.g., a plurality of neoantigens.
[0174] As used herein the term “candidate neoantigen” is a mutation or other aberration giving rise to a new sequence that may represent a neoantigen.
[0175] As used herein the term “coding region” is the portion(s) of a gene that encode protein.
[0176] As used herein the term “coding mutation” is a mutation occurring in a coding region.
[0177] As used herein the term “ORF” means open reading frame.
[0178] As used herein the term “NEO-ORF” is a tumor-specific ORF arising from a mutation or other aberration such as splicing.
[0179] As used herein the term “missense mutation” is a mutation causing a substitution from one amino acid to another.
[0180] 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.
[0181] As used herein the term “frameshift mutation” is a mutation causing a change in the frame of the protein.
[0182] As used herein the term “indel” is an insertion or deletion of one or more nucleic acids.
[0183] 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.
[0184] 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).
[0185] 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).
[0186] 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.
[0187] As used herein the term “non-stop or read-through” is a mutation causing the removal of the natural stop codon.
[0188] As used herein the term “epitope” is the specific portion of an antigen typically bound by an antibody or T cell receptor.
[0189] As used herein the term “immunogenic” is the ability to elicit an immune response, e.g., via T cells, B cells, or both.
[0190] As used herein the term “HLA binding affinity” “MHC binding affinity” means affinity of binding between a specific antigen and a specific MHC allele.
[0191] As used herein the term “bait” is a nucleic acid probe used to enrich a specific sequence of DNA or RNA from a sample.
[0192] As used herein the term “variant” is a difference between a subject's nucleic acids and the reference human genome used as a control.
[0193] As used herein the term “variant call” is an algorithmic determination of the presence of a variant, typically from sequencing.
[0194] As used herein the term “polymorphism” is a germline variant, i.e., a variant found in all DNA-bearing cells of an individual.
[0195] As used herein the term “somatic variant” is a variant arising in non-germline cells of an individual.

[0196] 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.
[0197] As used herein the term “HLA type” is the complement of HLA gene alleles.
[0198] 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.
[0199] 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.
[0200] 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.
[0201] 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.
[0202] 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.
[0203] 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.
[0204] As used herein the term “proteome” is the set of all proteins expressed and/or translated by a cell, group of cells, or individual.
[0205] 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).
[0206] 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.
[0207] As used herein the term “dextramers” is a dextran-based peptide-MHC multimers used for antigen-specific T-cell staining in flow cytometry.
[0208] 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.
[0209] 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).
[0210] 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.
[0211] 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.
[0212] 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.
[0213] 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.
[0214] 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, and smoking history.
[0215] The term “antigen-encoding nucleic acid sequences derived from a tumor” refers to nucleic acid sequences directly extracted 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.
[0216] 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.
[0217] The term “alphavirus backbone” refers to minimal sequence(s) of an alphavirus that allow for self-

replication of the viral genome. Minimal sequences can include conserved sequences for nonstructural protein-mediated amplification, a nonstructural protein 1 (nsP1) gene, a nsP2 gene, a nsP3 gene, a nsP4 gene, and a polyA sequence, as well as sequences for expression of subgenomic viral RNA including a 26S promoter element.
[0218] The term “sequences for nonstructural protein-mediated amplification” includes alphavirus conserved sequence elements (CSE) well known to those in the art. CSEs include, but are not limited to, an alphavirus 5′ UTR, a 51-nt CSE, a 24-nt CSE, or other 26S subgenomic promoter sequence, a 19-nt CSE, and an alphavirus 3′ UTR.
[0219] 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.
[0220] 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.
[0221] 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
[0222] 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.
[0223] 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.
[0224] 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.
[0225] 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.
[0226] 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. Methods of Identifying Neoantigens
[0227] Disclosed herein are methods for identifying neoantigens from a tumor of a subject that are likely to be presented on the cell surface of the tumor or immune cells, 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 tumor nucleotide sequencing data from the tumor cell of the subject, wherein the tumor nucleotide sequencing data is used to obtain data representing peptide

sequences of each of a set of neoantigens, and wherein the peptide sequence of each neoantigen comprises at least one alteration that makes it distinct from the corresponding wild-type peptide sequence; inputting the peptide sequence of each neoantigen into one or more presentation models to generate a set of numerical likelihoods that each of the neoantigens is presented by one or more MHC alleles on the tumor cell surface of the tumor cell of the subject or cells present in the tumor, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and selecting a subset of the set of neoantigens based on the set of numerical likelihoods to generate a set of selected neoantigens.

[0228] The presentation model can comprise a statistical regression or a machine learning (e.g., deep learning) model trained on a set of reference data (also referred to as a training data set) comprising a set of corresponding labels, wherein the set of reference data is obtained from each of a plurality of distinct subjects where optionally some subjects can have a tumor, and wherein the set of reference data comprises at least one of: data representing exome nucleotide sequences from tumor tissue, data representing exome nucleotide sequences from normal tissue, data representing transcriptome nucleotide sequences from tumor tissue, data representing proteome sequences from tumor tissue, and data representing MHC peptidome sequences from tumor tissue, and data representing MHC peptidome sequences from normal tissue. The reference data can further comprise mass spectrometry data, sequencing data, RNA sequencing data, and proteomics data for single-allele cell lines engineered to express a predetermined MHC allele that are subsequently exposed to synthetic protein, normal and tumor human cell lines, and fresh and frozen primary samples, and T cell assays (e.g., ELISPOT). In certain aspects, the set of reference data includes each form of reference data.

[0229] The presentation model can comprise a set of features derived at least in part from the set of reference data, and wherein the set of features comprises at least one of allele dependent-features and allele-independent features. In certain aspects each feature is included.

[0230] Also disclosed herein are methods for generating an output for constructing a personalized cancer vaccine by identifying one or more neoantigens from one or more tumor cells of a subject that are likely to be presented on a surface of the tumor cells. As an example, one such method may comprise the steps of: obtaining at least one of exome, transcriptome, or whole genome nucleotide sequencing data from the tumor cells and normal cells of the subject, wherein the nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of neoantigens identified by comparing the nucleotide sequencing data from the tumor cells and the nucleotide sequencing data from the normal cells, and wherein the peptide sequence of each neoantigen comprises at least one alteration that makes it distinct from the corresponding wild-type, peptide sequence identified from the normal cells of the subject; encoding the peptide sequences of each of the neoantigens into a corresponding numerical vector, each numerical vector including information regarding a plurality of amino acids that make up the peptide sequence and a set of positions of the amino acids in the peptide sequence; inputting the numerical vectors, using a computer processor, into a deep learning presentation model to generate a set of presentation likelihoods for the set of neoantigens, each presentation likelihood in the set representing the likelihood that a corresponding neoantigen is presented by one or more class II MHC alleles on the surface of the tumor cells of the subject, the deep learning presentation model; selecting a subset of the set of neoantigens based on the set of presentation likelihoods to generate a set of selected neoantigens; and generating the output for constructing the personalized cancer vaccine based on the set of selected neoantigens.

[0231] In some embodiments, the presentation model comprises a plurality of parameters identified at least based on a training data set and a function representing a relation between the numerical vector received as an input and the presentation likelihood generated as output based on the numerical vector and the parameters. In certain embodiments, the training data set comprises labels obtained by mass spectrometry measuring presence of peptides bound to at least one class II MHC allele identified as present in at least one of a plurality of samples, training peptide sequences encoded as numerical vectors including information regarding a plurality of amino acids that make up the peptide sequence and a set of positions of the amino acids in the peptide sequence, and at least one HLA allele associated with the training peptide sequences.

[0232] Dendritic cell presentation to naïve T cell features can comprise at least one of: A feature described above. The dose and type of antigen in the vaccine. (e.g., peptide, mRNA, virus, etc.): (1) The route by which dendritic cells (DCs) take up the antigen type (e.g., endocytosis, micropinocytosis); and/or (2) The efficacy with which the antigen is taken up by DCs. The dose and type of adjuvant in the vaccine. The length of the vaccine antigen sequence. The number and sites of vaccine administration. Baseline patient immune functioning (e.g., as measured by history of recent infections, blood counts, etc). For RNA vaccines: (1) the turnover rate of the mRNA protein product in the dendritic cell; (2) the rate of translation of the mRNA after uptake by dendritic cells as measured in in vitro or in vivo experiments; and/or (3) the number or rounds of translation of the mRNA after uptake by dendritic cells as measured by in vivo or in vitro experiments. The presence of protease cleavage motifs in the peptide, optionally giving additional weight to proteases typically expressed in dendritic cells (as measured

by RNA-seq or mass spectrometry). The level of expression of the proteasome and immunoproteasome in typical activated dendritic cells (which may be measured by RNA-seq, mass spectrometry, immunohistochemistry, or other standard techniques). The expression levels of the particular MHC allele in the individual in question (e.g., as measured by RNA-seq or mass spectrometry), optionally measured specifically in activated dendritic cells or other immune cells. The probability of peptide presentation by the particular MHC allele in other individuals who express the particular MHC allele, optionally measured specifically in activated dendritic cells or other immune cells. The probability of peptide presentation by MHC alleles in the same family of molecules (e.g., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) in other individuals, optionally measured specifically in activated dendritic cells or other immune cells.

[0233] Immune tolerance escape features can comprise at least one of: Direct measurement of the self-peptidome via protein mass spectrometry performed on one or several cell types. Estimation of the self-peptidome by taking the union of all k-mer (e.g. 5-25) substrings of self-proteins. Estimation of the self-peptidome using a model of presentation similar to the presentation model described above applied to all non-mutation self-proteins, optionally accounting for germline variants.

[0234] Ranking can be performed using the plurality of neoantigens provided by at least one model based at least in part on the numerical likelihoods. Following the ranking a selecting can be performed to select a subset of the ranked neoantigens according to a selection criteria. After selecting a subset of the ranked peptides can be provided as an output.

[0235] A number of the set of selected neoantigens may be 20.

[0236] The presentation model may represent dependence between presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and likelihood of presentation on the tumor 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.

[0237] A method disclosed herein can also include applying the one or more presentation models to the peptide sequence of the corresponding neoantigen to generate a dependency score for each of the one or more MHC alleles indicating whether the MHC allele will present the corresponding neoantigen based on at least positions of amino acids of the peptide sequence of the corresponding neoantigen.

[0238] A method disclosed herein can also include transforming the dependency scores to generate a corresponding per-allele likelihood for each MHC allele indicating a likelihood that the corresponding MHC allele will present the corresponding neoantigen; and combining the per-allele likelihoods to generate the numerical likelihood.

[0239] The step of transforming the dependency scores can model the presentation of the peptide sequence of the corresponding neoantigen as mutually exclusive.

[0240] A method disclosed herein can also include transforming a combination of the dependency scores to generate the numerical likelihood.

[0241] The step of transforming the combination of the dependency scores can model the presentation of the peptide sequence of the corresponding neoantigen as interfering between MHC alleles.

[0242] The set of numerical likelihoods can be further identified by at least an allele noninteracting feature, and a method disclosed herein can also include applying an allele noninteracting one of the one or more presentation models to the allele noninteracting features to generate a dependency score for the allele noninteracting features indicating whether the peptide sequence of the corresponding neoantigen will be presented based on the allele noninteracting features.

[0243] A method disclosed herein can also include combining the dependency score for each MHC allele in the one or more MHC alleles with the dependency score for the allele noninteracting feature; transforming the combined dependency scores for each MHC allele to generate a corresponding per-allele likelihood for the MHC allele indicating a likelihood that the corresponding MHC allele will present the corresponding neoantigen; and combining the per-allele likelihoods to generate the numerical likelihood.

[0244] A method disclosed herein can also include transforming a combination of the dependency scores for each of the MHC alleles and the dependency score for the allele noninteracting features to generate the numerical likelihood.

[0245] A set of numerical parameters for the presentation model can be trained based on a training data set including at least a set of training peptide sequences identified as present in a plurality of samples and one or more MHC alleles associated with each training peptide sequence, wherein the training peptide sequences are identified through mass spectrometry on isolated peptides eluted from MHC alleles derived from the plurality of samples.

[0246] The samples can also include cell lines engineered to express a single MHC class I or class II allele.

[0247] The samples can also include cell lines engineered to express a plurality of MHC class I or class II alleles.

[0248] The samples can also include human cell lines obtained or derived from a plurality of patients.
[0249] The samples can also include fresh or frozen tumor samples obtained from a plurality of patients.
[0250] The samples can also include fresh or frozen tissue samples obtained from a plurality of patients.
[0251] The samples can also include peptides identified using T-cell assays.
[0252] The training data set can further include data associated with: peptide abundance of the set of training peptides present in the samples; peptide length of the set of training peptides in the samples.
[0253] The training data set may be generated by comparing the set of training peptide sequences via alignment to a database comprising a set of known protein sequences, wherein the set of training protein sequences are longer than and include the training peptide sequences.
[0254] The training data set may be generated based on performing or having performed nucleotide sequencing on a cell line to obtain at least one of exome, transcriptome, or whole genome sequencing data from the cell line, the sequencing data including at least one nucleotide sequence including an alteration.
[0255] The training data set may be generated based on obtaining at least one of exome, transcriptome, and whole genome normal nucleotide sequencing data from normal tissue samples.
[0256] The training data set may further include data associated with proteome sequences associated with the samples.
[0257] The training data set may further include data associated with MHC peptidome sequences associated with the samples.
[0258] The training data set may further include data associated with peptide-MHC binding affinity measurements for at least one of the isolated peptides.
[0259] The training data set may further include data associated with peptide-MHC binding stability measurements for at least one of the isolated peptides.
[0260] The training data set may further include data associated with transcriptomes associated with the samples.
[0261] The training data set may further include data associated with genomes associated with the samples.
[0262] The training peptide sequences may be of lengths within a range of k-mers where k is between 8-15, inclusive for MHC class I or 6-30 inclusive for MHC class II.
[0263] A method disclosed herein can also include encoding the peptide sequence using a one-hot encoding scheme.
[0264] A method disclosed herein can also include encoding the training peptide sequences using a left-padded one-hot encoding scheme.
[0265] A method of treating a subject having a tumor, comprising performing the steps of any of the neoantigen identification methods described herein, and further comprising obtaining a tumor vaccine comprising the set of selected neoantigens, and administering the tumor vaccine to the subject.
[0266] A method disclosed herein can also include identifying one or more T cells that are antigen-specific for at least one of the neoantigens in the subset. In some embodiments, the identification comprises co-culturing the one or more T cells with one or more of the neoantigens in the subset under conditions that expand the one or more antigen-specific T cells. In further embodiments, the identification comprises contacting the one or more T cells with a tetramer comprising one or more of the neoantigens in the subset under conditions that allow binding between the T cell and the tetramer. In even further embodiments, the method disclosed herein can also include identifying one or more T cell receptors (TCR) of the one or more identified T cells. In certain embodiments, identifying the one or more T cell receptors comprises sequencing the T cell receptor sequences of the one or more identified T cells. The method disclosed herein can further comprise genetically engineering a plurality of T cells to express at least one of the one or more identified T cell receptors; culturing the plurality of T cells under conditions that expand the plurality of T cells; and infusing the expanded T cells into the subject. In some embodiments, genetically engineering the plurality of T cells to express at least one of the one or more identified T cell receptors comprises cloning the T cell receptor sequences of the one or more identified T cells into an expression vector; and transfecting each of the plurality of T cells with the expression vector. In some embodiments, the method disclosed herein further comprises culturing the one or more identified T cells under conditions that expand the one or more identified T cells; and infusing the expanded T cells into the subject.
[0267] Also disclosed herein is an isolated T cell that is antigen-specific for at least one selected neoantigen in the subset.
[0268] Also disclosed herein is a methods for manufacturing a tumor vaccine, comprising the steps of: obtaining at least one of exome, transcriptome or whole genome tumor nucleotide sequencing data from the tumor cell of the subject, wherein the tumor nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of neoantigens, and wherein the peptide sequence of each neoantigen comprises at least one alteration that makes it distinct from the corresponding wild-type peptide sequence; inputting the peptide sequence of each neoantigen into one or more presentation models to generate a set of numerical likelihoods that

each of the neoantigens is presented by one or more MHC alleles on the tumor cell surface of the tumor 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 neoantigens based on the set of numerical likelihoods to generate a set of selected neoantigens; and producing or having produced a tumor vaccine comprising the set of selected neoantigens.

[0269] Also disclosed herein is a tumor vaccine including a set of selected neoantigens selected by performing the method comprising the steps of: obtaining at least one of exome, transcriptome or whole genome tumor nucleotide sequencing data from the tumor cell of the subject, wherein the tumor nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of neoantigens, and wherein the peptide sequence of each neoantigen comprises at least one alteration that makes it distinct from the corresponding wild-type peptide sequence; inputting the peptide sequence of each neoantigen into one or more presentation models to generate a set of numerical likelihoods that each of the neoantigens is presented by one or more MHC alleles on the tumor cell surface of the tumor 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 neoantigens based on the set of numerical likelihoods to generate a set of selected neoantigens; and producing or having produced a tumor vaccine comprising the set of selected neoantigens.

[0270] The tumor vaccine may include one or more of a nucleotide sequence, a polypeptide sequence, RNA, DNA, a cell, a plasmid, or a vector.

[0271] The tumor vaccine may include one or more neoantigens presented on the tumor cell surface.

[0272] The tumor vaccine may include one or more neoantigens that is immunogenic in the subject.

[0273] The tumor vaccine may not include one or more neoantigens that induce an autoimmune response against normal tissue in the subject.

[0274] The tumor vaccine may include an adjuvant.

[0275] The tumor vaccine may include an excipient.

[0276] A method disclosed herein may also include selecting neoantigens that have an increased likelihood of being presented on the tumor cell surface relative to unselected neoantigens based on the presentation model.

[0277] A method disclosed herein may also include selecting neoantigens that have an increased likelihood of being capable of inducing a tumor-specific immune response in the subject relative to unselected neoantigens based on the presentation model.

[0278] A method disclosed herein may also include selecting neoantigens that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected neoantigens based on the presentation model, optionally wherein the APC is a dendritic cell (DC).

[0279] A method disclosed herein may also include selecting neoantigens that have a decreased likelihood of being subject to inhibition via central or peripheral tolerance relative to unselected neoantigens based on the presentation model.

[0280] A method disclosed herein may also include selecting neoantigens that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected neoantigens based on the presentation model.

[0281] The exome or transcriptome nucleotide sequencing data may be obtained by performing sequencing on the tumor tissue.

[0282] The sequencing may be next generation sequencing (NGS) or any massively parallel sequencing approach.

[0283] The set of numerical likelihoods may be further identified by at least MHC-allele interacting features comprising at least one of: the predicted affinity with which the MHC allele and the neoantigen encoded peptide bind; the predicted stability of the neoantigen encoded peptide-MHC complex; the sequence and length of the neoantigen encoded peptide; the probability of presentation of neoantigen encoded peptides with similar sequence in cells from other individuals expressing the particular MHC allele as assessed by mass-spectrometry proteomics or other means; the expression levels of the particular MHC allele in the subject in question (e.g. as measured by RNA-seq or mass spectrometry); the overall neoantigen encoded peptide-sequence-independent probability of presentation by the particular MHC allele in other distinct subjects who express the particular MHC allele; the overall neoantigen encoded peptide-sequence-independent probability of presentation by MHC alleles in the same family of molecules (e.g., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) in other distinct subjects.

[0284] The set of numerical likelihoods are further identified by at least MHC-allele noninteracting features comprising at least one of: the C- and N-terminal sequences flanking the neoantigen encoded peptide within its source protein sequence; the presence of protease cleavage motifs in the neoantigen encoded peptide, optionally weighted according to the expression of corresponding proteases in the tumor cells (as measured by RNA-seq or mass spectrometry); the turnover rate of the source protein as measured in the appropriate cell type; the length of the source protein, optionally considering the specific splice variants (“isoforms”) most highly expressed in the

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

[0285] The at least one alteration may be a frameshift or nonframeshift indel, missense or nonsense substitution, splice site alteration, genomic rearrangement or gene fusion, or any genomic or expression alteration giving rise to a neoORF.

[0286] The tumor cell may be 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, and T cell lymphocytic leukemia, non-small cell lung cancer, and small cell lung cancer.

[0287] A method disclosed herein may also include obtaining a tumor vaccine comprising the set of selected neoantigens or a subset thereof, optionally further comprising administering the tumor vaccine to the subject.

[0288] At least one of neoantigens in the set of selected neoantigens, when in polypeptide form, may include at least one of: a binding affinity with MHC with an IC50 value of less than 1000 nM, for MHC Class I polypeptides a length of 8-15, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids, for MHC Class II polypeptides a length of 6-30, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 amino acids, presence of sequence motifs within or near the polypeptide in the parent protein sequence promoting proteasome cleavage, and presence of sequence motifs promoting TAP transport. For MHC Class II, presence of sequence motifs within or near the peptide promoting cleavage by extracellular or lysosomal proteases (e.g., cathepsins) or HLA-DM catalyzed HLA binding.

[0289] Also disclosed herein is a methods for generating a model for identifying one or more neoantigens that are likely to be presented on a tumor cell surface of a tumor cell, comprising the steps of: receiving mass spectrometry data comprising data associated with a plurality of isolated peptides eluted from major histocompatibility complex (MHC) derived from a plurality of samples; obtaining a training data set by at least identifying a set of training peptide sequences present in the samples and one or more MHCs associated with each training peptide sequence; training a set of numerical parameters of a presentation model using the training data

set comprising the training peptide sequences, the presentation model providing a plurality of numerical likelihoods that peptide sequences from the tumor cell are presented by one or more MHC alleles on the tumor cell surface.
[0290] The presentation model may represent dependence between: presence of a particular amino acid at a particular position of a peptide sequence; and likelihood of presentation, by one of the MHC alleles on the tumor cell, of the peptide sequence containing the particular amino acid at the particular position.
[0291] The samples can also include cell lines engineered to express a single MHC class I or class II allele.
[0292] The samples can also include cell lines engineered to express a plurality of MHC class I or class II alleles.
[0293] The samples can also include human cell lines obtained or derived from a plurality of patients.
[0294] The samples can also include fresh or frozen tumor samples obtained from a plurality of patients.
[0295] The samples can also include peptides identified using T-cell assays.
[0296] The training data set may further include data associated with: peptide abundance of the set of training peptides present in the samples; peptide length of the set of training peptides in the samples.
[0297] A method disclosed herein can also include obtaining a set of training protein sequences based on the training peptide sequences by comparing the set of training peptide sequences via alignment to a database comprising a set of known protein sequences, wherein the set of training protein sequences are longer than and include the training peptide sequences.
[0298] A method disclosed herein can also include performing or having performed mass spectrometry on a cell line to obtain at least one of exome, transcriptome, or whole genome nucleotide sequencing data from the cell line, the nucelotide sequencing data including at least one protein sequence including a mutation.
[0299] A method disclosed herein can also include: encoding the training peptide sequences using a one-hot encoding scheme.
[0300] A method disclosed herein can also include obtaining at least one of exome, transcriptome, and whole genome normal nucleotide sequencing data from normal tissue samples; and training the set of parameters of the presentation model using the normal nucleotide sequencing data.
[0301] The training data set may further include data associated with proteome sequences associated with the samples.
[0302] The training data set may further include data associated with MHC peptidome sequences associated with the samples.
[0303] The training data set may further include data associated with peptide-MHC binding affinity measurements for at least one of the isolated peptides.
[0304] The training data set may further include data associated with peptide-MHC binding stability measurements for at least one of the isolated peptides.
[0305] The training data set may further include data associated with transcriptomes associated with the samples.
[0306] The training data set may further include data associated with genomes associated with the samples.
[0307] A method disclosed herein may also include logistically regressing the set of parameters.
[0308] The training peptide sequences may be lengths within a range of k-mers where k is between 8-15, inclusive for MHC class I or 6-30, inclusive for MHC class II.
[0309] A method disclosed herein may also include encoding the training peptide sequences using a left-padded one-hot encoding scheme.
[0310] A method disclosed herein may also include determining values for the set of parameters using a deep learning algorithm.
[0311] Disclosed herein is are methods for identifying one or more neoantigens that are likely to be presented on a tumor cell surface of a tumor cell, comprising executing the steps of: receiving mass spectrometry data comprising data associated with a plurality of isolated peptides eluted from major histocompatibility complex (MHC) derived from a plurality of fresh or frozen tumor samples; obtaining a training data set by at least identifying a set of training peptide sequences present in the tumor samples and presented on one or more MHC alleles associated with each training peptide sequence; obtaining a set of training protein sequences based on the training peptide sequences; and training a set of numerical parameters of a presentation model using the training protein sequences and the training peptide sequences, the presentation model providing a plurality of numerical likelihoods that peptide sequences from the tumor cell are presented by one or more MHC alleles on the tumor cell surface.
[0312] The presentation model may represent dependence between: presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and likelihood of presentation on the tumor 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.
[0313] A method disclosed herein can also include selecting a subset of neoantigens, wherein the subset of

neoantigens is selected because each has an increased likelihood that it is presented on the cell surface of the tumor relative to one or more distinct tumor neoantigens.
[0314] A method disclosed herein can also include selecting a subset of neoantigens, wherein the subset of neoantigens is selected because each has an increased likelihood that it is capable of inducing a tumor-specific immune response in the subject relative to one or more distinct tumor neoantigens.
[0315] A method disclosed herein can also include selecting a subset of neoantigens, wherein the subset of neoantigens is selected because each has an increased likelihood that it is capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to one or more distinct tumor neoantigens, optionally wherein the APC is a dendritic cell (DC).
[0316] A method disclosed herein can also include selecting a subset of neoantigens, wherein the subset of neoantigens is selected because each has a decreased likelihood that it is subject to inhibition via central or peripheral tolerance relative to one or more distinct tumor neoantigens.
[0317] A method disclosed herein can also include selecting a subset of neoantigens, wherein the subset of neoantigens is selected because each has a decreased likelihood that it is capable of inducing an autoimmune response to normal tissue in the subject relative to one or more distinct tumor neoantigens.
[0318] A method disclosed herein can also include selecting a subset of neoantigens, wherein the subset of neoantigens is selected because each has a decreased likelihood that it will be differentially post-translationally modified in tumor cells versus APCs, optionally wherein the APC is a dendritic cell (DC).
[0319] The practice of the methods herein will employ, unless otherwise indicated, conventional methods of protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., T. E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A. L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences,* 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry* 3.sup.rd Ed. (Plenum Press) Vols A and B(1992).
III. Identification of Tumor Specific Mutations in Neoantigens
[0320] 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.
[0321] 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 nonframeshift indel, missense or nonsense substitution, splice site alteration, genomic rearrangement or gene fusion, or any genomic or expression alteration giving rise to a neoORF.
[0322] 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.
[0323] 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.
[0324] 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.
[0325] 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.

[0326] 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.

[0327] 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.

[0328] 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.

[0329] 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)).

[0330] 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 consists of 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.

[0331] 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.
[0332] 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.
[0333] 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.
[0334] 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.
[0335] 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.
[0336] 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.
[0337] 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. Neoantigens
[0338] Neoantigens can include nucleotides or polypeptides. For example, a neoantigen can be an RNA sequence that encodes for a polypeptide sequence. Neoantigens useful in vaccines can therefore include nucleotide sequences or polypeptide sequences.
[0339] 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.
[0340] One or more polypeptides encoded by a neoantigen nucleotide sequence can comprise at least one of: a binding affinity with MHC with an IC50 value of less than 1000 nM, 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.
[0341] One or more neoantigens can be presented on the surface of a tumor.
[0342] One or more neoantigens can be is immunogenic in a subject having a tumor, e.g., capable of eliciting a T cell response or a B cell response in the subject.
[0343] One or more neoantigens that induce an autoimmune response in a subject can be excluded from consideration in the context of vaccine generation for a subject having a tumor.
[0344] The size of at least one neoantigenic peptide molecule 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 neoantigenic peptide molecules are equal to or less than 50 amino acids.

[0345] Neoantigenic peptides and polypeptides can be: for MHC Class 115 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.

[0346] 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 amino acids—thus bypassing the need for computational or in vitro test-based selection of the strongest HLA-presented shorter peptide. In both cases, use of a longer peptide allows endogenous processing by patient cells and may lead to more effective antigen presentation and induction of T cell responses.

[0347] Neoantigenic peptides and polypeptides can be presented on an HLA protein. In some aspects neoantigenic peptides and polypeptides are presented on an HLA protein with greater affinity than a wild-type peptide. In some aspects, a neoantigenic 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.

[0348] In some aspects, neoantigenic peptides and polypeptides do not induce an autoimmune response and/or invoke immunological tolerance when administered to a subject.

[0349] Also provided are compositions comprising at least two or more neoantigenic peptides. In some embodiments the composition contains at least two distinct peptides. At least two distinct peptides can be derived from the same polypeptide. By distinct polypeptides is meant that the peptide vary by length, amino acid sequence, or both. The peptides are derived from any polypeptide known to or have been found to contain a tumor specific mutation. Suitable polypeptides from which the neoantigenic 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. In some aspects the tumor specific mutation is a driver mutation for a particular cancer type.

[0350] Neoantigenic 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, neoantigenic 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).

[0351] 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 trichloracetic 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.

[0352] The peptides and polypeptides can be modified to provide desired attributes other than improved serum half-life. For instance, the ability of the peptides to induce CTL activity can be enhanced by linkage to a sequence which contains at least one epitope that is capable of inducing a T helper cell response. Immunogenic peptides/T helper conjugates can be linked by a spacer molecule. The spacer is typically comprised of relatively small, neutral molecules, such as amino acids or amino acid mimetics, which are substantially uncharged under physiological conditions. The spacers are typically selected from, e.g., Ala, Gly, or other neutral spacers of nonpolar amino acids or neutral polar amino acids. It will be understood that the optionally present spacer need not be comprised of the same residues and thus can be a hetero- or homo-oligomer. When present, the spacer will usually be at least one or two residues, more usually three to six residues. Alternatively, the peptide can be linked to the T helper peptide without a spacer.
[0353] A neoantigenic 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 neoantigenic 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.
[0354] 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.
[0355] In a further aspect a neoantigen includes a nucleic acid (e.g. polynucleotide) that encodes a neoantigenic peptide or portion thereof. The polynucleotide can be, e.g., DNA, cDNA, PNA, CNA, RNA (e.g., mRNA), either single- and/or double-stranded, or native or stabilized forms of polynucleotides, such as, e.g., polynucleotides with a phosphorothiate backbone, or combinations thereof and it may or may not contain introns. A 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

[0356] 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. Vaccine compositions typically comprise a plurality of neoantigens, e.g., selected using a method described herein. Vaccine compositions can also be referred to as vaccines.
[0357] 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 neoantigen 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 neoantigen sequences, 6, 7, 8, 9, 10 11, 12, 13, or 14 different neoantigen sequences, or 12, 13 or 14 different neoantigen sequences.
[0358] 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.
[0359] The vaccine composition can be capable of raising a specific cytotoxic T-cells response and/or a specific helper T-cell response.
[0360] 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.
[0361] Adjuvants are any substance whose admixture into a vaccine composition increases or otherwise modifies the immune response to a neoantigen. Carriers can be scaffold structures, for example a polypeptide or a polysaccharide, to which a neoantigen, is capable of being associated. Optionally, adjuvants are conjugated covalently or non-covalently.
[0362] 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.
[0363] Suitable adjuvants include, but are not limited to 1018 ISS, alum, aluminium salts, Amplivax, AS15, BCG, CP-870,893, CpG7909, CyaA, dSLIM, GM-CSF, IC30, IC31, Imiquimod, ImuFact IMP321, IS Patch, ISS, ISCOMATRIX, Juvlmmune, 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).
[0364] 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.
[0365] 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, sorafinib, 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).
[0366] 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.
[0367] 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 diptheria toxoid are suitable carriers. Alternatively, the carrier can be dextrans for example sepharose.
[0368] 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.
[0369] Neoantigens 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 neoantigen 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 neoantigens, and thereby elicit a host immune (e.g., CTL) response against the peptide(s). Vaccinia vectors and methods useful in immunization protocols are described in, e.g., U.S. Pat. No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al. (*Nature* 351:456-460 (1991)). A wide variety of other vaccine vectors useful for therapeutic administration or immunization of neoantigens, e.g., *Salmonella typhi* vectors, and the like will be apparent to those skilled in the art from the description herein.
V.A. Neoantigen Cassette
[0370] The methods employed for the selection of one or more neoantigens, the cloning and construction of a "cassette" and its insertion into a viral vector are within the skill in the art given the teachings provided herein. By "neoantigen cassette" is meant the combination of a selected neoantigen or plurality of neoantigens and the other regulatory elements necessary to transcribe the neoantigen(s) and express the transcribed product. A neoantigen or plurality of neoantigens 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 neoantigen(s) in a cell transfected with the viral vector. Thus the neoantigen cassette can also contain a selected promoter which is linked to the neoantigen(s) and located, with other, optional regulatory elements, within the selected viral sequences of the recombinant vector.
[0371] Useful promoters can be constitutive promoters or regulated (inducible) promoters, which will enable control of the amount of neoantigen(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.
[0372] The neoantigen 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 neoantigen-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. A neoantigen cassette can also contain such an intron, located between the promoter/enhancer sequence and the neoantigen(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.
[0373] A neoantigen cassette can have one or more neoantigens. 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 neoantigens. Neoantigens can be linked directly to one another. Neoantigens can also be linked to one another with linkers. Neoantigens can be in any orientation relative to one another including N to C or C to N.

[0374] As above stated, the neoantigen cassette can be located in the site of any selected deletion in the viral vector, such as the site of the E1 gene region deletion or E3 gene region deletion, among others which may be selected.

[0375] The neoantigen cassette can be described using the following formula to describe the ordered sequence of each element, from 5′ to 3′:

[00001]$$(P_a\text{-}(L5_b\text{-}N_c\text{-}L3_d)_X)_Z\text{-}(P2_h\text{-}(G5_e\text{-}U_f)_Y)_W\text{-}G3_g$$

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 Nc is a epitope encoding nucleic acid sequence, where for each Y the corresponding Uf is an antigen-encoding nucleic acid sequence. 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.

[0376] 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 (i.e. only the promoter nucleotide sequence provided by the RNA alphavirus backbone is present), 20 MHC class I epitope are present, a 5′ linker is present for each N, a 3′ linker is present for each N, 2 MHC class II epitopes are present, a linker is present linking the two MHC class II epitopes, a linker is present linking the 5′ end of the two MHC class II epitopes to the 3′ linker of the final MHC class I epitope, and a linker is present linking the 3′ end of the two MHC class II epitopes to the to the RNA alphavirus backbone. Examples of linking the 3′ end of the neoantigen cassette to the RNA alphavirus backbone include linking directly to the 3′ UTR elements provided by the RNA alphavirus backbone, such as a 3′ 19-nt CSE. Examples of linking the 5′ end of the neoantigen cassette to the RNA alphavirus backbone include linking directly to a 26S promoter sequence, an alphavirus 5′ UTR, a 51-nt CSE, or a 24-nt CSE.

[0377] Other examples include: where a=1 describing where a promoter other than the promoter nucleotide sequence provided by the 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 RNA alphavirus 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 antigen-encoding nucleic acid sequences; and where g=0 describing the MHC class II antigen-encoding nucleic acid sequence, if present, is directly linked to the RNA alphavirus backbone.

[0378] 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 neoantigen 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 neoantigen 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.

[0379] In examples where more than one MHC class II epitope is present in the same neoantigen 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 neoantigen 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.

[0380] The promoter nucleotide sequences P and/or P2 can be the same as a promoter nucleotide sequence provided by the RNA alphavirus backbone. For example, the promoter sequence provided by the RNA alphavirus backbone, Pn and P2, can each comprise a 26S subgenomic promoter. The promoter nucleotide sequences P and/or P2 can be different from the promoter nucleotide sequence provided by the RNA alphavirus backbone, as well as can be different from each other.

[0381] 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.

[0382] 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.

[0383] 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.

[0384] For each X, each N can encodes a MHC class I epitope 7-15 amino acids in length. For each X, each N can also encodes a MHC class I epitope 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 amino acids in length. For each X, each N can also encodes a MHC class I epitope at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, or at least 30 amino acids in length.

V.B. Immune Checkpoints

[0385] Vectors described herein, such as C68 vectors described herein or alphavirus vectors described herein, can comprise a nucleic acid which encodes at least one neoantigen and the same or a separate vector can comprise a nucleic acid which encodes at least one immune modulator (e.g., an antibody such as an scFv) which binds to and blocks the activity of an immune checkpoint molecule. Vectors can comprise a neoantigen cassette and one or more nucleic acid molecules encoding a checkpoint inhibitor.

[0386] 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, γδ, and memory CD8+(αβ) 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), Nivolumamb (anti-PD1 antibody), CT-O11 (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

[0387] Truncal peptides, meaning those presented by all or most tumor subclones, can be prioritized for inclusion into the vaccine..sup.53 Optionally, if there are no truncal peptides predicted to be presented and immunogenic with high probability, or if the number of truncal peptides predicted to be presented and immunogenic with high probability is small enough that additional non-truncal peptides can be included in the vaccine, then further peptides can be prioritized by estimating the number and identity of tumor subclones and choosing peptides so as to maximize the number of tumor subclones covered by the vaccine..sup.54

V.C.2. Neoantigen Prioritization

[0388] After all of the above above neoantigen filters are applied, more candidate neoantigens may still be

available for vaccine inclusion than the vaccine technology can support. Additionally, uncertainty about various aspects of the neoantigen analysis may remain and tradeoffs may exist between different properties of candidate vaccine neoantigens. Thus, in place of predetermined filters at each step of the selection process, an integrated multi-dimensional model can be considered that places candidate neoantigens in a space with at least the following axes and optimizes selection using an integrative approach. [0389] 1. Risk of auto-immunity or tolerance (risk of germline) (lower risk of auto-immunity is typically preferred) [0390] 2. Probability of sequencing artifact (lower probability of artifact is typically preferred) [0391] 3. Probability of immunogenicity (higher probability of immunogenicity is typically preferred) [0392] 4. Probability of presentation (higher probability of presentation is typically preferred) [0393] 5. Gene expression (higher expression is typically preferred) [0394] 6. Coverage of HLA genes (larger number of HLA molecules involved in the presentation of a set of neoantigens may lower the probability that a tumor will escape immune attack via downregulation or mutation of HLA molecules) [0395] 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 escape)

[0396] Additionally, optionally, neoantigens 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. 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).

V.D. Alphavirus

V.D.1. Alphavirus Biology

[0397] Alphaviruses are members of the family Togaviridae, and are positive-sense single stranded RNA viruses. Alphaviruses can also be referred to as self-replicating RNA or srRNA. 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 Microbrial Review 1994). A natural alphavirus genome is typically around 12 kb 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).

[0398] A model lifecycle of an alphavirus involves several distinct steps (Strauss Microbrial 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-stand 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.

[0399] 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

[0400] Alphaviruses have previously been engineered for use as expression vector systems (Pushko 1997, Rheme 2004). Alphaviruses offer several advantages, particularly in a vaccine setting where heterologous antigen expression can be desired. Due to its ability to self-replicate in the host cytosol, alphavirus vectors are generally able to produce high copy numbers of the expression cassette within a cell resulting in a high level of heterologous antigen production. Additionally, the vectors are generally transient, resulting in improved biosafety as well as reduced induction of immunological tolerance to the vector. The public, in general, also lacks pre-existing immunity to alphavirus vectors as compared to other standard viral vectors, such as human adenovirus. Alphavirus based vectors also generally result in cytotoxic responses to infected cells. Cytotoxicity, to a certain

degree, can be important in a vaccine setting to properly illicit an immune response to the heterologous antigen expressed. However, the degree of desired cytotoxicity can be a balancing act, and thus several attenuated alphaviruses have been developed, including the TC-83 strain of VEE. Thus, an example of a neoantigen expression vector described herein can utilize an alphavirus backbone that allows for a high level of neoantigen expression, elicits a robust immune response to neoantigen, does not elicit an immune response to the vector itself, and can be used in a safe manner. Furthermore, the neoantigen expression cassette can be designed to elicit different levels of an immune response through optimization of which alphavirus sequences the vector uses, including, but not limited to, sequences derived from VEEor its attenuated derivative TC-83.

[0401] 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.

[0402] 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 U.S. Pat. No. 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 a neoantigen expression vector described herein can utilize an alphavirus backbone wherein the structural proteins are replaced by a neoantigen 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

[0403] 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, 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.

V.D.4. Delivery Via Lipid Nanoparticle

[0404] 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 neoantigen cassettes.

[0405] 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 neoantigens of interest to target cells may be reduced if infectious particles are targeted by neutralizing antibodies.

[0406] 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 soulable vitamins.

[0407] 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.
[0408] 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

[0409] Vaccine compositions for delivery of one or more neoantigens (e.g., via a neoantigen 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 neoantigen delivery (See SEQ ID NO: 1). Use of C68 adenovirus derived vectors is described in further detail in U.S. Pat. No. 6,083,716, which is herein incorporated by reference in its entirety, for all purposes.
[0410] In a further aspect, provided herein is a recombinant adenovirus comprising the DNA sequence of a chimpanzee adenovirus such as C68 and a neoantigen 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 neoantigen cassette product in the cell. In this vector, the native chimpanzee E1 gene, and/or E3 gene, and/or E4 gene can be deleted. A neoantigen cassette can be inserted into any of these sites of gene deletion. The neoantigen cassette can include a neoantigen against which a primed immune response is desired.
[0411] In another aspect, provided herein is a mammalian cell infected with a chimpanzee adenovirus such as C68.
[0412] 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.
[0413] In still a further aspect, provided herein is a method for delivering a neoantigen 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 neoantigen cassette.
[0414] Still another aspect provides a method for eliciting 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 a neoantigen cassette that encodes one or more neoantigens from the tumor against which the immune response is targeted.
[0415] 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.
[0416] 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.

[0417] Also disclosed is a vector comprising a chimpanzee adenovirus DNA sequence obtained from SEQ ID NO: 1 and a neoantigen 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 neoantigen 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.

[0418] Also disclosed herein is a host cell transfected with a vector disclosed herein such as a C68 vector engineered to expression a neoantigen 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.

[0419] Also disclosed herein is a method for delivering a neoantigen 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 neoantigen cassette.

[0420] Also disclosed herein is a method for producing a neoantigen comprising introducing a vector disclosed herein into a mammalian cell, culturing the cell under suitable conditions and producing the neoantigen.

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

[0421] 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 U.S. Pat. No. 6,083,716) can be followed to generate a cell line which expresses any selected chimpanzee adenovirus gene.

[0422] 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 U.S. Pat. No. 6,083,716.

[0423] 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.

[0424] 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.

[0425] 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

[0426] The compositions disclosed herein can comprise viral vectors, that deliver at least one neoantigen to cells. Such vectors comprise a chimpanzee adenovirus DNA sequence such as C68 and a neoantigen 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. A neoantigen cassette comprises at least one neoantigen 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.

[0427] 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.

[0428] 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

[0429] 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].

[0430] In the construction of useful chimpanzee adenovirus C68 vectors for delivery of a neoantigen 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

[0431] 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

[0432] 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.

[0433] 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 neoantigen(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.

[0434] 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.

[0435] 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.

[0436] 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.

[0437] 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.

[0438] The cassette comprising neoantigen(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

[0439] Depending upon the chimpanzee adenovirus gene content of the viral vectors employed to carry the neoantigen 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.

[0440] 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.

[0441] 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.

[0442] 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 neoantigen 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

[0443] Assembly of the selected DNA sequences of the adenovirus, the neoantigen 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.

[0444] For example, following the construction and assembly of the desired neoantigen 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-neoantigen sequences in the vector to be replicated and packaged into virion capsids, resulting in the recombinant viral vector particles.

[0445] The resulting recombinant chimpanzee C68 adenoviruses are useful in transferring a neoantigen 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

[0446] The resulting recombinant chimpanzee C68 adenovirus containing the neoantigen 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 neoantigen(s) to a subject in vivo or ex vivo.

[0447] The above-described recombinant vectors are administered to humans according to published methods for gene therapy. A chimpanzee viral vector bearing a neoantigen 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.

[0448] The chimpanzee adenoviral vectors are administered in sufficient amounts to transduce the human cells and to provide sufficient levels of neoantigen 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.
[0449] 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 neoantigen(s) can be monitored to determine the frequency of dosage administration.
[0450] 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 a neoantigen 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.
[0451] 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.
[0452] The levels of immunity to neoantigen(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

[0453] Also provided is a method of inducing 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 neoantigens such as a plurality of neoantigens identified using methods disclosed herein.
[0454] 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.
[0455] A neoantigen can be administered in an amount sufficient to induce a CTL response.
[0456] A neoantigen can be administered alone or in combination with other therapeutic agents. The therapeutic agent is for example, a chemotherapeutic agent, radiation, or immunotherapy. Any suitable therapeutic treatment for a particular cancer can be administered.
[0457] 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.
[0458] The optimum amount of each neoantigen to be included in a vaccine composition and the optimum dosing regimen can be determined. For example, a neoantigen 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.
[0459] A vaccine can be compiled so that the selection, number and/or amount of neoantigens present in the composition is/are tissue, cancer, 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. The selection can be dependent on the specific type of cancer, the status of the 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 neoantigens according to the expression of the neoantigen in the particular patient or adjustments for secondary treatments following a first round or scheme of treatment.
[0460] For a composition to be used as a vaccine for cancer, neoantigens 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 of a patient expresses high amounts of a certain

neoantigen, the respective pharmaceutical composition for treatment of this cancer can be present in high amounts and/or more than one neoantigen specific for this particularly neoantigen or pathway of this neoantigen can be included.

[0461] Compositions comprising a neoantigen can be administered to an individual already suffering from cancer. In therapeutic applications, compositions are administered to a patient in an amount sufficient to elicit an effective CTL response to the tumor antigen and to cure or at least partially arrest symptoms and/or complications. An amount adequate to accomplish this is defined as “therapeutically effective dose.” Amounts effective for this use will depend on, e.g., the composition, the manner of administration, the stage and severity of the disease being treated, the weight and general state of health of the patient, and the judgment of the prescribing physician. It should be kept in mind that compositions can generally be employed in serious disease states, that is, life-threatening or potentially life threatening situations, especially when the cancer has metastasized. In such cases, in view of the minimization of extraneous substances and the relative nontoxic nature of a neoantigen, it is possible and can be felt desirable by the treating physician to administer substantial excesses of these compositions.

[0462] For therapeutic use, administration can begin at the detection or surgical removal of tumors. This is followed by boosting doses until at least symptoms are substantially abated and for a period thereafter.

[0463] 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 the site of surgical exiscion to induce a local immune response to the tumor. Disclosed herein are compositions for parenteral administration which comprise a solution of the neoantigen 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.

[0464] Neoantigens 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 neoantigen 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 neoantigen 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.

[0465] 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.

[0466] 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.

[0467] 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. Nos. 5,279,833; 9,106,309WOAWO 91/06309; and Felgner et al., Proc. Natl. Acad. Sci. USA 84: 7413-7414

(1987).
[0468] Neoantigens 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 neoantigen 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 neoantigens, and thereby elicit a host immune (e.g., CTL) response against the peptide(s). Vaccinia vectors and methods useful in immunization protocols are described in, e.g., U.S. Pat. No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al. (Nature 351:456-460 (1991)). A wide variety of other vaccine vectors useful for therapeutic administration or immunization of neoantigens, e.g., *Salmonella typhi* vectors, and the like will be apparent to those skilled in the art from the description herein.
[0469] 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 cloned into a desired expression vector.
[0470] 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.
[0471] Also disclosed is a method of manufacturing a tumor vaccine, comprising performing the steps of a method disclosed herein; and producing a tumor vaccine comprising a plurality of neoantigens or a subset of the plurality of neoantigens.
[0472] Neoantigens disclosed herein can be manufactured using methods known in the art. For example, a method of producing a neoantigen or a vector (e.g., a vector including at least one sequence encoding one or more neoantigens) disclosed herein can include culturing a host cell under conditions suitable for expressing the neoantigen or vector wherein the host cell comprises at least one polynucleotide encoding the neoantigen or vector, and purifying the neoantigen or vector. Standard purification methods include chromatographic techniques, electrophoretic, immunological, precipitation, dialysis, filtration, concentration, and chromatofocusing techniques.
[0473] 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 a neoantigen 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 neoantigen or vector.

In certain embodiments the isolated polynucleotide can be cDNA.

VII. Neoantigen Use and Administration

[0474] A vaccination protocol can be used to dose a subject with one or more neoantigens. A priming vaccine and a boosting vaccine can be used to dose the subject. The priming vaccine can be based on C68 (e.g., the sequences shown in SEQ ID NO:1 or 2) or srRNA (e.g., the sequences shown in SEQ ID NO:3 or 4) and the boosting vaccine can be based on C68 (e.g., the sequences shown in SEQ ID NO:1 or 2) or srRNA (e.g., the sequences shown in SEQ ID NO:3 or 4). Each vector typically includes a cassette that includes neoantigens. Cassettes can include about 20 neoantigens, 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) 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.

[0475] 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.sup.12 viral particles); one or more injections of self-replicating RNA (srRNA) at low vaccine dose selected from the range 0.001 to 1 ug RNA, in particular 0.1 or 1 ug can be used; or one or more injections of srRNA at high vaccine dose selected from the range 1 to 100 ug RNA, in particular 10 or 100 ug can be used.

[0476] 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.sup.12 viral particles); one or more injections of self-replicating RNA (srRNA) at low vaccine dose selected from the range 0.001 to 1 ug RNA, in particular 0.1 or 1 ug can be used; or one or more injections of srRNA at high vaccine dose selected from the range 1 to 100 ug RNA, in particular 10 or 100 ug can be used.

[0477] Anti-CTLA-4 (e.g., tremelimumab) can also be administered to the subject. For example, anti-CTLA4 can be administered subcutaneously near the site of the intramuscular vaccine injection (ChAdV68 prime or srRNA low doses) to ensure drainage into the same lymph node. Tremelimumab is a selective human IgG2 mAb inhibitor of CTLA-4. Target Anti-CTLA-4 (tremelimumab) subcutaneous dose is typically 70-75 mg (in particular 75 mg) with a dose range of, e.g., 1-100 mg or 5-420 mg.

[0478] 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.

[0479] Immune monitoring can be performed before, during, and/or after vaccine administration. Such monitoring can inform safety and efficacy, among other parameters.

[0480] 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).

[0481] T cell responses can be assessed as part of an immune monitoring protocol. 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.

VIII. Neoantigen Identification

VIII.A. Neoantigen Candidate Identification

[0482] Research methods for NGS analysis of tumor and normal exome and transcriptomes have been described and applied in the neoantigen identification space..sup.6,14,15 The example below considers certain

optimizations for greater sensitivity and specificity for neoantigen identification in the clinical setting. These optimizations can be grouped into two areas, those related to laboratory processes and those related to the NGS data analysis.

VIII.A.1. Laboratory Process Optimizations

[0483] The process improvements presented here address challenges in high-accuracy neoantigen discovery from clinical specimens with low tumor content and small volumes by extending concepts developed for reliable cancer driver gene assessment in targeted cancer panels.sup.16 to the whole-exome and -transcriptome setting necessary for neoantigen identification. Specifically, these improvements include: [0484] 1. Targeting deep (>500×) unique average coverage across the tumor exome to detect mutations present at low mutant allele frequency due to either low tumor content or subclonal state. [0485] 2. Targeting uniform coverage across the tumor exome, with <5% of bases covered at <100×, so that the fewest possible neoantigens are missed, by, for instance: [0486] a. Employing DNA-based capture probes with individual probe QC.sup.17 [0487] b. Including additional baits for poorly covered regions [0488] 3. Targeting uniform coverage across the normal exome, where <5% of bases are covered at <20× so that the fewest neoantigens possible remain unclassified for somatic/germline status (and thus not usable as TSNAs) [0489] 4. To minimize the total amount of sequencing required, sequence capture probes will be designed for coding regions of genes only, as non-coding RNA cannot give rise to neoantigens. Additional optimizations include: [0490] a. supplementary probes for HLA genes, which are GC-rich and poorly captured by standard exome sequencing.sup.18 [0491] b. exclusion of genes predicted to generate few or no candidate neoantigens, due to factors such as insufficient expression, suboptimal digestion by the proteasome, or unusual sequence features. [0492] 5. Tumor RNA will likewise be sequenced at high depth (>100M reads) in order to enable variant detection, quantification of gene and splice-variant (“isoform”) expression, and fusion detection. RNA from FFPE samples will be extracted using probe-based enrichment.sup.19, with the same or similar probes used to capture exomes in DNA.

VIII.A.2. NGS Data Analysis Optimizations

[0493] Improvements in analysis methods address the suboptimal sensitivity and specificity of common research mutation calling approaches, and specifically consider customizations relevant for neoantigen identification in the clinical setting. These include: [0494] 1. Using the HG38 reference human genome or a later version for alignment, as it contains multiple MHC regions assemblies better reflective of population polymorphism, in contrast to previous genome releases. [0495] 2. Overcoming the limitations of single variant callers.sup.20 by merging results from different programs-5 [0496] a. Single-nucleotide variants and indels will be detected from tumor DNA, tumor RNA and normal DNA with a suite of tools including: programs based on comparisons of tumor and normal DNA, such as Strelka.sup.21 and Mutect.sup.22; and programs that incorporate tumor DNA, tumor RNA and normal DNA, such as UNCeqR, which is particularly advantageous in low-purity samples.sup.23 [0497] b. Indels will be determined with programs that perform local re-assembly, such as Strelka and ABRA.sup.24 [0498] c. Structural rearrangements will be determined using dedicated tools such as Pindel.sup.25 or Breakseq.sup.26 [0499] 3. In order to detect and prevent sample swaps, variant calls from samples for the same patient will be compared at a chosen number of polymorphic sites. [0500] 4. Extensive filtering of artefactual calls will be performed, for instance, by: [0501] a. Removal of variants found in normal DNA, potentially with relaxed detection parameters in cases of low coverage, and with a permissive proximity criterion in case of indels [0502] b. Removal of variants due to low mapping quality or low base quality.sup.27. [0503] c. Removal of variants stemming from recurrent sequencing artifacts, even if not observed in the corresponding normal.sup.27. Examples include variants primarily detected on one strand. [0504] d. Removal of variants detected in an unrelated set of controls.sup.27 [0505] 5. Accurate HLA calling from normal exome using one of seq2HLA.sup.28, ATHLATES.sup.29 or Optitype and also combining exome and RNA sequencing data.sup.28. Additional potential optimizations include the adoption of a dedicated assay for HLA typing such as long-read DNA sequencing.sup.30, or the adaptation of a method for joining RNA fragments to retain continuity 31. [0506] 6. Robust detection of neo-ORFs arising from tumor-specific splice variants will be performed by assembling transcripts from RNA-seq data using CLASS.sup.32, Bayesembler.sup.33, StringTie.sup.34 or a similar program in its reference-guided mode (i.e., using known transcript structures rather than attempting to recreate transcripts in their entirety from each experiment). While Cufflinks.sup.35 is commonly used for this purpose, it frequently produces implausibly large numbers of splice variants, many of them far shorter than the full-length gene, and can fail to recover simple positive controls. Coding sequences and nonsense-mediated decay potential will be determined with tools such as SpliceR.sup.36 and MAMBA.sup.37, with mutant sequences re-introduced. Gene expression will be determined with a tool such as Cufflinks.sup.35 or Express (Roberts and Pachter, 2013). Wild-type and mutant-specific expression counts and/or relative levels will be determined with tools developed for these purposes, such as ASE.sup.38 or HTSeq.sup.39. Potential filtering steps include: [0507] a. Removal of candidate neo-ORFs deemed to be insufficiently expressed. [0508] b. Removal of candidate neo-ORFs predicted

to trigger non-sense mediated decay (NMD). [0509] 7. Candidate neoantigens observed only in RNA (e.g., neoORFs) that cannot directly be verified as tumor-specific will be categorized as likely tumor-specific according to additional parameters, for instance by considering: [0510] a. Presence of supporting tumor DNA-only cis-acting frameshift or splice-site mutations [0511] b. Presence of corroborating tumor DNA-only trans-acting mutation in a splicing factor. For instance, in three independently published experiments with R625-mutant SF3B1, the genes exhibiting the most differentially splicing were concordant even though one experiment examined uveal melanoma patients 4, the second a uveal melanoma cell line.sup.41, and the third breast cancer patients.sup.42 [0512] c. For novel splicing isoforms, presence of corroborating "novel" splice-junction reads in the RNASeq data. [0513] d. For novel re-arrangements, presence of corroborating juxta-exon reads in tumor DNA that are absent from normal DNA [0514] e. Absence from gene expression compendium such as GTEx.sup.43 (i.e. making germline origin less likely) [0515] 8. Complementing the reference genome alignment-based analysis by comparing assembled DNA tumor and normal reads (or k-mers from such reads) directly to avoid alignment and annotation based errors and artifacts. (e.g. for somatic variants arising near germline variants or repeat-context indels)

[0516] In samples with poly-adenylated RNA, the presence of viral and microbial RNA in the RNA-seq data will be assessed using RNA CoMPASS' or a similar method, toward the identification of additional factors that may predict patient response.

VIII.B. Isolation and Detection of HLA Peptides

[0517] 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.

[0518] 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.

TABLE-US-00001 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

[0519] 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.

[0520] 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.

[0521] MS2 spectra from each analysis are searched against a protein database using Comet (61, 62) and the peptide identification are scored using Percolator (63-65). Additional sequencing is performed using PEAKS studio (Bioinformatics Solutions Inc.) and other search engines or sequencing methods can be used including spectral matching and de novo sequencing (97).

VIII.B.1. MS Limit of Detection Studies in Support of Comprehensive HLA Peptide Sequencing

[0522] Using the peptide YVYVADVAAK (SEQ ID NO: 59) 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 μmol, 100 fmol, 10 fmol, 1 fmol, and 100 amol. (Table 1) The results are shown in FIG. **1**F. These results indicate that the lowest limit of detection (LoD) is in the attomol range (10.sup.−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.sup.−15).

TABLE-US-00002 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

IX.A. System Overview

[0523] FIG. **2**A is an overview of an environment **100** for identifying likelihoods of peptide presentation in

patients, in accordance with an embodiment. The environment **100** provides context in order to introduce a presentation identification system **160**, itself including a presentation information store **165**.

[0524] The presentation identification system **160** is one or computer models, embodied in a computing system as discussed below with respect to FIG. **14**, that receives peptide sequences associated with a set of MHC alleles and determines likelihoods that the peptide sequences will be presented by one or more of the set of associated MHC alleles. The presentation identification system **160** may be applied to both class I and class II MHC alleles. This is useful in a variety of contexts. One specific use case for the presentation identification system **160** is that it is able to receive nucleotide sequences of candidate neoantigens associated with a set of MHC alleles from tumor cells of a patient **110** and determine likelihoods that the candidate neoantigens will be presented by one or more of the associated MHC alleles of the tumor and/or induce immunogenic responses in the immune system of the patient **110**. Those candidate neoantigens with high likelihoods as determined by system **160** can be selected for inclusion in a vaccine **118**, such an anti-tumor immune response can be elicited from the immune system of the patient **110** providing the tumor cells.

[0525] The presentation identification system **160** determines presentation likelihoods through one or more presentation models. Specifically, the presentation models generate likelihoods of whether given peptide sequences will be presented for a set of associated MHC alleles, and are generated based on presentation information stored in store **165**. For example, the presentation models may generate likelihoods of whether a peptide sequence "YVYVADVAAK (SEQ ID NO: 59) will be presented for the set of alleles HLA-A*02:01, HLA-A*03:01, HLA-B*07:02, HLA-B*08:03, HLA-C*01:04, HLA-A*06:03, HLA-B*01:04 on the cell surface of the sample. The presentation information **165** contains information on whether peptides bind to different types of MHC alleles such that those peptides are presented by MHC alleles, which in the models is determined depending on positions of amino acids in the peptide sequences. The presentation model can predict whether an unrecognized peptide sequence will be presented in association with an associated set of MHC alleles based on the presentation information **165**. As previously mentioned, the presentation models may be applied to both class I and class II MHC alleles.

IX.B. Presentation Information

[0526] FIG. **2** illustrates a method of obtaining presentation information, in accordance with an embodiment. The presentation information **165** includes two general categories of information: allele-interacting information and allele-noninteracting information. Allele-interacting information includes information that influence presentation of peptide sequences that are dependent on the type of MHC allele. Allele-noninteracting information includes information that influence presentation of peptide sequences that are independent on the type of MHC allele.

IX.B.1. Allele-Interacting Information

[0527] Allele-interacting information primarily includes identified peptide sequences that are known to have been presented by one or more identified MHC molecules from humans, mice, etc. Notably, this may or may not include data obtained from tumor samples. The presented peptide sequences may be identified from cells that express a single MHC allele. In this case the presented peptide sequences are generally collected from single-allele cell lines that are engineered to express a predetermined MHC allele and that are subsequently exposed to synthetic protein. Peptides presented on the MHC allele are isolated by techniques such as acid-elution and identified through mass spectrometry. FIG. **2**B shows an example of this, where the example peptide YEMFNDKS (SEQ ID NO:60), presented on the predetermined MHC allele HLA-A*01:01, is isolated and identified through mass spectrometry. FIG. **2**D shows another example of this, where the example peptide YEMFNDKSQRAPDDKMF (SEQ ID NO: 61), presented on the predetermined MHC allele HLA-DRB1*12:01, is isolated and identified through mass spectrometry. Since in these situations peptides are identified through cells engineered to express a single predetermined MHC protein, the direct association between a presented peptide and the MHC protein to which it was bound to is definitively known.

[0528] The presented peptide sequences may also be collected from cells that express multiple MHC alleles. Typically in humans, 6 different types of MHC-I and up to 12 different types of MHC-II molecules are expressed for a cell. Such presented peptide sequences may be identified from multiple-allele cell lines that are engineered to express multiple predetermined MHC alleles. Such presented peptide sequences may also be identified from tissue samples, either from normal tissue samples or tumor tissue samples. In this case particularly, the MHC molecules can be immunoprecipitated from normal or tumor tissue. Peptides presented on the multiple MHC alleles can similarly be isolated by techniques such as acid-elution and identified through mass spectrometry. FIG. **2**C shows an example of this, where the six example peptides, YEMFNDKSF (SEQ ID NO: 62), HROEIFSHDFJ (SEQ ID NO: 63), FJIEJFOESS (SEQ ID NO: 64), NEIOREIREI (SEQ ID NO: 65), JFKSIFEMMSJDSSU (SEQ ID NO: 66), and KNFLENFIESOFI (SEQ ID NO: 67), are presented on identified MHC alleles HLA-A*01:01, HLA-A*02:01, HLA-B*07:02, HLA-B*08:01, HLA-C*01:03, and HLA-C*01:04 and are isolated and identified through mass spectrometry. In another example, FIG. **2**C shows where the six example peptides, YEMFNDKSF

(SEQ ID NO: 62), HROEIFSHDFJ (SEQ ID NO: 63), FJIEJFOESS (SEQ ID NO: 64), NEIOREIREI (SEQ ID NO: 65), JFKSIFEMMSJDSSUIFLKSJFIEIFJ (SEQ ID NO: 68), and KNFLENFIESOFI (SEQ ID NO: 67), are presented on identified class I MHC alleles HLA-A*01:01, HLA-A*02:01, HLA-B*07:02, HLA-B*08:01, and class II MHC alleles HLA-DRB1*10:01, HLA-DRB1:11:01 and are isolated and identified through mass spectrometry. In contrast to single-allele cell lines, in these examples the direct association between a presented peptide and the MHC protein to which it was bound to may be unknown since the bound peptides are isolated from the MHC molecules before being identified.

[0529] Allele-interacting information can also include mass spectrometry ion current which depends on both the concentration of peptide-MHC molecule complexes, and the ionization efficiency of peptides. The ionization efficiency varies from peptide to peptide in a sequence-dependent manner. Generally, ionization efficiency varies from peptide to peptide over approximately two orders of magnitude, while the concentration of peptide-MHC complexes varies over a larger range than that.

[0530] Allele-interacting information can also include measurements or predictions of binding affinity between a given MHC allele and a given peptide (94, 95, 96). One or more affinity models can generate such predictions. For example, going back to the example shown in FIG. **1**D, presentation information **165** may include a binding affinity prediction of 1000 nM between the peptide YEMFNDKSF (SEQ ID NO: 62) and the class I allele HLA-A*01:01. Few peptides with IC50>1000 nm are presented by the MHC, and lower IC50 values increase the probability of presentation. Presentation information **165** may include a binding affinity prediction between the peptide KNFLENFIESOFI (SEQ ID NO: 67) and the class II allele HLA-DRB1:11:01.

[0531] Allele-interacting information can also include measurements or predictions of stability of the MHC complex. One or more stability models that can generate such predictions. More stable peptide-MHC complexes (i.e., complexes with longer half-lives) are more likely to be presented at high copy number on tumor cells and on antigen-presenting cells that encounter vaccine antigen. For example, going back to the example shown in FIG. **2**C, presentation information **165** may include a stability prediction of a half-life of 1h for the class I molecule HLA-A*01:01. Presentation information **165** may also include a stability prediction of a half-life for the class II molecule HLA-DRB1:11:01.

[0532] Allele-interacting information can also include the measured or predicted rate of the formation reaction for the peptide-MHC complex. Complexes that form at a higher rate are more likely to be presented on the cell surface at high concentration.

[0533] Allele-interacting information can also include the sequence and length of the peptide. MHC class I molecules typically prefer to present peptides with lengths between 8 and 15 peptides. 60-80% of presented peptides have length 9. Histograms of presented peptide lengths from several cell lines are shown in FIG. **5**. MHC class II molecules typically prefer to present peptides with lengths between 6-30 peptides.

[0534] Allele-interacting information can also include the presence of kinase sequence motifs on the neoantigen encoded peptide, and the absence or presence of specific post-translational modifications on the neoantigen encoded peptide. The presence of kinase motifs affects the probability of post-translational modification, which may enhance or interfere with MHC binding.

[0535] Allele-interacting information can also include the expression or activity levels of proteins involved in the process of post-translational modification, e.g., kinases (as measured or predicted from RNA seq, mass spectrometry, or other methods).

[0536] Allele-interacting information can also include the probability of presentation of peptides with similar sequence in cells from other individuals expressing the particular MHC allele as assessed by mass-spectrometry proteomics or other means.

[0537] Allele-interacting information can also include the expression levels of the particular MHC allele in the individual in question (e.g. as measured by RNA-seq or mass spectrometry). Peptides that bind most strongly to an MHC allele that is expressed at high levels are more likely to be presented than peptides that bind most strongly to an MHC allele that is expressed at a low level.

[0538] Allele-interacting information can also include the overall neoantigen encoded peptide-sequence-independent probability of presentation by the particular MHC allele in other individuals who express the particular MHC allele.

[0539] Allele-interacting information can also include the overall peptide-sequence-independent probability of presentation by MHC alleles in the same family of molecules (e.g., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) in other individuals. For example, HLA-C molecules are typically expressed at lower levels than HLA-A or HLA-B molecules, and consequently, presentation of a peptide by HLA-C is a priori less probable than presentation by HLA-A or HLA-B. For another example, HLA-DP is typically expressed at lower levels than HLA-DR or HLA-DQ; consequently, presentation of a peptide by HLA-DP is a prior less probable than presentation by HLA-DR or HLA-DQ.

[0540] Allele-interacting information can also include the protein sequence of the particular MHC allele.
[0541] Any MHC allele-noninteracting information listed in the below section can also be modeled as an MHC allele-interacting information.
IX.B.2. Allele-noninteracting Information
[0542] Allele-noninteracting information can include C-terminal sequences flanking the neoantigen encoded peptide within its source protein sequence. For MHC-I, C-terminal flanking sequences may impact proteasomal processing of peptides. However, the C-terminal flanking sequence is cleaved from the peptide by the proteasome before the peptide is transported to the endoplasmic reticulum and encounters MHC alleles on the surfaces of cells. Consequently, MHC molecules receive no information about the C-terminal flanking sequence, and thus, the effect of the C-terminal flanking sequence cannot vary depending on MHC allele type. For example, going back to the example shown in FIG. **2**C, presentation information **165** may include the C-terminal flanking sequence FOEIFNDKSLDKFJI (SEQ ID NO: 69) of the presented peptide FJIEJFOESS (SEQ ID NO: 64) identified from the source protein of the peptide.
[0543] Allele-noninteracting information can also include mRNA quantification measurements. For example, mRNA quantification data can be obtained for the same samples that provide the mass spectrometry training data. As later described in reference to FIG. **13**H, RNA expression was identified to be a strong predictor of peptide presentation. In one embodiment, the mRNA quantification measurements are identified from software tool RSEM. Detailed implementation of the RSEM software tool can be found at Bo Li and Colin N. Dewey. *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome*. BMC Bioinformatics, 12:323, August 2011. In one embodiment, the mRNA quantification is measured in units of fragments per kilobase of transcript per Million mapped reads (FPKM).
[0544] Allele-noninteracting information can also include the N-terminal sequences flanking the peptide within its source protein sequence.
[0545] Allele-noninteracting information can also include the source gene of the peptide sequence. The source gene may be defined as the Ensembl protein family of the peptide sequence. In other examples, the source gene may be defined as the source DNA or the source RNA of the peptide sequence. The source gene can, for example, be represented as a string of nucleotides that encode for a protein, or alternatively be more categorically represented based on a named set of known DNA or RNA sequences that are known to encode specific proteins. In another example, allele-noninteracting information can also include the source transcript or isoform or set of potential source transcripts or isoforms of the peptide sequence drawn from a database such as Ensembl or RefSeq.
[0546] Allele-noninteracting information can also include the presence of protease cleavage motifs in the peptide, optionally weighted according to the expression of corresponding proteases in the tumor cells (as measured by RNA-seq or mass spectrometry). Peptides that contain protease cleavage motifs are less likely to be presented, because they will be more readily degraded by proteases, and will therefore be less stable within the cell.
[0547] Allele-noninteracting information can also include the turnover rate of the source protein as measured in the appropriate cell type. Faster turnover rate (i.e., lower half-life) increases the probability of presentation; however, the predictive power of this feature is low if measured in a dissimilar cell type.
[0548] Allele-noninteracting information can also include the length of the source protein, optionally considering the specific splice variants ("isoforms") most highly expressed in the tumor cells as measured by RNA-seq or proteome mass spectrometry, or as predicted from the annotation of germline or somatic splicing mutations detected in DNA or RNA sequence data.
[0549] Allele-noninteracting information can also include the level of expression of the proteasome, immunoproteasome, thymoproteasome, or other proteases in the tumor cells (which may be measured by RNA-seq, proteome mass spectrometry, or immunohistochemistry). Different proteasomes have different cleavage site preferences. More weight will be given to the cleavage preferences of each type of proteasome in proportion to its expression level.
[0550] Allele-noninteracting information can also include the expression of the source gene of the peptide (e.g., as measured by RNA-seq or mass spectrometry). Possible optimizations include adjusting the measured expression to account for the presence of stromal cells and tumor-infiltrating lymphocytes within the tumor sample. Peptides from more highly expressed genes are more likely to be presented. Peptides from genes with undetectable levels of expression can be excluded from consideration.
[0551] Allele-noninteracting information can also include the probability that the source mRNA of the neoantigen encoded peptide will be subject to nonsense-mediated decay as predicted by a model of nonsense-mediated decay, for example, the model from Rivas et al, Science 2015.
[0552] Allele-noninteracting information can also include the typical tissue-specific expression of the source gene of the peptide during various stages of the cell cycle. Genes that are expressed at a low level overall (as measured

by RNA-seq or mass spectrometry proteomics) but that are known to be expressed at a high level during specific stages of the cell cycle are likely to produce more presented peptides than genes that are stably expressed at very low levels.
[0553] Allele-noninteracting information can also include a comprehensive catalog of features of the source protein as given in e.g. uniProt or PDB www.rcsb.org/pdb/home/home.do. These features may include, among others: the secondary and tertiary structures of the protein, subcellular localization 11, Gene ontology (GO) terms. Specifically, this information may contain annotations that act at the level of the protein, e.g., 5′ UTR length, and annotations that act at the level of specific residues, e.g., helix motif between residues 300 and 310. These features can also include turn motifs, sheet motifs, and disordered residues.
[0554] Allele-noninteracting information can also include features describing the properties of the domain of the source protein containing the peptide, for example: secondary or tertiary structure (e.g., alpha helix vs beta sheet); Alternative splicing.
[0555] Allele-noninteracting information can also include features describing the presence or absence of a presentation hotspot at the position of the peptide in the source protein of the peptide.
[0556] Allele-noninteracting information can also include the probability of presentation of peptides from the source protein of the peptide in question in other individuals (after adjusting for the expression level of the source protein in those individuals and the influence of the different HLA types of those individuals).
[0557] Allele-noninteracting information can also include the probability that the peptide will not be detected or over-represented by mass spectrometry due to technical biases.
[0558] The expression of various gene modules/pathways as measured by a gene expression assay such as RNASeq, microarray(s), targeted panel(s) such as Nanostring, or single/multi-gene representatives of gene modules measured by assays such as RT-PCR (which need not contain the source protein of the peptide) that are informative about the state of the tumor cells, stroma, or tumor-infiltrating lymphocytes (TILs).
[0559] Allele-noninteracting information can also include the copy number of the source gene of the peptide in the tumor cells. For example, peptides from genes that are subject to homozygous deletion in tumor cells can be assigned a probability of presentation of zero.
[0560] Allele-noninteracting information can also include the probability that the peptide binds to the TAP or the measured or predicted binding affinity of the peptide to the TAP. Peptides that are more likely to bind to the TAP, or peptides that bind the TAP with higher affinity are more likely to be presented by MHC-I.
[0561] Allele-noninteracting information can also include the expression level of TAP in the tumor cells (which may be measured by RNA-seq, proteome mass spectrometry, immunohistochemistry). For MHC-I, higher TAP expression levels increase the probability of presentation of all peptides.
[0562] Allele-noninteracting information can also include the presence or absence of tumor mutations, including, but not limited to: [0563] i. Driver mutations in known cancer driver genes such as EGFR, KRAS, ALK, RET, ROS1, TP53, CDKN2A, CDKN2B, NTRK1, NTRK2, NTRK3 [0564] ii. In genes encoding the proteins involved in the antigen presentation machinery (e.g., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 or any of the genes coding for components of the proteasome or immunoproteasome). Peptides whose presentation relies on a component of the antigen-presentation machinery that is subject to loss-of-function mutation in the tumor have reduced probability of presentation.
[0565] Presence or absence of functional germline polymorphisms, including, but not limited to: [0566] i. In genes encoding the proteins involved in the antigen presentation machinery (e.g., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 or any of the genes coding for components of the proteasome or immunoproteasome)
[0567] Allele-noninteracting information can also include tumor type (e.g., NSCLC, melanoma).
[0568] Allele-noninteracting information can also include known functionality of HLA alleles, as reflected by, for instance HLA allele suffixes. For example, the N suffix in the allele name HLA-A*24:09N indicates a null allele that is not expressed and is therefore unlikely to present epitopes; the full HLA allele suffix nomenclature is described at www.ebi.ac.uk/ipd/imgt/hla/nomenclature/suffixes.html.
[0569] Allele-noninteracting information can also include clinical tumor subtype (e.g., squamous lung cancer vs. non-squamous).
[0570] Allele-noninteracting information can also include smoking history.
[0571] Allele-noninteracting information can also include history of sunburn, sun exposure, or exposure to other mutagens.

[0572] Allele-noninteracting information can also include the typical expression of the source gene of the peptide in the relevant tumor type or clinical subtype, optionally stratified by driver mutation. Genes that are typically expressed at high levels in the relevant tumor type are more likely to be presented.
[0573] Allele-noninteracting information can also include the frequency of the mutation in all tumors, or in tumors of the same type, or in tumors from individuals with at least one shared MHC allele, or in tumors of the same type in individuals with at least one shared MHC allele.
[0574] In the case of a mutated tumor-specific peptide, the list of features used to predict a probability of presentation may also include the annotation of the mutation (e.g., missense, read-through, frameshift, fusion, etc.) or whether the mutation is predicted to result in nonsense-mediated decay (NMD). For example, peptides from protein segments that are not translated in tumor cells due to homozygous early-stop mutations can be assigned a probability of presentation of zero. NMD results in decreased mRNA translation, which decreases the probability of presentation.
IX.C. Presentation Identification System
[0575] FIG. **3** is a high-level block diagram illustrating the computer logic components of the presentation identification system **160**, according to one embodiment. In this example embodiment, the presentation identification system **160** includes a data management module **312**, an encoding module **314**, a training module **316**, and a prediction module **320**. The presentation identification system **160** is also comprised of a training data store **170** and a presentation models store **175**. Some embodiments of the model management system **160** have different modules than those described here. Similarly, the functions can be distributed among the modules in a different manner than is described here.
IX.C.1. Data Management Module
[0576] The data management module **312** generates sets of training data **170** from the presentation information **165**. Each set of training data contains a plurality of data instances, in which each data instance i contains a set of independent variables z.sup.i that include at least a presented or non-presented peptide sequence p.sup.i, one or more associated MHC alleles a.sup.i associated with the peptide sequence p.sup.i, and a dependent variable y.sup.i that represents information that the presentation identification system **160** is interested in predicting for new values of independent variables.
[0577] In one particular implementation referred throughout the remainder of the specification, the dependent variable y.sup.i is a binary label indicating whether peptide p.sup.i was presented by the one or more associated MHC alleles a.sup.i. However, it is appreciated that in other implementations, the dependent variable y.sup.i can represent any other kind of information that the presentation identification system **160** is interested in predicting dependent on the independent variables z.sup.i. For example, in another implementation, the dependent variable y.sup.i may also be a numerical value indicating the mass spectrometry ion current identified for the data instance.
[0578] The peptide sequence p.sup.i for data instance i is a sequence of k.sub.i amino acids, in which k.sub.i may vary between data instances i within a range. For example, that range may be 8-15 for MHC class I or 6-30 for MHC class II. In one specific implementation of system **160**, all peptide sequences p.sup.i in a training data set may have the same length, e.g. 9. The number of amino acids in a peptide sequence may vary depending on the type of MHC alleles (e.g., MHC alleles in humans, etc.). The MHC alleles a.sup.i for data instance i indicate which MHC alleles were present in association with the corresponding peptide sequence p.sup.i.
[0579] The data management module **312** may also include additional allele-interacting variables, such as binding affinity b.sup.i and stability s.sup.i predictions in conjunction with the peptide sequences p.sup.i and associated MHC alleles a.sup.i contained in the training data **170**. For example, the training data **170** may contain binding affinity predictions b.sup.i between a peptide p.sup.i and each of the associated MHC molecules indicated in a.sup.i. As another example, the training data **170** may contain stability predictions s.sup.i for each of the MHC alleles indicated in a.sup.i.
[0580] The data management module **312** may also include allele-noninteracting variables w.sup.i, such as C-terminal flanking sequences and mRNA quantification measurements in conjunction with the peptide sequences p.sup.i.
[0581] The data management module **312** also identifies peptide sequences that are not presented by MHC alleles to generate the training data **170**. Generally, this involves identifying the “longer” sequences of source protein that include presented peptide sequences prior to presentation. When the presentation information contains engineered cell lines, the data management module **312** identifies a series of peptide sequences in the synthetic protein to which the cells were exposed to that were not presented on MHC alleles of the cells. When the presentation information contains tissue samples, the data management module **312** identifies source proteins from which presented peptide sequences originated from, and identifies a series of peptide sequences in the source protein that were not presented on MHC alleles of the tissue sample cells.
[0582] The data management module **312** may also artificially generate peptides with random sequences of amino

acids and identify the generated sequences as peptides not presented on MHC alleles. This can be accomplished by randomly generating peptide sequences allows the data management module **312** to easily generate large amounts of synthetic data for peptides not presented on MHC alleles. Since in reality, a small percentage of peptide sequences are presented by MHC alleles, the synthetically generated peptide sequences are highly likely not to have been presented by MHC alleles even if they were included in proteins processed by cells.

[0583] FIG. **4**A illustrates an example set of training data **170**A, according to one embodiment. Specifically, the first 3 data instances in the training data **170**A indicate peptide presentation information from a single-allele cell line involving the allele HLA-C*01:03 and 3 peptide sequences QCEIOWARE (SEQ ID NO: 70), FIEUHFWI (SEQ ID NO: 71), and FEWRHRJTRUJR (SEQ ID NO: 72). The fourth data instance in the training data **170**A indicates peptide information from a multiple-allele cell line involving the alleles HLA-B*07:02, HLA-C*01:03, HLA-A*01:01 and a peptide sequence QIEJOEIJE (SEQ ID NO: 73). The first data instance indicates that peptide sequence QCEIOWARE (SEQ ID NO: 70) was not presented by the allele HLA-C*01:03. As discussed in the prior two paragraphs, the peptide sequence may be randomly generated by the data management module **312** or identified from source protein of presented peptides. The training data **170**A also includes a binding affinity prediction of 1000 nM and a stability prediction of a half-life of 1h for the peptide sequence-allele pair. The training data **170**A also includes allele-noninteracting variables, such as the C-terminal flanking sequence of the peptide FJELFISBOSJFIE (SEQ ID NO: 74), and a mRNA quantification measurement of 10.sup.2 TPM. The fourth data instance indicates that peptide sequence QIEJOEIJE (SEQ ID NO: 73) was presented by one of the alleles HLA-B*07:02, HLA-C*01:03, or HLA-A*01:01. The training data **170**A also includes binding affinity predictions and stability predictions for each of the alleles, as well as the C-terminal flanking sequence of the peptide and the mRNA quantification measurement for the peptide.

[0584] FIG. **4**B illustrates another example set of training data **170**A, according to one embodiment. Specifically, the first data instances in the training data **170**A indicate peptide presentation information from a single-allele cell line involving the class II allele HLA-DRB3:01:01 and the peptide sequence QCEIOWAREFLKEIGJ (SEQ ID NO: 75). The first data instance indicates that peptide sequence QCEIOWAREFLKEIGJ (SEQ ID NO: 75) was not presented by the allele HLA-DRB3:01:01.

IX.C.2. Encoding Module

[0585] The encoding module **314** encodes information contained in the training data **170** into a numerical representation that can be used to generate the one or more presentation models. In one implementation, the encoding module **314** one-hot encodes sequences (e.g., peptide sequences or C-terminal flanking sequences) over a predetermined 20-letter amino acid alphabet. Specifically, a peptide sequence p.sup.i with k.sub.i amino acids is represented as a row vector of 20.Math.k.sub.i elements, where a single element among p.sup.i.sub.20.Math.(j−1)+1, p.sup.i.sub.20.Math.(j−1)+2, . . . , p.sup.i.sub.20.Math.j that corresponds to the alphabet of the amino acid at the j-th position of the peptide sequence has a value of 1. Otherwise, the remaining elements have a value of 0. As an example, for a given alphabet {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y}, the peptide sequence EAF of 3 amino acids for data instance i may be represented by the row vector of 60 elements p.sup.i=[0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]. The C-terminal flanking sequence c.sup.i can be similarly encoded as described above, as well as the protein sequence d.sub.h for MHC alleles, and other sequence data in the presentation information.

[0586] When the training data **170** contains sequences of differing lengths of amino acids, the encoding module **314** may further encode the peptides into equal-length vectors by adding a PAD character to extend the predetermined alphabet. For example, this may be performed by left-padding the peptide sequences with the PAD character until the length of the peptide sequence reaches the peptide sequence with the greatest length in the training data **170**. Thus, when the peptide sequence with the greatest length has k.sub.max amino acids, the encoding module **314** numerically represents each sequence as a row vector of (20+1).Math.k.sub.max elements. As an example, for the extended alphabet {PAD, A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y} and a maximum amino acid length of k.sub.max=5, the same example peptide sequence EAF of 3 amino acids may be represented by the row vector of 105 elements p.sup.i=[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]. The C-terminal flanking sequence c.sup.i or other sequence data can be similarly encoded as described above. Thus, each independent variable or column in the peptide sequence p.sup.i or c.sup.i represents presence of a particular amino acid at a particular position of the sequence.

[0587] Although the above method of encoding sequence data was described in reference to sequences having amino acid sequences, the method can similarly be extended to other types of sequence data, such as DNA or RNA sequence data, and the like.

[0588] The encoding module **314** also encodes the one or more MHC alleles a.sup.i for data instance i as a row vector of m elements, in which each element h=1, 2, . . . , m corresponds to a unique identified MHC allele. The

elements corresponding to the MHC alleles identified for the data instance i have a value of 1. Otherwise, the remaining elements have a value of 0. As an example, the alleles HLA-B*07:02 and HLA-C*01:03 for a data instance i corresponding to a multiple-allele cell line among m=4 unique identified MHC allele types {HLA-A*01:01, HLA-C*01:08, HLA-B*07:02, HLA-C*01:03} may be represented by the row vector of 4 elements a.sup.i=[0 0 1 1], in which a.sub.3.sup.i=1 and a.sub.4.sup.i=1. As another example, the elements corresponding to the MHC alleles identified for the data instance i have a value of 1. Otherwise, the remaining elements have a value of 0. As an example, the alleles HLA-B*07:02 and HLA-DRB1*10:01 for a data instance i corresponding to a multiple-allele cell line among m=4 unique identified MHC allele types {HLA-A*01:01, HLA-C*01:08, HLA-B*07:02, HLA-DRB1*10:01} may be represented by the row vector of 4 elements a.sup.i=[0 0 1 1], in which a.sub.3.sup.i=1 and a.sub.4.sup.i=1. Although the examples described herein with 4 identified MHC allele types, the number of MHC allele types can be hundreds or thousands in practice. As previously discussed, each data instance i typically contains at most 6 different MHC class I allele types in association with the peptide sequence p.sup.i, and/or at most 4 different MHC class II DR allele types in association with the peptide sequence p.sub.i, and/or at most 12 different MHC class II allele types in association with the peptide sequence p.sub.i.

[0589] The encoding module **314** also encodes the label y.sub.i for each data instance i as a binary variable having values from the set of {0, 1}, in which a value of 1 indicates that peptide x.sup.i was presented by one of the associated MHC alleles a.sup.i, and a value of 0 indicates that peptide x.sup.i was not presented by any of the associated MHC alleles a.sup.i. When the dependent variable y.sub.i represents the mass spectrometry ion current, the encoding module **314** may additionally scale the values using various functions, such as the log function having a range of [−∞, ∞] for ion current values between [0, ∞].

[0590] The encoding module **314** may represent a pair of allele-interacting variables x.sub.h.sup.i for peptide p.sub.i and an associated MHC allele h as a row vector in which numerical representations of allele-interacting variables are concatenated one after the other. For example, the encoding module **314** may represent x.sub.h.sup.i as a row vector equal to [p.sup.i], [p.sup.i b.sub.h.sup.i], [p.sup.i s.sub.h.sup.i], or [p.sup.i b.sub.h.sup.i s.sub.h.sup.i], where b.sub.h.sup.i is the binding affinity prediction for peptide p.sup.i and associated MHC allele h, and similarly for s.sub.h.sup.i for stability. Alternatively, one or more combination of allele-interacting variables may be stored individually (e.g., as individual vectors or matrices).

[0591] In one instance, the encoding module **314** represents binding affinity information by incorporating measured or predicted values for binding affinity in the allele-interacting variables x.sub.h.sup.i.

[0592] In one instance, the encoding module **314** represents binding stability information by incorporating measured or predicted values for binding stability in the allele-interacting variables x.sub.h.sup.i,

[0593] In one instance, the encoding module **314** represents binding on-rate information by incorporating measured or predicted values for binding on-rate in the allele-interacting variables x.sub.h.sup.i.

[0594] In one instance, for peptides presented by class I MHC molecules, the encoding module **314** represents peptide length as a vector T.sub.k=[custom-character(L.sub.k=8) custom-character(L.sub.k=9) custom-character(L.sub.k=10) custom-character(L.sub.k=11) custom-character(L.sub.k=12) custom-character(L.sub.k=13) custom-character(L.sub.k=14) custom-character(L.sub.k=15)] where custom-character is the indicator function, and L.sub.k denotes the length of peptide p.sub.k. The vector T.sub.k can be included in the allele-interacting variables x.sub.h.sup.i. In another instance, for peptides presented by class II MHC molecules, the encoding module **314** represents peptide length as a vector T.sub.k=[custom-character(L.sub.k=6) custom-character(L.sub.k=7) custom-character(L.sub.k=8) custom-character(L.sub.k=9) custom-character(L.sub.k=10) custom-character(L.sub.k=11) custom-character(L.sub.k=12) custom-character(L.sub.k=13) custom-character(L.sub.k=14) custom-character(L.sub.k=15) custom-character(L.sub.k=16) custom-character(L.sub.k=17) custom-character(L.sub.k=18) custom-character(L.sub.k=19) custom-character(L.sub.k=20) custom-character(L.sub.k=21) custom-character(L.sub.k=22) custom-character(L.sub.k=23) custom-character(L.sub.k=24) custom-character(L.sub.k=25) custom-character(L.sub.k=26) custom-character(L.sub.k=27) custom-character(L.sub.k=28) custom-character(L.sub.k=29) custom-character(L.sub.k=30)] where custom-character is the indicator function, and L.sub.k denotes the length of peptide p.sub.k. The vector T.sub.k can be included in the allele-interacting variables x.sub.h.sup.i.

[0595] In one instance, the encoding module **314** represents RNA expression information of MHC alleles by incorporating RNA-seq based expression levels of MHC alleles in the allele-interacting variables x.sub.h.sup.i.

[0596] Similarly, the encoding module **314** may represent the allele-noninteracting variables w.sup.i as a row vector in which numerical representations of allele-noninteracting variables are concatenated one after the other. For example, w.sup.i may be a row vector equal to [c.sup.i] or [c.sup.i m.sup.i w.sup.i] in which w.sup.i is a row vector representing any other allele-noninteracting variables in addition to the C-terminal flanking sequence of peptide p.sup.i and the mRNA quantification measurement m.sup.i associated with the peptide. Alternatively, one

or more combination of allele-noninteracting variables may be stored individually (e.g., as individual vectors or matrices).
[0597] In one instance, the encoding module **314** represents turnover rate of source protein for a peptide sequence by incorporating the turnover rate or half-life in the allele-noninteracting variables w.sup.i.
[0598] In one instance, the encoding module **314** represents length of source protein or isoform by incorporating the protein length in the allele-noninteracting variables w.sup.i.
[0599] In one instance, the encoding module **314** represents activation of immunoproteasome by incorporating the mean expression of the immunoproteasome-specific proteasome subunits including the β1.sub.i, β2.sub.i, β5.sub.i subunits in the allele-noninteracting variables w.sup.i.
[0600] In one instance, the encoding module **314** represents the RNA-seq abundance of the source protein of the peptide or gene or transcript of a peptide (quantified in units of FPKM, TPM by techniques such as RSEM) can be incorporating the abundance of the source protein in the allele-noninteracting variables w.sup.i.
[0601] In one instance, the encoding module **314** represents the probability that the transcript of origin of a peptide will undergo nonsense-mediated decay (NMD) as estimated by the model in, for example, Rivas et. al. *Science*, 2015 by incorporating this probability in the allele-noninteracting variables w.sup.i.
[0602] In one instance, the encoding module **314** represents the activation status of a gene module or pathway assessed via RNA-seq by, for example, quantifying expression of the genes in the pathway in units of TPM using e.g., RSEM for each of the genes in the pathway then computing a summary statistics, e.g., the mean, across genes in the pathway. The mean can be incorporated in the allele-noninteracting variables w.sup.i.
[0603] In one instance, the encoding module **314** represents the copy number of the source gene by incorporating the copy number in the allele-noninteracting variables w.sup.i.
[0604] In one instance, the encoding module **314** represents the TAP binding affinity by including the measured or predicted TAP binding affinity (e.g., in nanomolar units) in the allele-noninteracting variables w.sup.i.
[0605] In one instance, the encoding module **314** represents TAP expression levels by including TAP expression levels measured by RNA-seq (and quantified in units of TPM by e.g., RSEM) in the allele-noninteracting variables w.sup.i.
[0606] In one instance, the encoding module **314** represents tumor mutations as a vector of indicator variables (i.e., d.sup.k=1 if peptide p.sup.k comes from a sample with a KRAS G12D mutation and 0 otherwise) in the allele-noninteracting variables w.sup.i.
[0607] In one instance, the encoding module **314** represents germline polymorphisms in antigen presentation genes as a vector of indicator variables (i.e., d.sup.k=1 if peptide p.sup.k comes from a sample with a specific germline polymorphism in the TAP). These indicator variables can be included in the allele-noninteracting variables w.sup.i.
[0608] In one instance, the encoding module **314** represents tumor type as a length-one one-hot encoded vector over the alphabet of tumor types (e.g., NSCLC, melanoma, colorectal cancer, etc). These one-hot-encoded variables can be included in the allele-noninteracting variables w.sup.i.
[0609] In one instance, the encoding module **314** represents MHC allele suffixes by treating 4-digit HLA alleles with different suffixes. For example, HLA-A*24:09N is considered a different allele from HLA-A*24:09 for the purpose of the model. Alternatively, the probability of presentation by an N-suffixed MHC allele can be set to zero for all peptides, because HLA alleles ending in the N suffix are not expressed.
[0610] In one instance, the encoding module **314** represents tumor subtype as a length-one one-hot encoded vector over the alphabet of tumor subtypes (e.g., lung adenocarcinoma, lung squamous cell carcinoma, etc). These one hot-encoded variables can be included in the allele-noninteracting variables w.sup.i.
[0611] In one instance, the encoding module **314** represents smoking history as a binary indicator variable (d.sup.k=1 if the patient has a smoking history, and 0 otherwise), that can be included in the allele-noninteracting variables w.sup.i. Alternatively, smoking history can be encoded as a length-one one-hot-encoded variable over an alphabet of smoking severity. For example, smoking status can be rated on a 1-5 scale, where 1 indicates nonsmokers, and 5 indicates current heavy smokers. Because smoking history is primarily relevant to lung tumors, when training a model on multiple tumor types, this variable can also be defined to be equal to 1 if the patient has a history of smoking and the tumor type is lung tumors and zero otherwise.
[0612] In one instance, the encoding module **314** represents sunburn history as a binary indicator variable (d.sup.k=1 if the patient has a history of severe sunburn, and 0 otherwise), which can be included in the allele-noninteracting variables w.sup.i. Because severe sunburn is primarily relevant to melanomas, when training a model on multiple tumor types, this variable can also be defined to be equal to 1 if the patient has a history of severe sunburn and the tumor type is melanoma and zero otherwise.
[0613] In one instance, the encoding module **314** represents distribution of expression levels of a particular gene or transcript for each gene or transcript in the human genome as summary statistics (e,g., mean, median) of

distribution of expression levels by using reference databases such as TCGA. Specifically, for a peptide p.sup.k in a sample with tumor type melanoma, we can include not only the measured gene or transcript expression level of the gene or transcript of origin of peptide p.sup.k in the allele-noninteracting variables w.sup.i, but also the mean and/or median gene or transcript expression of the gene or transcript of origin of peptide p.sup.k in melanomas as measured by TCGA.

[0614] In one instance, the encoding module **314** represents mutation type as a length-one one-hot-encoded variable over the alphabet of mutation types (e.g., missense, frameshift, NMD-inducing, etc). These onehot-encoded variables can be included in the allele-noninteracting variables w.sup.i.

[0615] In one instance, the encoding module **314** represents protein-level features of protein as the value of the annotation (e.g., 5′ UTR length) of the source protein in the allele-noninteracting variables w.sup.i. In another instance, the encoding module **314** represents residue-level annotations of the source protein for peptide p.sup.i by including an indicator variable, that is equal to 1 if peptide p.sup.i overlaps with a helix motif and 0 otherwise, or that is equal to 1 if peptide p.sup.i is completely contained with within a helix motif in the allele-noninteracting variables w.sup.i. In another instance, a feature representing proportion of residues in peptide p.sup.i that are contained within a helix motif annotation can be included in the allele-noninteracting variables w.sup.i.

[0616] In one instance, the encoding module **314** represents type of proteins or isoforms in the human proteome as an indicator vector o.sup.k that has a length equal to the number of proteins or isoforms in the human proteome, and the corresponding element o.sup.k.sub.i is 1 if peptide p.sup.k comes from protein i and 0 otherwise.

[0617] In one instance, the encoding module **314** represents the source gene G=gene(p.sup.i) of peptide p.sup.i as a categorical variable with L possible categories, where L denotes the upper limit of the number of indexed source genes 1, 2, . . . , L.

[0618] The encoding module **314** may also represent the overall set of variables z.sup.i for peptide p.sup.i and an associated MHC allele h as a row vector in which numerical representations of the allele-interacting variables x.sup.i and the allele-noninteracting variables w.sup.i are concatenated one after the other. For example, the encoding module **314** may represent z.sub.h.sup.i as a row vector equal to [x.sub.h.sup.i w.sup.i] or [w.sup.i x.sub.h.sup.i].

X. Training Module

[0619] The training module **316** constructs one or more presentation models that generate likelihoods of whether peptide sequences will be presented by MHC alleles associated with the peptide sequences. Specifically, given a peptide sequence p.sup.k and a set of MHC alleles a.sup.k associated with the peptide sequence p.sup.k, each presentation model generates an estimate u.sub.k indicating a likelihood that the peptide sequence p.sup.k will be presented by one or more of the associated MHC alleles a.sup.k.

X.A. Overview

[0620] The training module **316** constructs the one more presentation models based on the training data sets stored in store **170** generated from the presentation information stored in 165. Generally, regardless of the specific type of presentation model, all of the presentation models capture the dependence between independent variables and dependent variables in the training data **170** such that a loss function is minimized. Specifically, the loss function custom-character(y.sub.i∈S, u.sub.i∈S; θ) represents discrepancies between values of dependent variables y.sub.i∈S for one or more data instances S in the training data **170** and the estimated likelihoods u.sub.i∈S for the data instances S generated by the presentation model. In one particular implementation referred throughout the remainder of the specification, the loss function (y.sub.i∈S, u.sub.i∈S; θ) is the negative log likelihood function given by equation (1a) as follows:

[00002] $$\ell(y_{i \in S}, u_{i \in S}; \theta) = \underset{i \in S}{\text{.Math.}} (y_i \log u_i + (1 - y_l)\log(1 - u_i)). \quad (1a)$$

[0621] However, in practice, another loss function may be used. For example, when predictions are made for the mass spectrometry ion current, the loss function is the mean squared loss given by equation 1b as follows:

[00003] $$\ell(y_{i \in S}, u_{i \in S}; \theta) = \underset{i \in S}{\text{.Math.}} (\text{.Math.}\ y_i - u_i \ \text{.Math.}\ {}_2^2). \quad (1b)$$

[0622] The presentation model may be a parametric model in which one or more parameters θ mathematically specify the dependence between the independent variables and dependent variables. Typically, various parameters of parametric-type presentation models that minimize the loss function (y.sub.i∈S, u.sub.i∈S; θ) are determined through gradient-based numerical optimization algorithms, such as batch gradient algorithms, stochastic gradient algorithms, and the like. Alternatively, the presentation model may be a non-parametric model in which the model structure is determined from the training data **170** and is not strictly based on a fixed set of parameters.

X.B. Per-Allele Models

[0623] The training module **316** may construct the presentation models to predict presentation likelihoods of

peptides on a per-allele basis. In this case, the training module **316** may train the presentation models based on data instances S in the training data **170** generated from cells expressing single MHC alleles.

[0624] In one implementation, the training module **316** models the estimated presentation likelihood u.sub.k for peptide p.sup.k for a specific allele h by:

[00004] $u_k^h = \Pr(p^k \text{ presented}; \text{MHCallele} h) = f(g_h(x_h^k; \theta_h))$, (2)

where peptide sequence x.sub.h.sup.k denotes the encoded allele-interacting variables for peptide p.sup.k and corresponding MHC allele h, $f(\cdot)$ is any function, and is herein throughout is referred to as a transformation function for convenience of description. Further, g.sub.h(·) is any function, is herein throughout referred to as a dependency function for convenience of description, and generates dependency scores for the allele-interacting variables x.sub.h.sup.k based on a set of parameters θh determined for MHC allele h. The values for the set of parameters θh for each MHC allele h can be determined by minimizing the loss function with respect to θh, where i is each instance in the subset S of training data **170** generated from cells expressing the single MHC allele h.

[0625] The output of the dependency function g.sub.h(x.sub.h.sup.k;θh) represents a dependency score for the MHC allele h indicating whether the MHC allele h will present the corresponding neoantigen based on at least the allele interacting features x.sub.h.sup.k, and in particular, based on positions of amino acids of the peptide sequence of peptide p.sup.k. For example, the dependency score for the MHC allele h may have a high value if the MHC allele h is likely to present the peptide p.sup.k, and may have a low value if presentation is not likely. The transformation function $f(\cdot)$ transforms the input, and more specifically, transforms the dependency score generated by g.sub.h(x.sub.h.sup.k;θh) in this case, to an appropriate value to indicate the likelihood that the peptide p.sup.k will be presented by an MHC allele.

[0626] In one particular implementation referred throughout the remainder of the specification, $f(\cdot)$ is a function having the range within [0, 1] for an appropriate domain range. In one example, $f(\cdot)$ is the expit function given by:

[00005] $f(z) = \frac{\exp(z)}{1 + \exp(z)}$. (4)

As another example, $f(\cdot)$ can also be the hyperbolic tangent function given by:

[00006] $f(z) = \tanh(z)$ (5)

when the values for the domain z is equal to or greater than 0. Alternatively, when predictions are made for the mass spectrometry ion current that have values outside the range [0, 1], $f(\cdot)$ can be any function such as the identity function, the exponential function, the log function, and the like.

[0627] Thus, the per-allele likelihood that a peptide sequence p.sup.k will be presented by a MHC allele h can be generated by applying the dependency function g.sub.h(·) for the MHC allele h to the encoded version of the peptide sequence p.sup.k to generate the corresponding dependency score. The dependency score may be transformed by the transformation function $f(\cdot)$ to generate a per-allele likelihood that the peptide sequence p.sup.k will be presented by the MHC allele h.

X.B.1 Dependency Functions for Allele Interacting Variables

[0628] In one particular implementation referred throughout the specification, the dependency function g.sub.h(·) is an affine function given by:

[00007] $g_h(x_h^i; \theta_h) = x_h^i$ .Math. θ_h . (6)

that linearly combines each allele-interacting variable in x.sub.h.sup.k with a corresponding parameter in the set of parameters θh determined for the associated MHC allele h.

[0629] In another particular implementation referred throughout the specification, the dependency function g.sub.h(·) is a network function given by:

[00008] $g_h(x_h^i; \theta_h) = NN_h(x_h^i; \theta_h)$. (7)

represented by a network model NN.sub.h(·) having a series of nodes arranged in one or more layers. A node may be connected to other nodes through connections each having an associated parameter in the set of parameters θh. A value at one particular node may be represented as a sum of the values of nodes connected to the particular node weighted by the associated parameter mapped by an activation function associated with the particular node. In contrast to the affine function, network models are advantageous because the presentation model can incorporate non-linearity and process data having different lengths of amino acid sequences. Specifically, through non-linear modeling, network models can capture interaction between amino acids at different positions in a peptide sequence and how this interaction affects peptide presentation.

[0630] In general, network models NN.sub.h(·) may be structured as feed-forward networks, such as artificial neural networks (ANN), convolutional neural networks (CNN), deep neural networks (DNN), and/or recurrent networks, such as long short-term memory networks (LSTM), bi-directional recurrent networks, deep bi-

directional recurrent networks, and the like.
[0631] In one instance referred throughout the remainder of the specification, each MHC allele in h=1, 2, . . . , m is associated with a separate network model, and NN.sub.h(·) denotes the output(s) from a network model associated with MHC allele h.
[0632] FIG. **5** illustrates an example network model NN.sub.3(·) in association with an arbitrary MHC allele h=3. As shown in FIG. **5**, the network model NN.sub.3(·) for MHC allele h=3 includes three input nodes at layer l=1, four nodes at layer l=2, two nodes at layer l=3, and one output node at layer l=4. The network model NN.sub.3(·) is associated with a set of ten parameters θ.sub.3 (1), θ.sub.3 (2), . . . , θ.sub.3 (10). The network model NN.sub.3(·) receives input values (individual data instances including encoded polypeptide sequence data and any other training data used) for three allele-interacting variables x.sub.3.sup.k(1), x.sub.3.sup.k(2), and x.sub.3.sup.k(3) for MHC allele h=3 and outputs the value NN.sub.3(x.sub.3.sup.k). The network function may also include one or more network models each taking different allele interacting variables as input.
[0633] In another instance, the identified MHC alleles h=1, 2, . . . , m are associated with a single network model NN.sub.H(·), and NN.sub.h(·) denotes one or more outputs of the single network model associated with MHC allele h. In such an instance, the set of parameters θh may correspond to a set of parameters for the single network model, and thus, the set of parameters θh may be shared by all MHC alleles.
[0634] FIG. **6**A illustrates an example network model NN.sub.H(·) shared by MHC alleles h=1, 2, . . . , m. As shown in FIG. **6**A, the network model NN.sub.H(·) includes m output nodes each corresponding to an MHC allele. The network model NN.sub.3(·) receives the allele-interacting variables x.sub.3.sup.k for MHC allele h=3 and outputs m values including the value NN.sub.3(x.sub.3.sup.k) corresponding to the MHC allele h=3.
[0635] In yet another instance, the single network model NN.sub.H(·) may be a network model that outputs a dependency score given the allele interacting variables x.sub.h.sup.k and the encoded protein sequence dh of an MHC allele h. In such an instance, the set of parameters θh may again correspond to a set of parameters for the single network model, and thus, the set of parameters θh may be shared by all MHC alleles. Thus, in such an instance, NN.sub.h(·) may denote the output of the single network model NN.sub.H(·) given inputs [x.sub.h.sup.kd.sub.h] to the single network model. Such a network model is advantageous because peptide presentation probabilities for MHC alleles that were unknown in the training data can be predicted just by identification of their protein sequence.
[0636] FIG. **6**B illustrates an example network model NN.sub.H(·) shared by MHC alleles. As shown in FIG. **6**B, the network model NN.sub.H(·) receives the allele interacting variables and protein sequence of MHC allele h=3 as input, and outputs a dependency score NN.sub.3(x.sub.3.sup.k) corresponding to the MHC allele h=3.
[0637] In yet another instance, the dependency function g.sub.h(·) can be expressed as:
[00009] $g_h(x_h^k;\theta_h) = g_h^{'}(x_h^k;\theta_h^{'}) + \theta_h^0$
where g′.sub.h(x.sub.h.sup.k;θ′.sub.h) is the affine function with a set of parameters θ′.sub.h, the network function, or the like, with a bias parameter θ.sub.h.sup.0 in the set of parameters for allele interacting variables for the MHC allele that represents a baseline probability of presentation for the MHC allele h.
[0638] In another implementation, the bias parameter θ.sub.h.sup.0 may be shared according to the gene family of the MHC allele h. That is, the bias parameter θ.sub.h.sup.0 for MHC allele h may be equal to θ.sub.gene(h).sup.0, where gene(h) is the gene family of MHC allele h. For example, class I MHC alleles HLA-A*02:01, HLA-A*02:02, and HLA-A*02:03 may be assigned to the gene family of "HLA-A," and the bias parameter θ.sub.h.sup.0 for each of these MHC alleles may be shared. As another example, class II MHC alleles HLA-DRB1:10:01, HLA-DRB1:11:01, and HLA-DRB3:01:01 may be assigned to the gene family of "HLA-DRB," and the bias parameter θ.sub.h.sup.0 for each of these MHC alleles may be shared.
[0639] Returning to equation (2), as an example, the likelihood that peptide p.sup.k will be presented by MHC allele h=3, among m=4 different identified MHC alleles using the affine dependency function g.sub.h(·), can be generated by:
[00010] $u_k^3 = f(x_3^k \text{ .Math. } \theta_3)$,
where x.sub.3.sup.k are the identified allele-interacting variables for MHC allele h=3, and θ.sub.3 are the set of parameters determined for MHC allele h=3 through loss function minimization.
[0640] As another example, the likelihood that peptide p.sup.k will be presented by MHC allele h=3, among m=4 different identified MHC alleles using separate network transformation functions g.sub.h(·), can be generated by:
[00011] $u_k^3 = f(NN_3(x_3^k;\theta_3))$,
where x.sub.3.sup.k are the identified allele-interacting variables for MHC allele h=3, and θ.sub.3 are the set of parameters determined for the network model NN.sub.3(·) associated with MHC allele h=3.
[0641] FIG. **7** illustrates generating a presentation likelihood for peptide p.sup.k in association with MHC allele h=3 using an example network model NN.sub.3(·). As shown in FIG. **7**, the network model NN.sub.3(·) receives

the allele-interacting variables x.sub.3.sup.k for MHC allele h=3 and generates the output NN.sub.3(x.sub.3.sup.k). The output is mapped by function $f(\cdot)$ to generate the estimated presentation likelihood u.sub.k.

X.B.2. Per-Allele with Allele-Noninteracting Variables

[0642] In one implementation, the training module **316** incorporates allele-noninteracting variables and models the estimated presentation likelihood u.sub.k for peptide p.sup.k by:

[00012] $$u_k^h = \Pr(p^k \text{ presented}) = f(g_w(w^k;\theta_w) + g_h(x_h^i;\theta_h)), \quad (8)$$

where w.sup.k denotes the encoded allele-noninteracting variables for peptide p.sup.k, g.sub.w(·) is a function for the allele-noninteracting variables w.sup.k based on a set of parameters θ.sub.w determined for the allele-noninteracting variables. Specifically, the values for the set of parameters θh for each MHC allele h and the set of parameters θ.sub.w for allele-noninteracting variables can be determined by minimizing the loss function with respect to θ.sub.h and θ.sub.w, where i is each instance in the subset S of training data **170** generated from cells expressing single MHC alleles.

[0643] The output of the dependency function g.sub.w(w.sup.k;θ.sub.w) represents a dependency score for the allele noninteracting variables indicating whether the peptide p.sup.k will be presented by one or more MHC alleles based on the impact of allele noninteracting variables. For example, the dependency score for the allele noninteracting variables may have a high value if the peptide p.sup.k is associated with a C-terminal flanking sequence that is known to positively impact presentation of the peptide p.sup.k, and may have a low value if the peptide p.sup.k is associated with a C-terminal flanking sequence that is known to negatively impact presentation of the peptide p.sup.k.

[0644] According to equation (8), the per-allele likelihood that a peptide sequence p.sup.k will be presented by a MHC allele h can be generated by applying the function g.sub.h(·) for the MHC allele h to the encoded version of the peptide sequence p.sup.k to generate the corresponding dependency score for allele interacting variables. The function g.sub.w(·) for the allele noninteracting variables are also applied to the encoded version of the allele noninteracting variables to generate the dependency score for the allele noninteracting variables. Both scores are combined, and the combined score is transformed by the transformation function $f(\cdot)$ to generate a per-allele likelihood that the peptide sequence p.sup.k will be presented by the MHC allele h.

[0645] Alternatively, the training module **316** may include allele-noninteracting variables w.sup.k in the prediction by adding the allele-noninteracting variables w.sup.k to the allele-interacting variables x.sub.h.sup.k in equation (2). Thus, the presentation likelihood can be given by:

[00013] $$u_k^h = \Pr(p^k \text{ presented; allele} h) = f(g_h([x_h^k\, w^k];\theta_h)). \quad (9)$$

X.B.3 Dependency Functions for Allele-Noninteracting Variables

[0646] Similarly to the dependency function g.sub.h(·) for allele-interacting variables, the dependency function g.sub.w(·) for allele noninteracting variables may be an affine function or a network function in which a separate network model is associated with allele-noninteracting variables w.sup.k.

[0647] Specifically, the dependency function g.sub.w(·) is an affine function given by:

[00014] $g_w(w^k;\theta_w) = w^k$ .Math. θ_w .

that linearly combines the allele-noninteracting variables in w.sup.k with a corresponding parameter in the set of parameters θ.sub.w.

[0648] The dependency function g.sub.w(·) may also be a network function given by:

[00015] $g_h(w^k;\theta_w) = \mathrm{NN}_w(w^k;\theta_w)$.

represented by a network model NN.sub.w(·) having an associated parameter in the set of parameters θ.sup.w. The network function may also include one or more network models each taking different allele noninteracting variables as input.

[0649] In another instance, the dependency function g.sub.w(·) for the allele-noninteracting variables can be given by:

[00016] $$g_w(w^k;\theta_w) = g'_w(w^k;\theta'_w) + h(m^k;\theta_w^m), \quad (10)$$

where g′.sub.w(w.sup.k;θ′.sub.w) is the affine function, the network function with the set of allele noninteracting parameters θ′.sub.w, or the like, m.sup.k is the mRNA quantification measurement for peptide p.sup.k, h(·) is a function transforming the quantification measurement, and θ.sub.w.sup.m is a parameter in the set of parameters for allele noninteracting variables that is combined with the mRNA quantification measurement to generate a dependency score for the mRNA quantification measurement. In one particular embodiment referred throughout the remainder of the specification, h(·) is the log function, however in practice h(·) may be any one of a variety of different functions.

[0650] In yet another instance, the dependency function g.sub.w(·) for the allele-noninteracting variables can be given by:

[00017] $g_w(w^k;\theta_w) = g'_w(w^k;\theta'_w) + \theta^o_w$ .Math. o^k, (11)

where g′.sub.w(w.sup.k;θ′.sub.w) is the affine function, the network function with the set of allele noninteracting parameters θ′.sub.w, or the like, o.sup.k is the indicator vector described above representing proteins and isoforms in the human proteome for peptide p.sup.k, and θ.sub.w.sup.o is a set of parameters in the set of parameters for allele noninteracting variables that is combined with the indicator vector. In one variation, when the dimensionality of o.sup.k and the set of parameters θ.sub.w.sup.o are significantly high, a parameter regularization term, such as λ.Math.‖θ.sub.w.sup.o‖, where ‖·‖ represents L1 norm, L2 norm, a combination, or the like, can be added to the loss function when determining the value of the parameters. The optimal value of the hyperparameter λ can be determined through appropriate methods.

[0651] In yet another instance, the dependency function g.sub.w(·) for the allele-noninteracting variables can be given by:

[00018] $g_w(w^k;\theta_w) = g'_w(w^k;\theta'_w) + \sum_{l=1}^{L}$ .Math. (gene($p^k = l$)) .Math. θ^l_w, (12)

where g′.sub.w(w.sup.k;θ′.sub.w) is the affine function, the network function with the set of allele noninteracting parameters θ′.sub.w, or the like, custom-character(gene(p.sup.k=l)) is the indicator function that equals to 1 if peptide p.sup.k is from source gene 1 as described above in reference to allele noninteracting variables, and θ.sub.w.sup.l is a parameter indicating "antigenicity" of source gene l. In one variation, when L is significantly high, and thus, the number of parameters θ.sub.w.sup.l=1, 2, . . . , L are significantly high, a parameter regularization term, such as λ.Math.‖θ.sub.w.sup.l‖, where ‖·‖ represents L1 norm, L2 norm, a combination, or the like, can be added to the loss function when determining the value of the parameters. The optimal value of the hyperparameter λ can be determined through appropriate methods.

[0652] In practice, the additional terms of any of equations (10), (11), and (12) may be combined to generate the dependency function g.sub.w(·) for allele noninteracting variables. For example, the term h(·) indicating mRNA quantification measurement in equation (10) and the term indicating source gene antigenicity in equation (12) may be summed together along with any other affine or network function to generate the dependency function for allele noninteracting variables.

[0653] Returning to equation (8), as an example, the likelihood that peptide p.sup.k will be presented by MHC allele h=3, among m=4 different identified MHC alleles using the affine transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00019] $u^3_k = f(w^k$ .Math. $\theta_w + x^k_3$ .Math. $\theta_3)$,

where w.sup.k are the identified allele-noninteracting variables for peptide p.sup.k, and θ.sub.w are the set of parameters determined for the allele-noninteracting variables.

[0654] As another example, the likelihood that peptide p.sup.k will be presented by MHC allele h=3, among m=4 different identified MHC alleles using the network transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00020] $u^3_k = f(\mathrm{NN}_w(w^k;\theta_w) + \mathrm{NN}_3(x^k_3;\theta_3))$

where w.sup.k are the identified allele-interacting variables for peptide p.sup.k, and θ.sub.w are the set of parameters determined for allele-noninteracting variables.

[0655] FIG. **8** illustrates generating a presentation likelihood for peptide p.sup.k in association with MHC allele h=3 using example network models NN.sub.3(·) and NN.sub.w(·). As shown in FIG. **8**, the network model NN.sub.3(·) receives the allele-interacting variables x.sub.3.sup.k for MHC allele h=3 and generates the output NN.sub.3(x.sub.3.sup.k). The network model NN.sub.w(·) receives the allele-noninteracting variables w.sup.k for peptide p.sup.k and generates the output NN.sub.w(w.sup.k). The outputs are combined and mapped by function $f(\cdot)$ to generate the estimated presentation likelihood u.sub.k.

X.C. Multiple-Allele Models

[0656] The training module **316** may also construct the presentation models to predict presentation likelihoods of peptides in a multiple-allele setting where two or more MHC alleles are present. In this case, the training module **316** may train the presentation models based on data instances S in the training data **170** generated from cells expressing single MHC alleles, cells expressing multiple MHC alleles, or a combination thereof.

X.C.1. Example 1: Maximum of Per-Allele Models

[0657] In one implementation, the training module **316** models the estimated presentation likelihood u.sub.k for peptide p.sup.k in association with a set of multiple MHC alleles H as a function of the presentation likelihoods u.sub.k.sup.h∈H determined for each of the MHC alleles h in the set H determined based on cells expressing single-alleles, as described above in conjunction with equations (2)-(11). Specifically, the presentation likelihood

u.sub.k can be any function of u.sub.k.sup.h∈H. In one implementation, as shown in equation (12), the function is the maximum function, and the presentation likelihood u.sub.k can be determined as the maximum of the presentation likelihoods for each MHC allele h in the set H.

[00021] u_k = Pr(p^k presented; alleles H) = max($u_k^{h \in H}$).

X.C.2. Example 2.1: Function-of-Sums Models

[0658] In one implementation, the training module **316** models the estimated presentation likelihood u.sub.k for peptide p.sup.k by:

[00022] u_k = Pr(p^k presented) = f(.Math. $_{h=1}^{m}$ a_h^k .Math. $g_h(x_h^k;\theta_h)$), (13)

where elements a.sub.h.sup.k are 1 for the multiple MHC alleles H associated with peptide sequence p.sup.k and x.sub.h.sup.k denotes the encoded allele-interacting variables for peptide p.sup.k and the corresponding MHC alleles. The values for the set of parameters θ.sub.h for each MHC allele h can be determined by minimizing the loss function with respect to θ.sub.h, where i is each instance in the subset S of training data **170** generated from cells expressing single MHC alleles and/or cells expressing multiple MHC alleles. The dependency function g.sub.h may be in the form of any of the dependency functions g.sub.h introduced above in sections X.B.1.

[0659] According to equation (13), the presentation likelihood that a peptide sequence p.sup.k will be presented by one or more MHC alleles h can be generated by applying the dependency function g.sub.h(·) to the encoded version of the peptide sequence p.sup.k for each of the MHC alleles H to generate the corresponding score for the allele interacting variables. The scores for each MHC allele h are combined, and transformed by the transformation function $f(\cdot)$ to generate the presentation likelihood that peptide sequence p.sup.k will be presented by the set of MHC alleles H.

[0660] The presentation model of equation (13) is different from the per-allele model of equation (2), in that the number of associated alleles for each peptide p.sup.k can be greater than 1. In other words, more than one element in a.sub.h.sup.k can have values of 1 for the multiple MHC alleles H associated with peptide sequence p.sub.k.

[0661] As an example, the likelihood that peptide p.sup.k will be presented by MHC alleles h=2, h=3, among m=4 different identified MHC alleles using the affine transformation functions g.sub.h(·), can be generated by:

[00023] $u_k = f(x_2^k$ .Math. $\theta_2 + x_3^k$ .Math. $\theta_3)$,

where x.sub.2.sup.k, x.sub.3.sup.k are the identified allele-interacting variables for MHC alleles h=2, h=3, and θ.sub.2, θ.sub.3 are the set of parameters determined for MHC alleles h=2, h=3.

[0662] As another example, the likelihood that peptide p.sup.k will be presented by MHC alleles h=2, h=3, among m=4 different identified MHC alleles using the network transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00024] $u_k = f(\mathrm{NN}_2(x_2^k;\theta_2) + \mathrm{NN}_3(x_3^k;\theta_3))$,

where NN.sub.2(·), NN.sub.3(·) are the identified network models for MHC alleles h=2, h=3, and θ.sub.2, θ.sub.3 are the set of parameters determined for MHC alleles h=2, h=3.

[0663] FIG. **9** illustrates generating a presentation likelihood for peptide p.sup.k in association with MHC alleles h=2, h=3 using example network models NN.sub.2(·) and NN.sub.3(·). As shown in FIG. **9**, the network model NN.sub.2(·) receives the allele-interacting variables x.sup.k for MHC allele h=2 and generates the output NN.sub.2(x.sup.k) and the network model NN.sub.3(·) receives the allele-interacting variables x3.sup.k for MHC allele h=3 and generates the output NN.sub.3(x.sub.3.sup.k). The outputs are combined and mapped by function $f(\cdot)$ to generate the estimated presentation likelihood u.sub.k.

X.C.3. Example 2.2: Function-of-Sums Models with Allele-Noninteracting Variables

[0664] In one implementation, the training module **316** incorporates allele-noninteracting variables and models the estimated presentation likelihood u.sub.k for peptide p.sup.k by:

[00025] u_k = Pr(p^k presented) = $f(g_w(w^k;\theta_w)$ + .Math. $_{h=1}^{m}$ a_h^k .Math. $g_h(x_h^k;\theta_h))$, (14)

where w.sup.k denotes the encoded allele-noninteracting variables for peptide p.sub.k. Specifically, the values for the set of parameters θ.sub.h for each MHC allele h and the set of parameters θ.sub.w for allele-noninteracting variables can be determined by minimizing the loss function with respect to θ.sub.h and θ.sub.w, where i is each instance in the subset S of training data **170** generated from cells expressing single MHC alleles and/or cells expressing multiple MHC alleles. The dependency function g.sub.w may be in the form of any of the dependency functions g.sub.w introduced above in sections X.B.3.

[0665] Thus, according to equation (14), the presentation likelihood that a peptide sequence P.sup.k will be presented by one or more MHC alleles H can be generated by applying the function g.sub.h(·) to the encoded version of the peptide sequence p.sup.k for each of the MHC alleles H to generate the corresponding dependency score for allele interacting variables for each MHC allele h. The function g.sub.w(·) for the allele noninteracting

variables is also applied to the encoded version of the allele noninteracting variables to generate the dependency score for the allele noninteracting variables. The scores are combined, and the combined score is transformed by the transformation function $f(\cdot)$ to generate the presentation likelihood that peptide sequence p.sup.k will be presented by the MHC alleles H.

[0666] In the presentation model of equation (14), the number of associated alleles for each peptide p.sup.k can be greater than 1. In other words, more than one element in a.sub.h.sup.k can have values of 1 for the multiple MHC alleles H associated with peptide sequence p.sup.k.

[0667] As an example, the likelihood that peptide p.sup.k will be presented by MHC alleles h=2, h=3, among m=4 different identified MHC alleles using the affine transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00026] $u_k = f(w^k$ .Math. $\theta_w + x_2^k$ .Math. $\theta_2 + x_3^k$ .Math. $\theta_3)$,

where w.sup.k are the identified allele-noninteracting variables for peptide p.sup.k, and θ.sub.w are the set of parameters determined for the allele-noninteracting variables.

[0668] As another example, the likelihood that peptide p.sup.k will be presented by MHC alleles h=2, h=3, among m=4 different identified MHC alleles using the network transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00027] $u_k = f(\mathrm{NN}_w(w^k;\theta_w) + \mathrm{NN}_2(x_2^k;\theta_2) + \mathrm{NN}_3(x_3^k;\theta_3))$

where w.sup.k are the identified allele-interacting variables for peptide p.sup.k, and θ.sub.w are the set of parameters determined for allele-noninteracting variables.

[0669] FIG. **10** illustrates generating a presentation likelihood for peptide p.sup.k in association with MHC alleles h=2, h=3 using example network models NN.sub.2(·), NN.sub.3(·), and NN.sub.w(·). As shown in FIG. **10**, the network model NN.sub.2(·) receives the allele-interacting variables x.sub.2.sup.k for MHC allele h=2 and generates the output NN.sub.2(x.sub.2.sup.k). The network model NN.sub.3(·) receives the allele-interacting variables x.sub.3.sup.k for MHC allele h=3 and generates the output NN.sub.3(x.sup.k). The network model NN.sub.w(·) receives the allele-noninteracting variables w.sup.k for peptide p.sup.k and generates the output NN.sub.w(w.sup.k). The outputs are combined and mapped by function $f(\cdot)$ to generate the estimated presentation likelihood u.sub.k.

[0670] Alternatively, the training module **316** may include allele-noninteracting variables w.sup.k in the prediction by adding the allele-noninteracting variables w.sup.k to the allele-interacting variables x.sub.h.sup.k in equation (15). Thus, the presentation likelihood can be given by:

[00028] $u_k = \Pr(p^k \text{ presented}) = f(\overset{m}{\underset{h=1}{.\mathrm{Math.}}} a_h^k$ .Math. $g_h([x_h^k w^k];\theta_h))$. (15)

X.C.4. Example 3.1: Models Using Implicit Per-Allele Likelihoods

[0671] In another implementation, the training module **316** models the estimated presentation likelihood u.sub.k for peptide p.sup.k by:

[00029] $u_k = \Pr(p^k \text{ presented}) = r(s(v = [a_1^k$ .Math. $u_k'^1(\theta)$ .Math. a_m^k .Math. $u_k'^m(\theta)]))$, (16)

where elements a.sub.h.sup.k are 1 for the multiple MHC alleles h∈H associated with peptide sequence p.sup.k, u′.sub.k.sup.h is an implicit per-allele presentation likelihood for MHC allele h, vector v is a vector in which element v.sub.h corresponds to a.sub.h.sup.k.Math.u′.sub.k.sup.h, s(·) is a function mapping the elements of v, and r(·) is a clipping function that clips the value of the input into a given range. As described below in more detail, s(·) may be the summation function or the second-order function, but it is appreciated that in other embodiments, s(·) can be any function such as the maximum function. The values for the set of parameters θ for the implicit per-allele likelihoods can be determined by minimizing the loss function with respect to θ, where i is each instance in the subset S of training data **170** generated from cells expressing single MHC alleles and/or cells expressing multiple MHC alleles.

[0672] The presentation likelihood in the presentation model of equation (17) is modeled as a function of implicit per-allele presentation likelihoods u′.sub.k.sup.h that each correspond to the likelihood peptide p.sup.k will be presented by an individual MHC allele h. The implicit per-allele likelihood is distinct from the per-allele presentation likelihood of section X.B in that the parameters for implicit per-allele likelihoods can be learned from multiple allele settings, in which direct association between a presented peptide and the corresponding MHC allele is unknown, in addition to single-allele settings. Thus, in a multiple-allele setting, the presentation model can estimate not only whether peptide p.sup.k will be presented by a set of MHC alleles H as a whole, but can also provide individual likelihoods u′.sub.k.sup.h∈H that indicate which MHC allele h most likely presented peptide p.sup.k. An advantage of this is that the presentation model can generate the implicit likelihoods without training data for cells expressing single MHC alleles.

[0673] In one particular implementation referred throughout the remainder of the specification, r(·) is a function

having the range [0, 1]. For example, r(·) may be the clip function:

[00030] $r(z) = \min(\max(z, 0), 1)$,

where the minimum value between z and 1 is chosen as the presentation likelihood u.sub.k. In another implementation, r(·) is the hyperbolic tangent function given by:

[00031] $r(z) = \tanh(z)$

when the values for the domain z is equal to or greater than 0.

X.C.5. Example 3.2: Sum-of-Functions Model

[0674] In one particular implementation, s(·) is a summation function, and the presentation likelihood is given by summing the implicit per-allele presentation likelihoods:

[00032] $u_k = \Pr(p^k \text{ presented}) = r(\sum_{h=1}^{m} a_h^k \cdot u_k'^h(\theta))$. (17)

[0675] In one implementation, the implicit per-allele presentation likelihood for MHC allele h is generated by:

[00033] $u_k'^h = f(g_h(x_h^k; \theta_h))$, (18)

such that the presentation likelihood is estimated by:

[00034] $u_k = \Pr(p^k \text{ presented}) = r(\sum_{h=1}^{m} a_h^k \cdot f(g_h(x_h^k; \theta_h)))$. (19)

[0676] According to equation (19), the presentation likelihood that a peptide sequence p.sup.k will be presented by one or more MHC alleles H can be generated by applying the function g.sub.h(·) to the encoded version of the peptide sequence p.sup.k for each of the MHC alleles H to generate the corresponding dependency score for allele interacting variables. Each dependency score is first transformed by the function $f(\cdot)$ to generate implicit per-allele presentation likelihoods u′.sub.k.sup.h. The per-allele likelihoods u′.sub.k.sup.h are combined, and the clipping function may be applied to the combined likelihoods to clip the values into a range [0, 1] to generate the presentation likelihood that peptide sequence p.sup.k will be presented by the set of MHC alleles H. The dependency function g.sub.h may be in the form of any of the dependency functions g.sub.h introduced above in sections X.B.1.

[0677] As an example, the likelihood that peptide p.sup.k will be presented by MHC alleles h=2, h=3, among m=4 different identified MHC alleles using the affine transformation functions g.sub.h(·), can be generated by:

[00035] $u_k = r(f(x_2^k \cdot \theta_2) + f(x_3^k \cdot \theta_3))$,

where x.sub.2.sup.k, x.sub.3.sup.k are the identified allele-interacting variables for MHC alleles h=2, h=3, and θ.sub.2, θ.sub.3 are the set of parameters determined for MHC alleles h=2, h=3.

[0678] As another example, the likelihood that peptide p.sup.k will be presented by MHC alleles h=2, h=3, among m=4 different identified MHC alleles using the network transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00036] $u_k = r(f(\mathrm{NN}_2(x_2^k; \theta_2)) + f(\mathrm{NN}_3(x_3^k; \theta_3)))$,

where NN.sub.2(·), NN.sub.3(·) are the identified network models for MHC alleles h=2, h=3, and θ.sub.2, θ.sub.3 are the set of parameters determined for MHC alleles h=2, h=3.

[0679] FIG. **11** illustrates generating a presentation likelihood for peptide p.sup.k in association with MHC alleles h=2, h=3 using example network models NN.sub.2(·) and NN.sub.3(·). As shown in FIG. **9**, the network model NN.sub.2(·) receives the allele-interacting variables x.sub.2.sup.k for MHC allele h=2 and generates the output NN.sub.2(x.sub.2.sup.k) and the network model NN.sub.3(·) receives the allele-interacting variables x.sub.3.sup.k for MHC allele h=3 and generates the output NN.sub.3(x.sub.3.sup.k). Each output is mapped by function $f(\cdot)$ and combined to generate the estimated presentation likelihood u.sub.k.

[0680] In another implementation, when the predictions are made for the log of mass spectrometry ion currents, r(·) is the log function and $f(\cdot)$ is the exponential function.

X.C.6. Example 3.3: Sum-of-Functions Models with Allele-Noninteracting Variables

[0681] In one implementation, the implicit per-allele presentation likelihood for MHC allele h is generated by:

[00037] $u_k'^h = f(g_h(x_h^k; \theta_h) + g_w(w^k; \theta_w))$, (20)

such that the presentation likelihood is generated by:

[00038] $u_k = \Pr(p^k \text{ presented}) = r(\sum_{h=1}^{m} a_h^k \cdot f(g_w(w^k; \theta_w) + g_h(x_h^k; \theta_h)))$, (21)

to incorporate the impact of allele noninteracting variables on peptide presentation.

[0682] According to equation (21), the presentation likelihood that a peptide sequence p.sup.k will be presented by one or more MHC alleles H can be generated by applying the function g.sub.h(·) to the encoded version of the peptide sequence p.sup.k for each of the MHC alleles H to generate the corresponding dependency score for allele interacting variables for each MHC allele h. The function g.sub.w(·) for the allele noninteracting variables is also

applied to the encoded version of the allele noninteracting variables to generate the dependency score for the allele noninteracting variables. The score for the allele noninteracting variables are combined to each of the dependency scores for the allele interacting variables. Each of the combined scores are transformed by the function $f(\cdot)$ to generate the implicit per-allele presentation likelihoods. The implicit likelihoods are combined, and the clipping function may be applied to the combined outputs to clip the values into a range [0,1] to generate the presentation likelihood that peptide sequence p.sup.k will be presented by the MHC alleles H. The dependency function g.sub.w may be in the form of any of the dependency functions g.sub.w introduced above in sections X.B.3.

[0683] As an example, the likelihood that peptide p.sup.k will be presented by MHC alleles h=2, h=3, among m=4 different identified MHC alleles using the affine transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00039] $u_k = r(f(w^k \text{ .Math. } \theta_w + x_2^k \text{ .Math. } \theta_2) + f(w^k \text{ .Math. } \theta_w + x_3^k \text{ .Math. } \theta_3))$,

where w.sup.k are the identified allele-noninteracting variables for peptide p.sup.k, and θ.sub.w are the set of parameters determined for the allele-noninteracting variables.

[0684] As another example, the likelihood that peptide p.sup.k will be presented by MHC alleles h=2, h=3, among m=4 different identified MHC alleles using the network transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00040] $u_k = r(f(\text{NN}_w(w^k;\theta_w) + \text{NN}_2(x_2^k;\theta_2)) + f(\text{NN}_w(w^k;\theta_w) + \text{NN}_3(x_3^k;\theta_3)))$

where w.sup.k are the identified allele-interacting variables for peptide p.sup.k, and θ.sub.w are the set of parameters determined for allele-noninteracting variables.

[0685] FIG. **12** illustrates generating a presentation likelihood for peptide p.sup.k in association with MHC alleles h=2, h=3 using example network models NN.sub.2(·), NN.sub.3(·), and NN.sub.w(·). As shown in FIG. **12**, the network model NN.sub.2(·) receives the allele-interacting variables x.sub.2.sup.k for MHC allele h=2 and generates the output NN.sub.2(x.sub.2.sup.k). The network model NN.sub.w(·) receives the allele-noninteracting variables w.sup.k for peptide p.sup.k and generates the output NN.sub.w(w.sup.k). The outputs are combined and mapped by function $f(\cdot)$. The network model NN.sub.3(·) receives the allele-interacting variables x.sub.3.sup.k for MHC allele h=3 and generates the output NN.sub.3(x.sub.3.sup.k), which is again combined with the output NN.sub.w(w.sup.k) of the same network model NN.sub.w(·) and mapped by function $f(\cdot)$. Both outputs are combined to generate the estimated presentation likelihood u.sub.k.

[0686] In another implementation, the implicit per-allele presentation likelihood for MHC allele h is generated by:

[00041] $u_k^{'h} = f(g_h([x_h^k w^k];\theta_h))$. (22)

such that the presentation likelihood is generated by:

[00042] $u_k = \Pr(p^k \text{ presented}) = r(\overset{m}{\underset{h=1}{\text{.Math.}}} a_h^k \text{ .Math. } f(g_h([x_h^k w^k];\theta_h)))$.

X.C.7. Example 4: Second Order Models

[0687] In one implementation, s(·) is a second-order function, and the estimated presentation likelihood u.sub.k for peptide p.sup.k is given by:

[00043]

$$u_k = \Pr(p^k \text{ presented}) = \overset{m}{\underset{h=1}{\text{.Math.}}} a_h^k \text{ .Math. } u_k^{'h}(\theta) - \overset{m}{\underset{h=1}{\text{.Math.}}} \underset{j<h}{\text{.Math.}} a_h^k \text{ .Math. } a_j^k \text{ .Math. } u_k^{'h}(\theta) \text{ .Math. } u_k^{'j}(\theta) \quad (23)$$

where elements u′.sub.k.sup.h are the implicit per-allele presentation likelihood for MHC allele h. The values for the set of parameters θ for the implicit per-allele likelihoods can be determined by minimizing the loss function with respect to θ, where i is each instance in the subset S of training data **170** generated from cells expressing single MHC alleles and/or cells expressing multiple MHC alleles. The implicit per-allele presentation likelihoods may be in any form shown in equations (18), (20), and (22) described above.

[0688] In one aspect, the model of equation (23) may imply that there exists a possibility peptide p.sup.k will be presented by two MHC alleles simultaneously, in which the presentation by two HLA alleles is statistically independent.

[0689] According to equation (23), the presentation likelihood that a peptide sequence p.sup.k will be presented by one or more MHC alleles H can be generated by combining the implicit per-allele presentation likelihoods and subtracting the likelihood that each pair of MHC alleles will simultaneously present the peptide p.sup.k from the summation to generate the presentation likelihood that peptide sequence p.sup.k will be presented by the MHC alleles H.

[0690] As an example, the likelihood that peptide p.sup.k will be presented by HLA alleles h=2, h=3, among m=4 different identified HLA alleles using the affine transformation functions g.sub.h(·), can be generated by:

[00044] $u_k = f(x_2^k \text{ .Math. } \theta_2) + f(x_3^k \text{ .Math. } \theta_3) - f(x_2^k \text{ .Math. } \theta_2) \text{ .Math. } f(x_3^k \text{ .Math. } \theta_3)$,

where x.sub.2.sup.k, x.sub.3.sup.k are the identified allele-interacting variables for HLA alleles h=2, h=3, and θ.sub.2, θ.sub.3 are the set of parameters determined for HLA alleles h=2, h=3.

[0691] As another example, the likelihood that peptide p.sup.k will be presented by HLA alleles h=2, h=3, among m=4 different identified HLA alleles using the network transformation functions g.sub.h(·), g.sub.w(·), can be generated by:

[00045] $u_k = f(NN_2(x_2^k;\theta_2)) + f(NN_3(x_3^k;\theta_3)) - f(NN_2(x_2^k;\theta_2))$ .Math. $f(NN_3(x_3^k;\theta_3))$,

where NN.sub.2(·), NN.sub.3(·) are the identified network models for HLA alleles h=2, h=3, and θ.sub.2, θ.sub.3 are the set of parameters determined for HLA alleles h=2, h=3.

XL.A Example 5: Prediction Module

[0692] The prediction module **320** receives sequence data and selects candidate neoantigens in the sequence data using the presentation models. Specifically, the sequence data may be DNA sequences, RNA sequences, and/or protein sequences extracted from tumor tissue cells of patients. The prediction module **320** processes the sequence data into a plurality of peptide sequences p.sup.k having 8-15 amino acids for MHC-I or 6-30 amino acids for MHC-II. For example, the prediction module **320** may process the given sequence "IEFROEIFJEF" (SEQ ID NO: 76) into three peptide sequences having 9 amino acids "IEFROEIFJ," (SEQ ID NO: 77), "EFROEIFJE," (SEQ ID NO: 78), and "FROEIFJEF." (SEQ ID NO: 79) In one embodiment, the prediction module **320** 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.

[0693] The presentation module **320** applies one or more of the presentation models to the processed peptide sequences to estimate presentation likelihoods of the peptide sequences. Specifically, the prediction module **320** may select one or more candidate neoantigen peptide sequences that are likely to be presented on tumor HLA molecules by applying the presentation models to the candidate neoantigens. In one implementation, the presentation module **320** selects candidate neoantigen sequences that have estimated presentation likelihoods above a predetermined threshold. In another implementation, the presentation model selects the N candidate neoantigen 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 neoantigens for a given patient can be injected into the patient to induce immune responses.

XI.B. Example 6: Cassette Design Module

XI.B.1 Overview

[0694] The cassette design module **324** generates a vaccine cassette sequence based on the v selected candidate peptides for injection into a patient. Specifically, for a set of selected peptides p.sup.k, k=1, 2, . . . , v for inclusion in a vaccine of capacity v, the cassette sequence is given by concatenation of a series of therapeutic epitope sequences p′.sup.k, k=1, 2, . . . , v that each include the sequence of a corresponding peptide p.sup.k. In one embodiment, the cassette design module **324** may concatenate the epitopes directly adjacent to one another. For example, a vaccine cassette C may be represented as:

[00046] $C = [\ p'^{t_1} \quad p'^{t_2} \quad$ .Math. $\quad p'^{t_v}\ ]$ (24)

where p′.sup.ti denotes the i-th epitope of the cassette. Thus, t.sub.i corresponds to an index k=1, 2, . . . , v for the selected peptide at the i-th position of the cassette. In another embodiment, the cassette design module **324** may concatenate the epitopes with one or more optional linker sequences in between adjacent epitopes. For example, a vaccine cassette C may be represented as:

[00047] $C = [\ p'^{t_1} \quad l_{(t_1,t_2)} \quad p'^{t_2} \quad l_{(t_2,t_3)} \quad$ .Math. $\quad l_{(t_{v-1},t_v)} \quad p'^{t_v}\ ]$ (25)

where l.sub.(ti,tj) denotes a linker sequence placed between the i-th epitope p′.sup.ti and the j=i+1-th epitope p′.sup.j=i+1 of the cassette. The cassette design module **324** determines which of the selected epitopes p′.sup.k, k=1, 2, . . . , v are arranged at the different positions of the cassette, as well as any linker sequences placed between the epitopes. A cassette sequence C can be loaded as a vaccine based on any of the methods described in the present specification.

[0695] In one embodiment, the set of therapeutic epitopes may be generated based on the selected peptides determined by the prediction module **320** 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.

[0696] In one embodiment, the therapeutic epitopes p′.sup.k may correspond to the selected peptides p.sup.k

themselves. In another embodiment, the therapeutic epitopes p′.sup.k may also include C- and/or N-terminal flanking sequences in addition to the selected peptides. For example, an epitope p′.sup.k included in the cassette may be represented as a sequence [n.sup.k p.sup.k c.sup.k] where c.sup.k is a C-terminal flanking sequence attached the C-terminus of the selected peptide p.sup.k, and n.sup.k is an N-terminal flanking sequence attached to the N-terminus of the selected peptide p.sup.k. In one instance referred throughout the remainder of the specification, the N- and C-terminal flanking sequences are the native N- and C-terminal flanking sequences of the therapeutic vaccine epitope in the context of its source protein. In one instance referred throughout the remainder of the specification, the therapeutic epitope p′.sup.k represents a fixed-length epitope. In another instance, the therapeutic epitope p′.sup.k 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 c.sup.k and the N-terminal flanking sequence n.sup.k can each have varying lengths of 2-5 residues, resulting in 16 possible choices for the epitope p′.sup.k.

[0697] In one embodiment, the cassette design module **324** generates 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. A junction epitope spanning epitopes p′.sup.ti and p′.sup.tj may include any epitope sequence that overlaps with both p′.sup.ti or p′.sup.tj that is different from the sequences of therapeutic epitopes p′.sup.ti and p′.sup.tj themselves. Specifically, each junction between epitope p′.sup.ti and an adjacent epitope p′.sup.tj of the cassette with or without an optional linker sequence l.sup.(ti,tj) may be associated with n.sub.(ti,tj) junction epitopes e.sub.n(ti,tj), n=1, 2, . . . , n.sub.(ti,tj). The junction epitopes may be sequences that at least partially overlap with both epitopes p′.sup.ti and p′.sup.tj, or may be sequences that at least partially overlap with linker sequences placed between the epitopes p′.sup.ti and p′.sup.tj. Junction epitopes may be presented by MHC class I, MHC class II, or both.

[0698] FIG. **38** shows two example cassette sequences, cassette 1 (C.sub.1) and cassette 2 (C.sub.2). Each cassette has a vaccine capacity of v=2, and includes therapeutic epitopes p′.sup.t1=p.sup.1=SINFEKL (SEQ ID NO: 80) and p′.sup.t2=p.sup.2=LLLLLVVVV (SEQ ID NO: 81), and a linker sequence l.sup.(t1,t2)=AAY in between the two epitopes. Specifically, the sequence of cassette C.sub.1 is given by [p.sup.1 l.sup.(t1,t2) p.sup.2], while the sequence of cassette C.sub.2 is given by [p.sup.2 l.sup.(t1,t2) p.sup.1]. Example junction epitopes e.sub.n.sup.(1,2) of cassette C.sub.1 may be sequences such as EKLAAYLLL (SEQ ID NO: 82), KLAAYLLLLL (SEQ ID NO: 83), and FEKLAAYL (SEQ ID NO: 84) that span across both epitopes p′.sup.1 and p′.sup.2 in the cassette, and may be sequences such as AAYLLLLL (SEQ ID NO: 85) and YLLLLLVVV (SEQ ID NO: 86) that span across the linker sequence and a single selected epitope in the cassette. Similarly, example junction epitopes e.sub.m.sup.(2,1) of cassette C2 may be sequences such as VVVVAAYSIN (SEQ ID NO: 87), VVVVAAY (SEQ ID NO: 88), and AYSINFEK (SEQ ID NO: 89). Although both cassettes involve the same set of sequences p.sup.1, l.sup.(c1,c2), and p.sup.2, the set of junction epitopes that are identified are different depending on the ordered sequence of the therapeutic epitopes within the cassette.

[0699] In one embodiment, the cassette design module **324** generates 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..sup.76

[0700] In one embodiment, the cassette design module **324** iterates through one or more candidate cassettes, and determines 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.

[0701] In one embodiment, the cassette design module **324** may determine a cassette sequence associated with the lowest junction epitope presentation score among the candidate cassette sequences. In one instance, the presentation score for a given cassette sequence C is determined based on a set of distance metrics d(e.sub.n.sup.(ti,tj), n=1, 2, . . . , n.sub.(ti,j))=d.sub.(ti,tj) each associated with a junction in the cassette C. Specifically, a distance metric d.sub.(ti,tj) specifies a likelihood that one or more of the junction epitopes spanning between the pair of adjacent therapeutic epitopes p′.sup.i and p′.sup.j will be presented. The junction epitope presentation score for cassette C can then be determined by applying a function (e.g., summation, statistical function) to the set of distance metrics for the cassette C. Mathematically, the presentation score is given by:

[00048] score = $h(d_{(t_1,t_2)}, d_{(t_2,t_3)}, \text{.Math.}, d_{(t_{v-1},t_v)})$ (26)

where h(·) is some function mapping the distance metrics of each junction to a score. In one particular instance referred throughout the remainder of the specification, the function h(·) is the summation across the distance metrics of the cassette.

[0702] The cassette design module **324** 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. In one particular embodiment referred throughout the remainder of the specification, the distance metric d(·) for a given junction may be given by the sum of the presentation likelihoods or the expected number presented junction epitopes as determined by the presentation models described in sections VII and VIII of the specification. However, it is appreciated that in other embodiments, the distance metric may be derived from other factors alone or in combination with the models like the one exemplified above, where these other factors may include deriving the distance metric from any one or more of (alone or in combination): HLA binding affinity or stability measurements or predictions for HLA class I or HLA class II, and a presentation or immunogenicity model trained on HLA mass spectrometry or T-cell epitope data, for HLA class I or HLA class II. In one embodiment, the distance metric may combine information about HLA class I and HLA class II presentation. For example, the distance metric could be the number of junction epitopes predicted to bind any of the patient's HLA class I or HLA class II alleles with binding affinity below a threshold. In another example, the distance metric could be the expected number of epitopes predicted to be presented by any of the patient's HLA class I or HLA class II alleles.

[0703] The cassette design module **324** 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 **324** 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 by setting the distance metric d.sub.(ti,tj) to a very large value (e.g., 100) for pairs of epitopes t.sub.i,t.sub.j where containing epitope t.sub.i to the N-terminus of epitope t.sub.j results in the formation of a junction self-epitope.

[0704] Returning to the example in FIG. **38** the cassette design module **324** determines (for example) a distance metric d.sub.(t1,t2)=d.sub.(1,2)=0.39 for the single junction (t.sub.1,t.sub.2) in cassette C.sub.1 given by the summation of presentation likelihoods of all possible junction epitopes e.sub.n.sup.(t1,t2)=e.sub.n.sup.(1,2) having lengths, for example, from 8 to 15 amino acids for MHC class I, or 6-30 amino acids for MHC class II. Since no other junctions are present in cassette C.sub.1, the junction epitope presentation score, which is a summation across the distance metrics for cassette C.sub.1, is also given by 0.39. The cassette design module **324** also determines a distance metric d.sub.(t1,t2)=d.sub.(2,1)=0.068 for the single junction in cassette C.sub.2 given by the summation of presentation likelihoods of all possible junction epitopes e.sub.n.sup.(t1,t2)=e.sub.n.sup.(2,1) having lengths from 8 to 15 for MHC class I, or 9-30 amino acids for MHC class II. In this example, the junction epitope presentation score for cassette C.sub.2 is also given by the distance metric of the single junction 0.068. The cassette design module **324** outputs the cassette sequence of C.sub.2 as the optimal cassette since the junction epitope presentation score is lower than the cassette sequence of C.sub.1.

[0705] In some cases, the cassette design module **324** can perform a brute force approach and iterates 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 v increases. For example, for a vaccine capacity of v=20 epitopes, the cassette design module **324** has to iterate through ˜10.sup.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 **324** 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, the cassette design module **324** 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.

[0706] In one embodiment, the cassette design module **324** generates 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 **324** 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 **324** may generate a subset of ~1 million candidate cassettes for a set of v=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.

[0707] In another embodiment, the cassette design module **324** determines 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.

[0708] The cassette design module **324** determines an improved cassette sequence by solving an asymmetric TSP, in which each node corresponds to a therapeutic epitope p′.sup.k. The distance from a node corresponding to epitope p′.sup.k to another node corresponding to epitope p′.sup.m is given by the junction epitope distance metric d.sub.(k,m), while the distance from the node corresponding to the epitope p′.sup.m to the node corresponding to epitope p′.sup.k is given by the distance metric d.sub.(m,k) that may be different from the distance metric d.sub.(k,m). By solving for an improved optimal cassette using an asymmetric TSP, the cassette design module **324** 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. Specifically, given the set of therapeutic epitopes k=1, 2, . . . , v, the cassette design module **324** determines the distance metrics d.sub.(k,m), km=1, 2, . . . , v for each possible ordered pair of therapeutic epitopes in the cassette. In other words, for a given pair k, m of epitopes, both the distance metric d.sub.(k,m) for concatenating therapeutic epitope p′.sup.m after epitope p′.sup.k and the distance metric d.sub.(m,k) for concatenating therapeutic epitope p′.sup.k after epitope p′.sup.m is determined, since these distance metrics may be different from each other.

[0709] In one embodiment, the cassette design module **324** solves the asymmetric TSP through an integer linear programming problem. Specifically, the cassette design module **324** generates a (v+1) x (v+1) path matrix P given by the following:

[00049] $$P = \begin{bmatrix} 0 & 0^{1\times v} \\ 0^{v\times 1} & D \end{bmatrix}. \quad (26)$$

The v×v matrix D is an asymmetric distance matrix, where each element D(k, m), k=1, 2, . . . , v; m=1, 2, . . . , v corresponds to the distance metric for a junction from epitope p′.sup.k to epitope p′-Rows k 2, . . . , v of P correspond to nodes of the original epitopes, while row 1 and column 1 corresponds to a "ghost node" that is at zero distance from all other nodes. The addition of the "ghost node" to the matrix encodes the notion that the vaccine cassette is linear rather than circular, so there is no junction between the first and last epitopes. In other words, the sequence is not circular, and the first epitope is not assumed to be concatenated after the last epitope in the sequence. Let x.sub.km denote a binary variable whose value is 1 if there is a directed path (i.e., an epitope-epitope junction in the cassette) where epitope p′.sup.k is concatenated to the N-terminus of epitope p′.sup.m and 0 otherwise. In addition, let E denote the set of all v therapeutic vaccine epitopes, and let S⊂E denote a subset of epitopes. For any such subset S, let out(S) denote the number of epitope-epitope junctions x.sub.km=1 where k is an epitope in S and m is an epitope in E\S. Given a known path matrix P, the cassette design module **324** finds a path matrix X that solves the following integer linear programming problem:

[00050] $$\min_{X} \ .\text{Math.}\ \sum_{k=1}^{v+1} \ .\text{Math.}\ \sum_{k\neq m,\, m=1}^{v+1} P_{km} \ .\text{Math.}\ x_{km} \quad (27)$$

in which P.sub.km denotes element P(k,m) of the path matrix P, subject to the following constraints:

[00051]
$$\sum_{k=1}^{v+1} \ .\text{Math.}\ x_{km} = 1, \qquad m = 1, 2, \ .\text{Math.}\ , v+1$$

$$\sum_{m=1}^{v+1} \ .\text{Math.}\ x_{km} = 1, \qquad k = 1, 2, \ .\text{Math.}\ , v+1$$

$$x_{kk} = 0, \qquad k = 1, 2, \ .\text{Math.}\ , v+1$$

$$\text{out}(S) \geq 1, \qquad S \subset E, 2 \leq \ .\text{Math.}\ S \ .\text{Math.}\ \leq \ .\text{Math.}\ V \ .\text{Math.}\ /2$$

The first two constraints guarantee that each epitope appears exactly once in the cassette. The last constraint

ensures that the cassette is connected. In other words, the cassette encoded by x is a connected linear protein sequence.

[0710] The solutions for x.sub.km, k,m=1, 2, . . . , v+1 in the integer linear programming problem of equation (27) indicates the closed sequence of nodes and ghost nodes that can be used to infer one or more sequences of therapeutic epitopes for the cassette that lower the presentation score of junction epitopes. Specifically, a value of x.sub.km=1 indicates that a "path" exists from node k to node m, or in other words, that therapeutic epitope p′.sup.m should be concatenated after therapeutic epitope p.sup.i in the improved cassette sequence. A solution of x.sub.km=0 indicates that no such path exists, or in other words, that therapeutic epitope p′.sup.m should not be concatenated after therapeutic epitope p.sup.i in the improved cassette sequence. Collectively, the values of x.sub.km in the integer programming problem of equation (27) represent a sequence of nodes and the ghost node, in which the path enters and exists each node exactly once. For example, the values of x.sub.ghost,1=1, x.sub.13=1, x.sub.32=1, and x.sub.2,ghost=1 (0 otherwise) may indicate a sequence ghost.fwdarw.1.fwdarw.3.fwdarw.2.fwdarw.ghost of nodes and ghost nodes.

[0711] Once the sequence has been solved for, the ghost nodes are deleted from the sequence to generate a refined sequence with only the original nodes corresponding to therapeutic epitopes in the cassette. The refined sequence indicates the order in which selected epitopes should be concatenated in the cassette to improve the presentation score. For example, continuing from the example in the previous paragraph, the ghost node may be deleted to generate a refined sequence 1.fwdarw.3.fwdarw.2. The refined sequence indicates one possible way to concatenate epitopes in the cassette, namely p.sup.1.fwdarw.p.sup.3.fwdarw.p.sup.2

[0712] In one embodiment, when the therapeutic epitopes p′.sup.k are variable-length epitopes, the cassette design module **324** determines candidate distance metrics corresponding to different lengths of the therapeutic epitopes p′.sup.k and p′.sup.m, and identifies the distance metric d.sub.(k,m) as the smallest candidate distance metric. For example, epitopes p′.sup.k=[n.sup.k p.sup.k c.sup.k] and p′.sup.m=[n.sup.m p.sup.mc.sup.m] may each include a corresponding N- and C-terminal flanking sequence that can vary from (in one embodiment) 2-5 amino acids. Thus, the junction between epitopes p′.sup.k and p′.sup.m is associated with 16 different sets of junction epitopes based on the 4 possible length values of n.sup.k and the 4 possible length values of c.sup.m that are placed in the junction. The cassette design module **324** may determine candidate distance metrics for each set of junction epitopes, and determine the distance metric d.sub.(k,m) as the smallest value. The cassette design module **324** can then construct the path matrix P and solve for the integer linear programming problem in equation (27) to determine the cassette sequence.

[0713] Compared to the random sampling approach, solving for the cassette sequence using the integer programming problem requires determination of v×(v−1) distance metrics each corresponding to a pair of therapeutic epitopes in the vaccine. 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.

XI.B.2. Comparison of Junction Epitope Presentation for Cassette Sequences Generated by Random Sampling Vs. Asymmetric TSP

[0714] Two cassette sequences including v=20 therapeutic epitopes were generated by random sampling 1,000,000 permutations (cassette sequence C.sub.1), and by solving the integer linear programming problem in equation (27) (cassette sequence C.sub.2). The distance metrics, and thus, the presentation score was determined based on the presentation model described in equation (14), in which f is the sigmoid function, x.sub.h.sup.i is the sequence of peptide p.sup.i, g.sub.h(·) is the neural network function, w includes the flanking sequence, the log transcripts per kilobase million (TPM) of peptide p.sup.i, the antigenicity of the protein of peptide p.sup.i, and the sample ID of origin of peptide p.sup.i, and g.sub.w(·) of the flanking sequence and the log TPM are neural network functions, respectively. Each of the neural network functions for g.sub.h(·) included one output node of a one-hidden-layer multilayer perceptron (MLP) with input dimensions 231 (11 residues×21 characters per residue, including pad characters), width 256, rectified linear unit (ReLU) activations in the hidden layer, linear activations in the output layer, and one output node per HLA allele in the training data set. The neural network function for the flanking sequence was a one hidden-layer MLP with input dimension 210 (5 residues of N-terminal flanking sequence+5 residues of C-terminal flanking sequence×21 characters per residue, including the pad characters), width 32, ReLU activations in the hidden layer and linear activation in the output layer. The neural network function for the RNA log TPM was a one hidden layer MLP with input dimension 1, width 16, ReLU activations in the hidden layer and linear activation in the output layer. The presentation models were constructed for HLA alleles HLA-A*02:04, HLA-A*02:07, HLA-B*40:01, HLA-B*40:02, HLA-C*16:02, and HLA-C*16:04. The presentation score indicating the expected number of presented junction epitopes of the two cassette sequences were compared. Results showed that the presentation score for the cassette sequence generated

by solving the equation of (27) was associated with a ˜4 fold improvement over the presentation score for the cassette sequence generated by random sampling.

[0715] Specifically, the v=20 epitopes were given by:

TABLE-US-00003 p′.sup.1 = (SEQ ID NO: 90) YNYSYWISIFAHTMWYNIWHVQWNK p′.sup.2 = (SEQ ID NO: 91) IEALPYVFLQDQFELRLLKGEQGNN p′.sup.3 = (SEQ ID NO: 92) DSEETNTNYLHYCHFHWTWAQQTTV p′.sup.4 = (SEQ ID NO: 93) GMLSQYELKDCSLGFSWNDPAKYLR p′.sup.5 = (SEQ ID NO: 94) VRIDKFLMYVWYSAPFSAYPLYQDA p′.sup.6 = (SEQ ID NO: 95) CVHIYNNYPRMLGIPFSVMVSGFAM p′.sup.7 = (SEQ ID NO: 96) FTFKGNIWIEMAGQFERTWNYPLSL p′.sup.8 = (SEQ ID NO: 97) ANDDTPDFRKCYIEDHSFRFSQTMN p′.sup.9 = (SEQ ID NO: 98) AAQYIACMVNRQMTIVYHLTRWGMK p′.sup.10 = (SEQ ID NO: 99) KYLKEFTQLLTFVDCYMWITFCGPD p′.sup.11 = (SEQ ID NO: 100) AMHYRTDIHGYWIEYRQVDNQMWNT p′.sup.12 = (SEQ ID NO: 101) THVNEHQLEAVYRFHOVHCRFPYEN p′.sup.13 = (SEQ ID NO: 102) QTFSECLFFHCLKVWNNVKYAKSLK p′.sup.14 = (SEQ ID NO: 103) SFSSWHYKESHIALLMSPKKNHNNT p′.sup.15 = (SEQ ID NO: 104) ILDGIMSRWEKVCTRQTRYSYCQCA p′.sup.16 = (SEQ ID NO: 105) YRAAQMSKWPNKYFDFPEFMAYMPI p′.sup.17 = (SEQ ID NO: 106) PRPGMPCQHHNTHGLNDRQAFDDFV p′.sup.18 = (SEQ ID NO: 107) HNIISDETEVWEQAPHITWVYMWCR p′.sup.19 = (SEQ ID NO: 108) AYSWPVVPMKWIPYRALCANHPPGT p′.sup.20 = (SEQ ID NO: 109) HVMPHVAMNICNWYEFLYRISHIGR.

In the first example, 1,000,000 different candidate cassette sequences were randomly generated with the 20 therapeutic epitopes. The presentation score was generated for each of the candidate cassette sequences. The candidate cassette sequence identified to have the lowest presentation score was:

TABLE-US-00004 (SEQ ID NO: 110) C.sub.1 =
THVNEHQLEAVYRFHQVHCRFPYENAMHYQMWNTYRAAQMSKWPN
KYFDFPEFMAYMPICVHIYNNYPRMLGIPFSVMVSGFAMAYSWPVVPMKW
IPYRALCANHPPGTANDDTPDFRKCYIEDHSFRFSQTMNIEALPYVFLQD
QFELRLLKGEQGNNDSEETNTNYLHYCHFHWTWAQQTTVILDGIMSRWEK
VCTRQTRYSYCQCAFTFKGNIWIEMAGQFERTWNYPLSLSFSSWHYKESH
IALLMSPKKNHNNTQTFSECLFFHCLKVWNNVKYAKSLKHVMPHVAMNIC
NWYEFLYRISHIGRHNIISDETEVWEQAPHITWVYMWCRVRIDKFLMYVW
YSAPFSAYPLYQDAKYLKEFTQLLTFVDCYMWITFCGPDAAQYIACMVNR
QMTIVYHLTRWGMKYNYSYWISIFAHTMWYNIWHVQWNKGMLSQYELKDC
SLGFSWNDPAKYLRPRPGMPCQHHNTHGLNDRQAFDDFV

with a presentation score of 6.1 expected number of presented junction epitopes. The median presentation score of the 1,000,000 random sequences was 18.3. The experiment shows that the expected number of presented junction epitopes can be significantly reduced by identifying a cassette sequence among randomly sampled cassettes.

[0716] In the second example, a cassette sequence C.sub.2 was identified by solving the integer linear programming problem in equation (27). Specifically, the distance metric of each potential junction between a pair of therapeutic epitopes was determined. The distance metrics were used to solve for the solution to the integer programming problem. The cassette sequence identified by this approach was:

TABLE-US-00005 (SEQ ID NO: 111) C.sub.2 =
IEALPYVFLQDQFELRLLKGEQGNNILDGIMSRWEKVCTRQTRYS
YCQCAHVMPHVAMNICNWYEFLYRISHIGRTHVNEHQLEAVYRFHQVHCR
FPYENFTFKGNIWIEMAGQFERTWNYPLSLAMHYQMWNTSFSSWHYKESH
IALLMSPKKNHNNTVRIDKFLMYVWYSAPFSAYPLYQDAQTFSECLFFHC
LKVWNNVKYAKSLKYRAAQMSKWPNKYFDFPEFMAYMPIAYSWPVVPMKW
IPYRALCANHPPGTCVHIYNNYPRMLGIPFSVMVSGFAMHNIISDETEVW
EQAPHITWVYMWCRAAQYIACMVNRQMTIVYHLTRWGMKYNYSYWISIFA
HTMWYNIWHVQWNKGMLSQYELKDCSLGFSWNDPAKYLRKYLKEFTQLLT
FVDCYMWITFCGPDANDDTPDFRKCYIEDHSFRFSQTMNDSEETNTNYLH
YCHFHWTWAQQTTVPRPGMPCQHHNTHGLNDRQAFDDFV

with a presentation score of 1.7. The presentation score of cassette sequence C.sub.2 showed a -4 fold improvement over the presentation score of cassette sequence C.sub.1, and a -11 fold improvement over the

median presentation score of the 1,000,000 randomly generated candidate cassettes. The run-time for generating cassette C.sub.1 was 20 seconds on a single thread of a 2.30 GHz Intel Xeon E5-2650 CPU. The run-time for generating cassette C.sub.2 was 1 second on a single thread of the same CPU. Thus in this example, the cassette sequence identified by solving the integer programming problem of equation (27) produces a -4-fold better solution at 20-fold reduced computational cost.

[0717] The results show that the integer programming problem can potentially provide a cassette sequence with a lower number of presented junction epitopes than one identified from random sampling, potentially with less computation resources.

XI.B.3. Comparison of Junction Epitope Presentation for Cassette Sequence Selection Generated by MHCflurry and the Presentation Model

[0718] In this example, cassette sequences including v=20 therapeutic epitopes were selected based off tumor/normal exome sequencing, tumor transcriptome sequencing and HLA typing of a lung cancer sample were generated by random sampling 1,000,000 permutations, and by solving the integer linear programming problem in equation (27). The distance metrics, and thus, the presentation score were determined based on the number of junction epitopes predicted by MHCflurry, an HLA-peptide binding affinity predictor, to bind the patient's HLAs with affinity below a variety of thresholds (e.g., 50-1000 nM, or higher, or lower). In this example, the 20 nonsynoymous somatic mutations chosen as therapeutic epitopes were selected from among the 98 somatic mutations identified in the tumor sample by ranking the mutations according to the presentation model in Section XI.B above. However, it is appreciated that in other embodiments, the therapeutic epitopes may be selected based on other criteria; such as those based stability, or combinations of criteria such as presentation score, affinity, and so on. In addition, it is appreciated that the criteria used for prioritizing therapuetic epitopes for inclusion in the vac cine need not be the same as the criteria used for determining the distance metric D(k, m) used in the cassette design module **324**.

[0719] The patient's HLA class I alleles were HLA-A*01:01, HLA-A*03:01, HLA-B*07:02, HLA-B*35:03, HLA-C*07:02, HLA-C*14:02.

[0720] Specifically in this example, the v=20 therapuetic epitopes were TABLE-US-00006 (SEQ ID NO: 112) SSTPYLYYGTSSVSYQFPMVPGGDR (SEQ ID NO: 113) EMAGKIDLLRDSYIFQLFWREAAEP (SEQ ID NO: 114) ALKQRTWQALAHKYNSQPSVSLRDF (SEQ ID NO: 115) VSSHSSQATKDSAVGLKYSASTPVR (SEQ ID NO: 116) KEAIDAWAPYLPEYIDHVISPGVTS (SEQ ID NO: 117) SPVITAPPSSPVFDTSDIRKEPMNI (SEQ ID NO: 118) PAEVAEQYSEKLVYMPHTFFIGDHA (SEQ ID NO: 119) MADLDKLNIHSIIQRLLEVRGS (SEQ ID NO: 120) AAAYNEKSGRITLLSLLFQKVFAQI (SEQ ID NO: 121) KIEEVRDAMENEIRTQLRRQAAAHT (SEQ ID NO: 122) DRGHYVLCDFGSTTNKFQNPQTEGV (SEQ ID NO: 123) QVDNRKAEAEEAIKRLSYISQKVSD (SEQ ID NO: 124) CLSDAGVRKMTAAVRVMKRGLENLT (SEQ ID NO: 125) LPPRSLPSDPFSQVPASPQSQSSSQ (SEQ ID NO: 126) ELVLEDLQDGDVKMGGSFRGAFSNS (SEQ ID NO: 127) VTMDGVREEDLASFSLRKRWESEPH (SEQ ID NO: 128) IVGVMFFERAFDEGADAIYDHINEG (SEQ ID NO: 129) TVTPTPTPTGTQSPTPTPITTTTTV (SEQ ID NO: 130) QEEMPPRPCGGHTSSSLPKSHLEPS (SEQ ID NO: 131) PNIQAVLLPKKTDSHHKAKGK

[0721] Results from this example in the table below compare the number of junction epitopes predicted by MHCflurry to bind the patient's HLAs with affinity below the value in the threshold column (where nM stands for nanoMolar) as found via three example methods. For the first method, the optimal cassette found via the traveling salesman problem (ATSP) formulation described above with is run-time. For the second method, the optimal cassette as determined by taking the best cassette found after 1 million random samples. For the third method, the median number of junction epitopes was found in the 1 million random samples.

TABLE-US-00007 ATSP Random Sampling Median Threshold # Binding # Binding # Binding (nM) Junction Epitopes Junction Epitopes Junction Epitopes 50 0 0 3 100 0 0 7 150 0 1 12 500 15 26 55 1000 68 91 131

[0722] The results of this example illustrate that any one of a number of criteria may be used to identify whether or not a given cassette design meets design requirements. Specifically, as demonstrated by prior examples, the selected cassette sequence out of many candidates may be specified by the cassette sequence having a lowest junction epitope presentation score, or at least such a score below an identified threshold. This example represents that another criteria, such as binding affinity, may be used to specify whether or not a given cassette design meets design requirements. For this criteria, a threshold binding affinity (e.g., 50-1000, or greater or lower) may be set specifying that the cassette design sequence should have fewer than some threshold number of junction epitopes above the threshold (e.g., 0), and any one of a number of methods may be used (e.g., methods one through three

illustrated in the table) can be used to identify if a given candidate cassette sequence meets those requirements. These example methods further illustrate that depending on the method used, the thresholds may need to be set differently. Other criteria may be envisioned, such as those based stability, or combinations of criteria such as presentation score, affinity, and so on.

[0723] In another example, the same cassettes were generated using the same HLA type and 20 therapeutic epitopes from earlier in this section (XI.C), but instead of using distance metrics based off binding affinity prediction, the distance metric for epitopes m, k was the number of peptides spanning the m to k junction predicted to be presented by the patient's HLA class I alleles with probability of presentation above a series of thresholds (between probability of 0.005 and 0.5, or higher, or lower), where the probabilities of presentation were determined by the presentation model in Section XI.B above. This example further illustrates the breadth of criteria that may be considered in identifying whether a given candidate cassette sequence meets design requirements for use in the vaccine.

TABLE-US-00008 ATSP # Random Sampling Median Threshold Junction # Junction # Junction (probability) Epitopes Epitopes Epitopes 0.005 58 79 118 0.01 39 59 93 0.05 7 33 47 0.1 5 14 35 0.2 1 8 25 0.5 0 2 14

[0724] The examples above have identified that the criteria for determining whether a candidate cassette sequence may vary by implementation. Each of these examples has illustrated that the count of the number of junction epitopes falling above or below the criteria may be a count used in determining whether the candidate cassette sequence meets that criteria. For example, if the criteria is number of epitopes meeting or exceeding a threshold binding affinity for HLA, whether the candidate cassette sequence has greater or fewer than that number may determine whether the candidate cassette sequence meets the criteria for use as the selected cassette for the vaccine. Similarly if the criteria is the number of junction epitopes exceeding a threshold presentation likelihood.

[0725] However, in other embodiments, calculations other than counting can be performed to determine whether a candidate cassette sequence meets the design criteria. For example, rather than the count of epitopes exceeding/falling below some threshold, it may instead be determined what proportion of junction epitopes exceed or fall below the threshold, for example whether the top X % of junction epitopes have a presentation likelihood above some threshold Y, or whether X % percent of junction epitopes have an HLA binding affinity less than or greater than Z nM. These are merely examples, generally the criteria may be based on any attribute of either individual junction epitopes, or statistics derived from aggregations of some or all of the junction epitopes. Here, X can generally be any number between 0 and 100% (e.g., 75% or less) and Y can be any value between 0 and 1, and Z can be any number suitable to the criteria in question. These values may be determined empirically, and depend on the models and criteria used, as well as the quality of the training data used.

[0726] As such, in certain aspects, junction epitopes with high probabilities of presentation can be removed; junction epitopes with low probabilities of presentation can be retained; junction epitopes that bind tightly, i.e., junction epitopes with binding affinity below 1000 nM or 500 nM or some other threshold can be removed; and/or junction epitopes that bind weakly, i.e., junction epitopes with binding affinity above 1000 nM or 500 nM or some other threshold can be retained.

[0727] Although the examples above have identified candidate sequences using an implementation of the presentation model described above, these principles apply equally to an implementation where the epitopes for arrangement in the cassette sequences are identified based on other types of models as well, such as those based on affinity, stability, and so on.

XII. Example 7: Experimentation Results Showing Example Presentation Model Performance

[0728] The validity of the various presentation models described above were tested on test data T that were subsets of training data **170** that were not used to train the presentation models or a separate dataset from the training data **170** that have similar variables and data structures as the training data **170**.

[0729] A relevant metric indicative of the performance of a presentation models is:

[00052]PositivePredictiveValue(PPV) = $P(y_{i \in T} = 1 \text{ .Math. } u_{i \in T} \geq t) = \frac{\text{.Math.}_{i \in T} \mathbb{1}(y_i = 1, u_i \geq t)}{\text{.Math.}_{i \in T} \mathbb{1}(u_i \geq t)}$

that indicates the ratio of the number of peptide instances that were correctly predicted to be presented on associated HLA alleles to the number of peptide instances that were predicted to be presented on the HLA alleles. In one implementation, a peptide p.sup.i in the test data T was predicted to be presented on one or more associated HLA alleles if the corresponding likelihood estimate u.sub.i is greater or equal to a given threshold value t.

Another relevant metric indicative of the performance of presentation models is:

[00053]Recall = $P(u_{i \in T} \geq t \text{ .Math. } y_{i \in T} = 1) = \frac{\text{.Math.}_{i \in T} \mathbb{1}(y_i = 1, u_i \geq t)}{\text{.Math.}_{i \in T} \mathbb{1}(u_i \geq t)}$

that indicates the ratio of the number of peptide instances that were correctly predicted to be presented on associated HLA alleles to the number of peptide instances that were known to be presented on the HLA alleles.

Another relevant metric indicative of the performance of presentation models is the area-under-curve (AUC) of

the receiver operating characteristic (ROC). The ROC plots the recall against the false positive rate (FPR), which is given by:

[00054]FPR = $P(u_{i \in T} \geq t \mid y_{i \in T} = 0) = \frac{\text{.Math.}_{i \in T}\ \mathbb{1}(y_i = 1, u_i \geq t)}{\text{.Math.}_{i \in T}\ \mathbb{1}(u_i \geq t)}$.

XII.A. Comparison of Presentation Model Performance on Mass Spectrometry Data Against State-of-the-Art Model

[0730] FIG. **13**A compares performance results of an example presentation model, as presented herein, and state-of-the-art models for predicting peptide presentation on multiple-allele mass spectrometry data. Results showed that the example presentation model performed significantly better at predicting peptide presentation than state-of-the-art models based on affinity and stability predictions.

[0731] Specifically, the example presentation model shown in FIG. **13**A as "MS" was the maximum of per-alleles presentation model shown in equation (12), using the affine dependency function g.sub.h(·) and the expit function $f(\cdot)$. The example presentation model was trained based on a subset of the single-allele HLA-A*02:01 mass spectrometry data from the IEDB data set (data set "D1") (data can be found at www.iedb.org/doc/mhc_ligand_full.zip) and a subset of the single-allele HLA-B*07:02 mass spectrometry from the IEDB data set (data set "D2") (data can be found at www.iedb.org/doc/mhc_ligand_full.zip). All peptides from source protein that contain presented peptides in the test set were eliminated from the training data such that the example presentation model could not simply memorize the sequences of presented antigens.

[0732] The model shown in FIG. **13**A as "Affinity" was a model similar to the current state-of-the-art model that predicts peptide presentation based on affinity predictions NETMHCpan. Implementation of NETMHCpan is provided in detail at www.cbs.dtu.dk/services/NetMHCpan/. The model shown in FIG. **13**A as "Stability" was a model similar to the current state-of-the-art model that predicts peptide presentation based on stability predictions NETMHCstab. Implementation of NETMHCstab is provided in detail at www.cbs.dtu.dk/services/NetMHCstab-1.0/. The test data that is a subset of the multiple-allele JY cell line HLA-A*02:01 and HLA-B*07:02 mass spectrometry data from the Bassani-Sternberg data set (data set "D3") (data can be found at www.ebi.ac.uk/pride/archive/projects/PXD000394). The error bars (as indicated in solid lines) show 95% confidence intervals.

[0733] As shown in the results of FIG. **13**A, the example presentation model trained on mass spectrometry data had a significantly higher PPV value at 10% recall rate relative to the state-of-the-art models that predict peptide presentation based on MHC binding affinity predictions or MHC binding stability predictions. Specifically, the example presentation model had approximately 14% higher PPV than the model based on affinity predictions, and had approximately 12% higher PPV than the model based on stability predictions.

[0734] These results demonstrate that the example presentation model had significantly better performance than the state-of-the-art models that predict peptide presentation based on MHC binding affinity or MHC binding stability predictions even though the example presentation model was not trained based on protein sequences that contained presented peptides.

XII.B. Comparison of Presentation Model Performance on T-Cell Epitope Data Against State-of-the-Art Models

[0735] FIG. **13**B compares performance results of another example presentation model, as presented herein, and state-of-the-art models for predicting peptide presentation on T-cell epitope data. T-cell epitope data contains peptide sequences that were presented by MHC alleles on the cell surface, and recognized by T-cells. Results showed that even though the example presentation model is trained based on mass spectrometry data, the example presentation model performed significantly better at predicting T-cell epitopes than state-of-the-art models based on affinity and stability predictions. In other words, the results of FIG. **13**B indicated that not only did the example presentation model perform better than state-of-the-art models at predicting peptide presentation on mass spectrometry test data, but the example presentation model also performed significantly better than state-of-the-art models at predicting epitopes that were actually recognized by T-cells. This is an indication that the variety of presentation models as presented herein can provide improved identification of antigens that are likely to induce immunogenic responses in the immune system.

[0736] Specifically, the example presentation model shown in FIG. **13**B as "MS" was the per-allele presentation model shown in equation (2), using the affine transformation function g.sub.h(·) and the expit function $f(\cdot)$ that was trained based on a subset of data set D1. All peptides from source protein that contain presented peptides in the test set were eliminated from the training data such that the presentation model could not simply memorize the sequences of presented antigens.

[0737] Each of the models were applied to the test data that is a subset of mass spectrometry data on HLA-A*02:01 T-cell epitope data (data set "D4") (data can be found at www.iedb.org/doc/tcell full v3.zip). The model shown in FIG. **13**B as "Affinity" was a model similar to the current state-of-the-art model that predicts peptide presentation based on affinity predictions NETMHCpan, and the model shown in FIG. **13**B as "Stability" was a model similar to the current state-of-the-art model that predicts peptide presentation based on stability predictions

NETMHCstab. The error bars (as indicated in solid lines) show 95% confidence intervals.

[0738] As shown in the results of FIG. **13**A, the per-allele presentation model trained on mass spectrometry data had a significantly higher PPV value at 10% recall rate than the state-of-the-art models that predict peptide presentation based on MHC binding affinity or MHC binding stability predictions even though the presentation model was not trained based on protein sequences that contained presented peptides. Specifically, the per-allele presentation model had approximately 9% higher PPV than the model based on affinity predictions, and had approximately 8% higher PPV than the model based on stability predictions.

[0739] These results demonstrated that the example presentation model trained on mass spectrometry data performed significantly better than state-of-the-art models on predicting epitopes that were recognized by T-cells.

XII.C. Comparison of Different Presentation Model Performances on Mass Spectrometry Data

[0740] FIG. **13**C compares performance results for an example function-of-sums model (equation (13)), an example sum-of-functions model (equation (19)), and an example second order model (equation (23)) for predicting peptide presentation on multiple-allele mass spectrometry data. Results showed that the sum-of-functions model and second order model performed better than the function-of-sums model. This is because the function-of-sums model implies that alleles in a multiple-allele setting can interfere with each other for peptide presentation, when in reality, the presentation of peptides are effectively independent.

[0741] Specifically, the example presentation model labeled as “sigmoid-of-sums” in FIG. **13**C was the function-of-sums model using a network dependency function g.sub.h(·), the identity function$f(\cdot)$, and the expit function r(·). The example model labeled as “sum-of-sigmoids” was the sum-of-functions model in equation (19) with a network dependency function g.sub.h(·), the expit function$f(\cdot)$, and the identity function r(·). The example model labeled as “hyperbolic tangent” was the sum-of-functions model in equation (19) with a network dependency function g.sub.h(·), the expit function$f(\cdot)$, and the hyperbolic tangent function r(·). The example model labeled as “second order” was the second order model in equation (23) using an implicit per-allele presentation likelihood form shown in equation (18) with a network dependency function g.sub.h(·) and the expit function $f(\cdot)$. Each model was trained based on a subset of data set D1, D2, and D3. The example presentation models were applied to a test data that is a random subset of data set D3 that did not overlap with the training data.

[0742] As shown in FIG. **13**C, the first column refers to the AUC of the ROC when each presentation model was applied to the test set, the second column refers to the value of the negative log likelihood loss, and the third column refers to the PPV at 10% recall rate. As shown in FIG. **13**C, the performance of presentation models “sum-of-sigmoids,” “hyperbolic tangent,” and “second order” were approximately tied at approximately 15-16% PPV at 10% recall, while the performance of the model “sigmoid-of-sums” was slightly lower at approximately 11%.

[0743] As discussed previously in section X.C.4., the results showed that the presentation models “sum-of-sigmoids,” “hyperbolic tangent,” and “second order” have high values of PPV compared to the “sigmoid-of-sums” model because the models correctly account for how peptides are presented independently by each MHC allele in a multiple-allele setting.

XII.D. Comparison of Presentation Model Performance with and without Training on Single-Allele Mass Spectrometry Data

[0744] FIG. **13**D compares performance results for two example presentation models that are trained with and without single-allele mass spectrometry data on predicting peptide presentation for multiple-allele mass spectrometry data. The results indicated that example presentation models that are trained without single-allele data achieve comparable performance to that of example presentation models trained with single-allele data.

[0745] The example model “with A2/B7 single-allele data” was the “sum-of-sigmoids” presentation model in equation (19) with a network dependency function g.sub.h(·), the expit function $f(\cdot)$, and the identity function r(·). The model was trained based on a subset of data set D3 and single-allele mass spectrometry data for a variety of MHC alleles from the IEDB database (data can be found at: www.iedb.org/doc/mhc_ligand_full.zip). The example model “without A2/B7 single-allele data” was the same model, but trained based on a subset of the multiple-allele D3 data set without single-allele mass spectrometry data for alleles HLA-A*02:01 and HLA-B*07:02, but with single-allele mass spectrometry data for other alleles. Within the multiple-allele training data, cell line HCC1937 expressed HLA-B*07:02 but not HLA-A*02:01, and cell line HCT116 expressed HLA-A*02:01 but not HLA-B*07:02. The example presentation models were applied to a test data that was a random subset of data set D3 and did not overlap with the training data.

[0746] As shown in FIG. **13**D, the predictions based on the implicit per-allele presentation likelihoods for MHC allele HLA-A*02:01 performed significantly better on single-allele test data for MHC allele HLA-A*02:01 rather than for MHC allele HLA-B*07:02. Similar results are shown for MHC allele HLA-B*07:02.

[0747] These results indicate that the implicit per-allele presentation likelihoods of the presentation model can correctly predict and distinguish binding motifs to individual MHC alleles, even though direct association

between the peptides and each individual MHC allele was not known in the training data.

XII.E. Comparison of Per-Allele Prediction Performance without Training on Single-Allele Mass Spectrometry Data

[0748] FIG. **13**E shows performance for the "without A2/B7 single-allele data" and "with A2/B7 single-allele data" example models shown in FIG. **13**D on single-allele mass spectrometry data for alleles HLA-A*02:01 and HLA-B*07:02 that were held out in the analysis shown in FIG. **13**D. Results indicate that even through the example presentation model is trained without single-allele mass spectrometry data for these two alleles, the model is able to learn binding motifs for each MHC allele.

[0749] The column "Correlation" refers to the correlation between the actual labels that indicate whether the peptide was presented on the corresponding allele in the test data, and the label for prediction. As shown in FIG. **13**E, "A2 model predicting B7" indicates the performance of the model when peptide presentation is predicted for single-allele HLA-B*07:02 data based on the implicit per-allele presentation likelihood estimate for MHC allele HLA-A*02:01. Similarly, "A2 model predicting A2" indicates the performance of the model when peptide presentation is predicted for single-allele HLA-A*02:01 based on the implicit per-allele presentation likelihood estimate for MHC allele HLA-A*02:01. "B7 model predicting B7" indicates the performance of the model when peptide presentation is predicted for single-allele HLA-B*07:02 data based on the implicit per-allele presentation likelihood estimate for MHC allele HLA-B*07:02. "B7 model predicting A2" indicates the performance of the model when peptide presentation is predicted for single-allele HLA-A*02:01 based on the implicit per-allele presentation likelihood estimate for MHC allele HLA-B*07:02.

[0750] As shown in FIG. **13**E, the predictive capacity of implicit per-allele likelihoods for an HLA allele is significantly higher for the intended allele, and significantly lower for the other HLA allele. Similarly to the results shown in FIG. **13**D, the example presentation models correctly learned to differentiate peptide presentation of individual alleles HLA-A*02:01 and HLA-B*07:02, even though direct association between peptide presentation and these alleles were not present in the multiple-allele training data.

XII.E. Frequently Occurring Anchor Residues in Per-Allele Predictions Match Known Canonical Anchor Motifs

[0751] FIG. **13**F shows the common anchor residues at positions 2 and 9 among nonamers predicted by the "without A2/B7 single-allele data" example model shown in FIG. **13**D. The peptides were predicted to be presented if the estimated likelihood was above 5%. Results show that most common anchor residues in the peptides identified for presentation on the MHC alleles HLA-A*02:01 and HLA-B*07:02 matched previously known anchor motifs for these MHC alleles. This indicates that the example presentation models correctly learned peptide binding based on particular positions of amino acids of the peptide sequences, as expected.

[0752] As shown in FIG. **13**F, amino acids L/M at position 2 and amino acids V/L at position 9 were known to be canonical anchor residue motifs (as shown in Table 4 of link.springer.com/article/10.1186/1745-7580-4-2) for HLA-A*02:01, and amino acid P at position 2 and amino acids L/V at position 9 were known to be canonical anchor residue motifs for HLA-B*07:02. The most common anchor residue motifs at positions 2 and 9 for peptides identified the model matched the known canonical anchor residue motifs for both HLA alleles.

XII.G. Comparison of Presentation Model Performances with and without Allele Noninteracting Variables

[0753] FIG. **13**G compares performance results between an example presentation model that incorporated C- and N-terminal flanking sequences as allele-interacting variables, and an example presentation model that incorporated C- and N-terminal flanking sequences as allele-noninteracting variables. Results showed that incorporating C- and N-terminal flanking sequences as allele noninteracting variables significantly improved model performance. More specifically, it is valuable to identify appropriate features for peptide presentation that are common across different MHC alleles, and model them such that statistical strength for these allele-noninteracting variables are shared across MHC alleles to improve presentation model performance.

[0754] The example "allele-interacting" model was the sum-of-functions model using the form of implicit per-allele presentation likelihoods in equation (22) that incorporated C- and N-terminal flanking sequences as allele-interacting variables, with a network dependency function g.sub.h(·) and the expit function $f(\cdot)$. The example "allele-noninteracting" model was the sum-of-functions model shown in equation (21) that incorporated C- and N-terminal flanking sequences as allele-noninteracting variables, with a network dependency function g.sub.h(·) and the expit function $f(\cdot)$. The allele-noninteracting variables were modeled through a separate network dependency function g.sub.w(·). Both models were trained on a subset of data set D3 and single-allele mass spectrometry data for a variety of MHC alleles from the IEDB database (data can be found at: www.iedb.org/doc/mhc_ligand_full.zip). Each of the presentation models was applied to a test data set that is a random subset of data set D3 that did not overlap with the training data.

[0755] As shown in FIG. **13**G, incorporating C- and N-terminal flanking sequences in the example presentation model as allele-noninteracting variables achieved an approximately 3% improvement in PPV value relative to modeling them as allele-interacting variables. This is because, in general, the "allele-noninteracting" example

presentation model was able to share statistical strength of allele-noninteracting variables across MHC alleles by modeling the effect with a separate network dependency function with very little addition in computing power.

XII.H. Dependency Between Presented Peptides and mRNA Quantification

[0756] FIG. **13**H shows the dependency between mRNA abundance and the frequency of peptides presented on a tumor cell as determined by mass-spectrometry. Results show that there is a strong dependency between mRNA expression and peptide presentation.

[0757] Specifically, the horizontal axis in FIG. **13**H indicates mRNA expression in terms of transcripts per million (TPM) quartiles. The vertical axis in FIG. **13**H indicates fraction of presented epitopes from genes in corresponding mRNA expression quartiles. Each solid line is a plot relating the two measurements from a tumor sample that is associated with corresponding mass spectrometry data and mRNA expression measurements. As shown in FIG. **13**H, there is a strong positive correlation between mRNA expression, and the fraction of peptides in the corresponding gene. Specifically, peptides from genes in the top quartile of RNA expression are more than 20 times likely to be presented than the bottom quartile. Moreover, essentially 0 peptides are presented from genes that are not detected through RNA.

[0758] The results indicate that the performance of the presentation model can be greatly improved by incorporating mRNA quantification measurements, as these measurements are strongly predictive of peptide presentation.

XII.I. Comparison of Presentation Model Performance with Incorporation of RNA Quantification Data

[0759] FIG. **13**I shows performance of two example presentation models, one of which is trained based on mass spectrometry tumor cell data, another of which incorporates mRNA quantification data and mass spectrometry tumor cell data. As expected from FIG. **13**H, results indicated that there is a significant improvement in performance by incorporating mRNA quantification measurements in the example presentation model, since the mRNA expression is a strong indicator of peptide presentation.

[0760] “MHCflurry+RNA filter” was a model similar to the current state-of-the-art model that predicts peptide presentation based on affinity predictions. It was implemented using MHCflurry along with a standard gene expression filter that removed all peptides from proteins with mRNA quantification measurements that were less than 3.2 FPKM. Implementation of MHCflurry is provided in detail at github.com/hammerlab/mhcflurry/, and at biorxiv.org/content/early/2016/05/22/054775. The “Example Model, no RNA” model was the “sum-of-sigmoids” example presentation model shown in equation (21) with the network dependency function g.sub.h(·), the network dependency function g.sub.w(·), and the expit function $f(\cdot)$. The “Example Model, no RNA” model incorporated C-terminal flanking sequences as allele-noninteracting variables through a network dependency function g.sub.w(·).

[0761] The “Example Model, with RNA” model was the “sum-of-sigmoids” presentation model shown in equation (19) with network dependency function g.sub.h(·), the network dependency function g.sub.w(·) in equation (10) incorporating mRNA quantification data through a log function, and the expit function $f(\cdot)$. The “Example Model, with RNA” model incorporated C-terminal flanking sequences as allele-noninteracting variables through the network dependency functions g.sub.w(·) and incorporated mRNA quantification measurements through the log function.

[0762] Each model was trained on a combination of the single-allele mass spectrometry data from the IEDB data set, 7 cell lines from the multiple-allele mass spectrometry data from the Bassani-Sternberg data set, and 20 mass spectrometry tumor samples. Each model was applied to a test set including 5,000 held-out proteins from 7 tumor samples that constituted 9,830 presented peptides from a total of 52,156,840 peptides.

[0763] As shown in the first two bar graphs of FIG. **13**I, the “Example Model, no RNA” model has a PPV value at 20% Recall of 21%, while that of the state-of-the-art model is approximately 3%, This indicates an initial performance improvement of 18% in PPV value, even without the incorporation of mRNA quantification measurements. As shown in the third bar graph of FIG. **13**I, the “Example Model, with RNA” model that incorporates mRNA quantification data into the presentation model shows a PPV value of approximately 30%, which is almost a 10% increase in performance compared to the example presentation model without mRNA quantification measurements.

[0764] Thus, results indicate that as expected from the findings in FIG. **13**H, mRNA expression is indeed a strong predictor of peptide prediction, that allows significant improvement in the performance of a presentation model with very little addition of computational complexity.

XII.J. Example of Parameters Determined for MHC Allele HLA-C*16:04

[0765] FIG. **13**J compares probability of peptide presentation for different peptide lengths between results generated by the “Example Model, with RNA” presentation model described in reference to FIG. **13**I, and predicted results by state-of-the-art models that do not account for peptide length when predicting peptide presentation. Results indicated that the “Example Model, with RNA” example presentation model from FIG. **13**I

captured variation in likelihoods across peptides of differing lengths.
[0766] The horizontal axis denoted samples of peptides with lengths 8, 9, 10, and 11. The vertical axis denoted the probability of peptide presentation conditioned on the lengths of the peptide. The plot "Truth (Blind Test Data)" showed the proportion of presented peptides according to the length of the peptide in a sample test data set. The presentation likelihood varied with the length of the peptide. For example, as shown in FIG. **13**J, a 10mer peptide with canonical HLA-A2 L/V anchor motifs was approximately 3 times less likely to be presented than a 9mer with the same anchor residues. The plot "Models Ignoring Length" indicated predicted measurements if state-of-the-art models that ignore peptide length were to be applied to the same test data set for presentation prediction. These models may be NetMHC versions before version 4.0, NetMHCpan versions before version 3.0, and MHCflurry, that do not take into account variation in peptide presentation according to peptide length. As shown in FIG. **13**J, the proportion of presented peptides would be constant across different values of peptide length, indicating that these models would fail to capture variation in peptide presentation according to length. The plot "Example Model, with RNA" indicated measurements generated from the "Example Model, with RNA" presentation model. As shown in FIG. **13**J, the measurements generated by the "Example Model, with RNA" model closely followed those shown in "Truth (Blind Test Data)" and correctly accounted for different degrees of peptide presentation for lengths 8, 9, 10, and 11.
[0767] Thus, the results showed that the example presentation models as presented herein generated improved predictions not only for 9mer peptides, but also for peptides of other lengths between 8-15, which account for up to 40% of the presented peptides in HLA class I alleles.
XII.K. Example of Parameters Determined for MHC Allele HLA-C*16:04
[0768] The following shows a set of parameters determined for a variation of the per-allele presentation model (equation (2)) for MHC allele HLA-C*16:04 denoted by h:
[00055] u_k = expit(relu(x_h^k .Math. $W_h^1 + b_h^1$) .Math. $W_h^2 + b_h^2$),
where relu(·) is the rectified linear unit (RELU) function, and W.sub.h.sup.1, b.sub.h.sup.1, W.sub.h.sup.2, and b.sub.h.sup.2 are the set of parameters θ determined for the model. The allele interacting variables x.sub.h.sup.k consist of peptide sequences. The dimensions of W.sub.h.sup.1 are (231×256), the dimensions of b.sub.h.sup.1(1×256), the dimensions of W.sub.h.sup.2 are (256×1), and b.sub.h.sup.2 is a scalar. For demonstration purposes, values for b.sub.h.sup.1, b.sub.h.sup.2, W.sub.h.sup.1, and W.sub.h.sup.2 are described in detail in PCT publication WO2017106638, herein incorporated by reference for all that it teaches.
XII.L. MHC II Example 1
[0769] Methods for determining MHC class II neoantigens are described in more detail in international application PCT/US2018/028438, herein incorporated by reference for all that it teaches.
[0770] FIG. **13**K is a histogram of lengths of peptides eluted from class II MHC alleles on human tumor cells and tumor infiltrating lymphocytes (TIL) using mass spectrometry. Specifically, mass spectrometry peptidomics was performed on HLA-DRB1*12:01 homozygote alleles ("Dataset 1") and HLA-DRB1*12:01, HLA-DRB1*10:01 multi-allele samples ("Dataset 2"). Results show that lengths of peptides eluted from class II MHC alleles range from 6-30 amino acids. The frequency distribution shown in FIG. **13**K is similar to that of lengths of peptides eluted from class II MHC alleles using state-of-the-art mass spectrometry techniques, as shown in FIG. **1**C of reference 91.
[0771] FIG. **13**L illustrates the dependency between mRNA quantification and presented peptides per residue for Dataset 1 and Dataset 2. Results show that there is a strong dependency between mRNA expression and peptide presentation for class II MHC alleles.
[0772] Specifically, the horizontal axis in FIG. **13**B indicates mRNA expression in terms of log.sub.10 transcripts per million (TPM) bins. The vertical axis in FIG. **13**L indicates peptide presentation per residue as a multiple of that of the lowest bin corresponding to mRNA expression between 10.sup.−2<log.sub.10TPM<10.sup.−1. One solid line is a plot relating mRNA quantification and peptide presentation for Dataset 1, and another is for Dataset 2. As shown in FIG. **13**L, there is a strong positive correlation between mRNA expression, and peptide presentation per residue in the corresponding gene. Specifically, peptides from genes in the range of 10.sup.1<log.sub.10TPM<10.sup.2 of RNA expression are more than 5 times likely to be presented than the bottom bin.
[0773] The results indicate that the performance of the presentation model can be greatly improved by incorporating mRNA quantification measurements, as these measurements are strongly predictive of peptide presentation.
[0774] FIG. **13**M compares performance results for example presentation models trained and tested using Dataset 1 and Dataset 2. For each set of model features of the example presentation models, FIG. **13**M depicts a PPV value at 10% recall when the features in the set of model features are classified as allele interacting features, and alternatively when the features in the set of model features are classified as allele non-interacting features

variables. As seen in FIG. **13**M, for each set of model features of the example presentation models, a PPV value at 10% recall that was identified when the features in the set of model features were classified as allele interacting features is shown on the left side, and a PPV value at 10% recall that was identified when the features in the set of model features were classified as allele non-interacting features is shown on the right side. Note that the feature of peptide sequence was always classified as an allele interacting feature for the purposes of FIG. **13**M. Results showed that the presentation models achieved a PPV value at 10% recall varying from 14% up to 29%, which are significantly (approximately 500-fold) higher than PPV for a random prediction.

[0775] Peptide sequences of lengths 9-20 were considered for this experiment. The data was split into training, validation, and testing sets. Blocks of peptides of 50 residue blocks from both Dataset 1 and Dataset 2 were assigned to training and testing sets. Peptides that were duplicated anywhere in the proteome were removed, ensuring that no peptide sequence appeared both in the training and testing set. The prevalence of peptide presentation in the training and testing set was increased by 50 times by removing non-presented peptides. This is because Dataset 1 and Dataset 2 are from human tumor samples in which only a fraction of the cells are class II HLA alleles, resulting in peptide yields that were roughly 10 times lower than in pure samples of class II HLA alleles, which is still an underestimate due to imperfect mass spectrometry sensitivity. The training set contained 1,064 presented and 3,810,070 non-presented peptides. The test set contained 314 presented and 807,400 non-presented peptides.

[0776] Example model 1 was the sum-of-functions model in equation (22) using a network dependency function g.sub.h(·), the expit function $f(\cdot)$, and the identity function r(·). The network dependency function g.sub.h(·) was structured as a multi-layer perceptron (MLP) with 256 hidden nodes and rectified linear unit (ReLU) activations. In addition to the peptide sequence, the allele interacting variables w contained the one-hot encoded C-terminal and N-terminal flanking sequence, a categorical variable indicating index of source gene G=gene(p.sup.i) of peptide p.sup.i, and a variable indicating mRNA quantification measurement. Example model 2 was identical to example model 1, except that the C-terminal and N-terminal flanking sequence was omitted from the allele interacting variables. Example model 3 was identical to example model 1, except that the index of source gene was omitted from the allele interacting variables. Example model 4 was identical to example model 1, except that the mRNA quantification measurement was omitted from the allele interacting variables.

[0777] Example model 5 was the sum-of-functions model in equation (20) with a network dependency function g.sub.h(·), the expit function $f(\cdot)$,the identity function r(·), and the dependency function g.sub.w(·) of equation (12). The dependency function g.sub.w(·) also included a network model taking mRNA quantification measurement as input, structured as a MLP with 16 hidden nodes and ReLU activations, and a network model taking C-flanking sequence as input, structured as a MLP with 32 hidden nodes and ReLU activations. The network dependency function g.sub.h(·) was structured as a multi-layer perceptron with 256 hidden nodes and rectified linear unit (ReLU) activations. Example model 6 was identical to example model 5, except that the network model for C-terminal and N-terminal flanking sequence was omitted. Example model 7 was identical to example model 5, except that the index of source gene was omitted from the allele noninteracting variables. Example model 8 was identical to example model 5, except that the network model for mRNA quantification measurement was omitted.

[0778] The prevalence of presented peptides in the test set was approximately 1/2400, and therefore, the PPV of a random prediction would also be approximately 1/2400=0.00042. As shown in FIG. **13**M, the best-performing presentation model achieved a PPV value of approximately 29%, which is roughly 500 times better than the PPV value of a random prediction.

XII.M. MHC II Example 2

[0779] FIG. **13**N is a histogram that depicts the quantity of peptides sequenced using mass spectrometry for each sample of a total of 39 samples comprising HLA class II molecules. Furthermore, for each sample of the plurality of samples, the histogram shown in FIG. **13**N depicts the quantity of peptides sequenced using mass spectrometry at different q-value thresholds. Specifically, for each sample of the plurality of samples, FIG. **13**N depicts the quantity of peptides sequenced using mass spectrometry with a q-value of less than 0.01, with a q-value of less than 0.05, and with a q-value of less than 0.2.

[0780] As noted above, each sample of the 39 samples of FIG. **13**N comprised HLA class II molecules. More specifically, each sample of the 39 samples of FIG. **13**N comprised HLA-DR molecules. The HLA-DR molecule is one type of HLA class II molecule. Even more specifically, each sample of the 39 samples of FIG. **13**N comprised HLA-DRB1 molecules, HLA-DRB3 molecules, HLA-DRB4 molecules, and/or HLA-DRB5 molecules. The HLA-DRB1 molecule, the HLA-DRB3 molecule, the HLA-DRB4 molecule, and the HLA-DRB5 molecule are types of the HLA-DR molecule.

[0781] While this particular experiment was performed using samples comprising HLA-DR molecules, and particularly HLA-DRB1 molecules, HLA-DRB3 molecules, HLA-DRB4 molecules, and HLA-DRB5 molecules,

in alternative embodiments, this experiment can be performed using samples comprising one or more of any type(s) of HLA class II molecules. For example, in alterative embodiments, identical experiments can be performed using samples comprising HLA-DP and/or HLA-DQ molecules. This ability to model any type(s) of MHC class II molecules using the same techniques, and still achieve reliable results, is well known by those skilled in the art. For instance, Jensen, Kamilla Kjaergaard, et al..sup.76 is one example of a recent scientific paper that uses identical methods for modeling binding affinity for HLA-DR molecules as well as for HLA-DQ and HLA-DP molecules. Therefore, one skilled in the art would understand that the experiments and models described herein can be used to separately or simultaneously model not only HLA-DR molecules, but any other MHC class II molecule, while still producing reliable results.

[0782] To sequence the peptides of each sample of the 39 total samples, mass spectrometry was performed for each sample. The resulting mass spectrum for the sample was then searched with Comet and scored with Percolator to sequence the peptides. Then, the quantity of peptides sequenced in the sample was identified for a plurality of different Percolator q-value thresholds. Specifically, for the sample, the quantity of peptides sequenced with a Percolator q-value of less than 0.01, with a Percolator q-value of less than 0.05, and with a Percolator q-value of less than 0.2 were determined.

[0783] For each sample of the 39 samples, the quantity of peptides sequenced at each of the different Percolator q-value thresholds is depicted in FIG. **13**N. For example, as seen in FIG. **13**N, for the first sample, approximately 4000 peptides with a q-value of less than 0.2 were sequenced using mass spectrometry, approximately 2800 peptides with a q-value of less than 0.05 were sequenced using mass spectrometry, and approximately 2300 peptides with a q-value of less than 0.01 were sequenced using mass spectrometry.

[0784] Overall, FIG. **13**N demonstrates the ability to use mass spectrometry to sequence a large quantity of peptides from samples containing MHC class II molecules, at low q-values. In other words, the data depicted in FIG. **13**N demonstrate the ability to reliably sequence peptides that may be presented by MHC class II molecules, using mass spectrometry.

[0785] FIG. **13**O is a histogram that depicts the quantity of samples in which a particular MHC class II molecule allele was identified. More specifically, for the 39 total samples comprising HLA class II molecules, FIG. **13**O depicts the quantity of samples in which certain MHC class II molecule alleles were identified.

[0786] As discussed above with regard to FIG. **13**N, each sample of the 39 samples of FIG. **13**N comprised HLA-DRB1 molecules, HLA-DRB3 molecules, HLA-DRB4 molecules, and/or HLA-DRB5 molecules. Therefore, FIG. **13**O depicts the quantity of samples in which certain alleles for HLA-DRB1, HLA-DRB3, HLA-DRB4, and HLA-DRB5 molecules were identified. To identify the HLA alleles present in a sample, HLA class II DR typing is performed for the sample. Then, to identify the quantity of samples in which a particular HLA allele was identified, the number of samples in which the HLA allele was identified using HLA class II DR typing is simply summed. For example, as depicted in FIG. **13**O, 19 samples of the 39 total samples contained the HLA class II molecule allele HLA-DRB4*01:03. In other words, 19 samples of the 39 total samples contained the allele HLA-DRB4*01:03 for the HLA-DRB4 molecule. Overall, FIG. **13**O depicts the ability to identify a wide range of HLA class II molecule alleles from the 39 samples comprising HLA class II molecules.

[0787] FIG. **13**P is a histogram that depicts the proportion of peptides presented by the MHC class II molecules in the 39 total samples, for each peptide length of a range of peptide lengths. To determine the length of each peptide in each sample of the 39 total samples, each peptide was sequenced using mass spectrometry as discussed above with regard to FIG. **13**N, and then the number of residues in the sequenced peptide was simply quantified.

[0788] As noted above, MHC class II molecules typically present peptides with lengths of between 9-20 amino acids. Accordingly, FIG. **13**P depicts the proportion of peptides presented by the MHC class II molecules in the 39 samples for each peptide length between 9-20 amino acids, inclusive. For example, as shown in FIG. **13**P, approximately 22% of the peptides presented by the MHC class II molecules in the 39 samples comprise a length of 14 amino acids.

[0789] Based on the data depicted in FIG. **13**P, modal lengths for the peptides presented by the MHC class II molecules in the 39 samples were identified to be 14 and 15 amino acids in length. These modal lengths identified for the peptides presented by the MHC class II molecules in the 39 samples are consistent with previous reports of modal lengths for peptides presented by MHC class II molecules. Additionally, as also consistent with previous reports, the data of FIG. **13**P indicates that more than 60% of the peptides presented by the MHC class II molecules from the 39 samples comprise lengths other than 14 and 15 amino acids. In other words, FIG. **13**P indicates that while peptides presented by MHC class II molecules are most frequently 14 or 15 amino acids in length, a large proportion of peptides presented by MHC class II molecules are not 14 or 15 amino acids in length. Accordingly, it is a poor assumption to assume that peptides of all lengths have equal probabilities of being presented by MHC class II molecules, or that only peptides that comprise a length of 14 or 15 amino acids are presented by MHC class II molecules. As discussed in detail below with regard to FIG. **13**T, these faulty

assumptions are currently used in many state-of-the-art models for predicting peptide presentation by MHC class II molecules, and therefore, the presentation likelihoods predicted by these models are often unreliable.
[0790] FIG. **13**Q is a line graph that depicts the relationship between gene expression and prevalence of presentation of the gene expression product by a MHC class II molecule, for genes present in the 39 samples. More specifically, FIG. **13**Q depicts the relationship between gene expression and the proportion of residues resulting from the gene expression that form the N-terminus of a peptide presented by a MHC class II molecule. To quantify gene expression in each sample of the 39 total samples, RNA sequencing is performed on the RNA included in each sample. In FIG. **13**Q, gene expression is measured by RNA sequencing in units of transcripts per million (TPM). To identify prevalence of presentation of gene expression products for each sample of the 39 samples, identification of HLA class II DR peptidomic data was performed for each sample.
[0791] As depicted in FIG. **13**Q, for the 39 samples, there is a strong correlation between gene expression level and presentation of residues of the expressed gene product by a MHC class II molecule. Specifically, as shown in FIG. **13**Q, peptides resulting from expression of the least-expressed genes are more than 100-fold less likely to be presented by a MHC class II molecule, than peptides resulting from expression of the most-expressed genes. In simpler terms, the products of more highly expressed genes are more frequently presented by MHC class II molecules.
[0792] FIGS. **13**H-J are line graphs that compare the performance of various presentation models at predicting the likelihood that peptides in a testing dataset of peptides will be presented by at least one of the MHC class II molecules present in the testing dataset. As shown in FIGS. **13**H-J, the performance of a model at predicting the likelihood that a peptide will be presented by at least one of the MHC class II molecules present in the testing dataset is determined by identifying a ratio of a true positive rate to a false positive rate for each prediction made by the model. These ratios identified for a given model can be visualized as a ROC (receiver operator characteristic) curve, in a line graph with an x-axis quantifying false positive rate and a y-axis quantifying true positive rate. An area under the curve (AUC) is used to quantify the performance of the model. Specifically, a model with a greater AUC has a higher performance (i.e., greater accuracy) relative to a model with a lesser AUC. In FIGS. **13**H-J, the blacked dashed line with a slope of 1 (i.e., a ratio of true positive rate to false positive rate of 1) depicts the expected curve for randomly guessing likelihoods of peptide presentation. The AUC for the dashed line is 0.5. ROC curves and the AUC metric are discussed in detail with regard to the top portion of Section XII. above.
[0793] FIG. **13**R is a line graph that compares the performance of five example presentation models at predicting the likelihood that peptides in a testing dataset of peptides will be presented by a MHC class II molecule, given different sets of allele interacting and allele non-interacting variables. In other words, FIG. **13**R quantifies the relative importance of various allele interacting and allele non-interacting variables for predicting the likelihood that a peptide will be presented by a MHC class II molecule.
[0794] The model architecture of each example presentation model of the five example presentations models used to generate the ROC curves of the line graph of FIG. **13**R, comprised an ensemble of five sum-of-sigmoids models. Each sum-of-sigmoids model in the ensemble was configured to model peptide presentation for up to four unique HLA-DR alleles per sample. Furthermore, each sum-of-sigmoids model in the ensemble was configured to make predictions of peptide presentation likelihood based on the following allele interacting and allele non-interacting variables: peptide sequence, flanking sequence, RNA expression in units of TPM, gene identifier, and sample identifier. The allele interacting component of each sum-of-sigmoids model in the ensemble was a one-hidden-layer MLP with ReLu activations as 256 hidden units.
[0795] Prior to using the example models to predict the likelihood that the peptides in a testing dataset of peptides will be presented by a MHC class II molecule, the example models were trained and validated. To train, validate, and finally test the example models, the data described above for the 39 samples was split into training, validation, and testing datasets.
[0796] To ensure that no peptides appeared in more than one of the training, validation, and testing datasets, the following procedure was performed. First all peptides from the 39 total samples that appeared in more than one location in the proteome were removed. Then, the peptides from the 39 total samples were partitioned into blocks of 10 adjacent peptides. Each block of the peptides from the 39 total samples was assigned uniquely to the training dataset, the validation dataset, or the testing dataset. In this way, no peptide appeared in more than one dataset of the training, validation, and testing datasets.
[0797] Out of the 28,081,944 peptides in the 39 total samples, the training dataset comprised 21,077 peptides presented by MHC class II molecules from 38 of the 39 total samples. The 21,077 peptides included in the training dataset were between lengths of 9 and 20 amino acids, inclusive. The example models used to generate the ROC curves in FIG. **13**R were trained on the training dataset using the ADAM optimizer and early stopping.
[0798] The validation dataset consisted of 2,346 peptides presented by MHC class II molecules from the same 38

samples used in the training dataset. The validation set was used only for early stopping.
[0799] The testing dataset comprised peptides presented by MHC class II molecules that were identified from a tumor sample using mass spectrometry. Specifically, the testing dataset comprised 203 peptides presented by MHC class II molecules-specifically HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB4*01:03, and HLA-DRB5*01:01 molecules—that were identified from the tumor sample. The peptides included in the testing dataset were held out of the training dataset described above.
[0800] As noted above, FIG. **13**R quantifies the relative importance of various allele interacting variables and allele non-interacting variables for predicting the likelihood that a peptide will be presented by a MHC class II molecule. As also noted above, the example models used to generate the ROC curves of the line graph of FIG. **13**R were configured to make predictions of peptide presentation likelihood based on the following allele interacting and allele non-interacting variables: peptide sequence, flanking sequence, RNA expression in units of TPM, gene identifier, and sample identifier. To quantify the relative importance of four of these five variables (peptide sequence, flanking sequence, RNA expression, and gene identifier) for predicting the likelihood that a peptide will be presented by a MHC class II molecule, each example model of the five the example models described above was tested using data from the testing dataset, with a different combination of the four variables. Specifically, for each peptide of the testing dataset, an example model 1 generated predictions of peptide presentation likelihood based on a peptide sequence, a flanking sequence, a gene identifier, and a sample identifier, but not on RNA expression. Similarly, for each peptide of the testing dataset, an example model 2 generated predictions of peptide presentation likelihood based on a peptide sequence, RNA expression, a gene identifier, and a sample identifier, but not on a flanking sequence. Similarly, for each peptide of the testing dataset, an example model 3 generated predictions of peptide presentation likelihood based on a flanking sequence, RNA expression, a gene identifier, and a sample identifier, but not on a peptide sequence. Similarly, for each peptide of the testing dataset, an example model 4 generated predictions of peptide presentation likelihood based on a flanking sequence, RNA expression, a peptide sequence, and a sample identifier, but not on a gene identifier. Finally, for each peptide of the testing dataset, an example model 5 generated predictions of peptide presentation likelihood based on all five variables of flanking sequence, RNA expression, peptide sequence, sample identifier, and gene identifier.
[0801] The performance of each of these five example models is depicted in the line graph of FIG. **13**R. Specifically, each of the five example models is associated with a ROC curve that depicts a ratio of a true positive rate to a false positive rate for each prediction made by the model. For instance, FIG. **13**R depicts a curve for the example model 1 that generated predictions of peptide presentation likelihood based on a peptide sequence, a flanking sequence, a gene identifier, and a sample identifier, but not on RNA expression. FIG. **13**R depicts a curve for the example model 2 that generated predictions of peptide presentation likelihood based on a peptide sequence, RNA expression, a gene identifier, and a sample identifier, but not on a flanking sequence. FIG. **13**R also depicts a curve for the example model 3 that generated predictions of peptide presentation likelihood based on a flanking sequence, RNA expression, a gene identifier, and a sample identifier, but not on a peptide sequence. FIG. **13**R also depicts a curve for the example model 4 that generated predictions of peptide presentation likelihood based on a flanking sequence, RNA expression, a peptide sequence, and a sample identifier, but not on a gene identifier. And finally FIG. **13**R depicts a curve for the example model 5 that generated predictions of peptide presentation likelihood based on all five variables of flanking sequence, RNA expression, peptide sequence, sample identifier, and gene identifier.
[0802] As noted above, the performance of a model at predicting the likelihood that a peptide will be presented by a MHC class II molecule is quantified by identifying an AUC for a ROC curve that depicts a ratio of a true positive rate to a false positive rate for each prediction made by the model. A model with a greater AUC has a higher performance (i.e., greater accuracy) relative to a model with a lesser AUC. As shown in FIG. **13**R, the curve for the example model 5 that generated predictions of peptide presentation likelihood based on all five variables of flanking sequence, RNA expression, peptide sequence, sample identifier, and gene identifier, achieved the highest AUC of 0.98. Therefore the example model 5 that used all five variables to generate predictions of peptide presentation achieved the best performance. The curve for the example model 2 that generated predictions of peptide presentation likelihood based on a peptide sequence, RNA expression, a gene identifier, and a sample identifier, but not on a flanking sequence, achieved the second highest AUC of 0.97. Therefore, the flanking sequence can be identified as the least important variable for predicting the likelihood that a peptide will be presented by a MHC class II molecule. The curve for the example model 4 generated predictions of peptide presentation likelihood based on a flanking sequence, RNA expression, a peptide sequence, and a sample identifier, but not on a gene identifier, achieved the third highest AUC of 0.96. Therefore, the gene identifier can be identified as the second least important variable for predicting the likelihood that a peptide will be presented by a MHC class II molecule. The curve for the example model 3 that generated predictions of

peptide presentation likelihood based on a flanking sequence, RNA expression, a gene identifier, and a sample identifier, but not on a peptide sequence, achieved the lowest AUC of 0.88. Therefore, the peptide sequence can be identified as the most important variable for predicting the likelihood that a peptide will be presented by a MHC class II molecule. The curve for the example model 1 that generated predictions of peptide presentation likelihood based on a peptide sequence, a flanking sequence, a gene identifier, and a sample identifier, but not on RNA expression, achieved the second lowest AUC of 0.95. Therefore, RNA expression can be identified as the second most important variable for predicting the likelihood that a peptide will be presented by a MHC class II molecule.

[0803] FIG. **13**S is a line graph that compares the performance of four different presentation models at predicting the likelihood that peptides in a testing dataset of peptides will be presented by a MHC class II molecule.

[0804] The first model tested in FIG. **13**S is referred to herein as a "full non-interacting model." The full non-interacting model is one embodiment of the presentation models described above in which allele-noninteracting variables w.sup.k and allele-interacting variables x.sub.h.sup.k are input into separate dependency functions such as, for example, a neural network, and then the outputs of these separate dependency functions are added. Specifically, the full non-interacting model is one embodiment of the presentation models described above in which allele-noninteracting variables w.sup.k are input into a dependency function g.sub.w, allele-interacting variables x.sub.h.sup.k are input into separate dependency function g.sub.h, and the outputs of the dependency function g.sub.w and the dependency function g.sub.h are added together. Therefore, in some embodiments, the full non-interacting model determines the likelihood of peptide presentation using equation 8 as shown above. Furthermore, embodiments of the full non-interacting model in which allele-noninteracting variables w.sup.k are input into a dependency function g.sub.w, allele-interacting variables x.sub.h.sup.k are input into separate dependency function g.sub.h, and the outputs of the dependency function g.sub.w and the dependency function g.sub.h are added, are discussed in detail above with regard to the top portion of Section X.B.2., the bottom portion of Section X.B.3., the top portion of Section X.C.3., and the top portion of Section X.C.6.

[0805] The second model tested in FIG. **13**S is referred to herein as a "full interacting model." The full interacting model is one embodiment of the presentation models described above in which allele-noninteracting variables w.sup.k are concatenated directly to allele-interacting variables x.sub.h.sup.k before being input into a dependency function such as, for example, a neural network. Therefore, in some embodiments, the full interacting model determines the likelihood of peptide presentation using equation 9 as shown above. Furthermore, embodiments of the full interacting model in which allele-noninteracting variables w.sup.k are concatenated with allele-interacting variables x.sub.h.sup.k before the variables are input into a dependency function are discussed in detail above with regard to the bottom portion of Section X.B.2., the bottom portion of Section X.C.2., and the bottom portion of Section X.C.5.

[0806] The third model tested in FIG. **13**S is referred to herein as a "CNN model." The CNN model comprises a convolutional neural network, and is similar to the full non-interacting model described above. However, the layers of the convolutional neural network of the CNN model differ from the layers of the neural network of the full non-interacting model. Specifically, the input layer of the convolutional neural network of the CNN model accepts a 20-mer peptide string and subsequently embeds the 20-mer peptide string as a (n, 20, 21) tensor. The next layers of the convolutional neural network of the CNN model comprise a 1-D convolutional kernel layer of size 5 with a stride of 1, a global max pooling layer, a dropout layer with p=0.2, and finally a dense 34-node layer with a ReLu activation.

[0807] The fourth and final model tested in FIG. **13**S is referred to herein as a "LSTM model." The LSTM model comprises a long short-term memory neural network. The input layer of the long short-term memory neural network of the LSTM model accepts a 20-mer peptide string and subsequently embeds the 20-mer peptide string as a (n, 20, 21) tensor. The next layers of the long short-term memory neural network of the LSTM model comprise a long short-term memory layer with 128 nodes, a dropout layer with p=0.2, and finally a dense 34-node layer with a ReLu activation.

[0808] Prior to using each of the four models of FIG. **13**S to predict the likelihood that the peptides in the testing dataset of peptides will be presented by a MHC class II molecule, the models were trained using the 38-sample training dataset described above and validated using the validation dataset described above. Following this training and validation of the models, each of the four models was tested using the held-out 39.sup.th sample testing dataset described above. Specifically, for each of the four models, each peptide of the testing dataset was input into the model, and the model subsequently output a presentation likelihood for the peptide.

[0809] The performance of each of the four models is depicted in the line graph in FIG. **13**S. Specifically, each of the four models is associated with a ROC curve that depicts a ratio of a true positive rate to a false positive rate for each prediction made by the model. For instance, FIG. **13**S depicts a ROC curve for the CNN model, a ROC curve for the full interacting model, a ROC curve for the LSTM model, and a ROC curve for the full non-

interacting model.
[0810] As noted above, the performance of a model at predicting the likelihood that a peptide will be presented by a MHC class II molecule is quantified by identifying an AUC for a ROC curve that depicts a ratio of a true positive rate to a false positive rate for each prediction made by the model. A model with a greater AUC has a higher performance (i.e., greater accuracy) relative to a model with a lesser AUC. As shown in FIG. **13**S, the curve for the full interacting model achieved the highest AUC of 0.982. Therefore the full interacting model achieved the best performance. The curve for the full non-interacting model achieved the second highest AUC of 0.977. Therefore, the full non-interacting model achieved the second best performance. The curve for the CNN model achieved the lowest AUC of 0.947. Therefore the CNN model achieved the worst performance. The curve for the LSTM model achieved the second lowest AUC of 0.952. Therefore, the LSTM model achieved the second worst performance. However, note that all models tested in FIG. **13**S have an AUC that is greater than 0.9. Accordingly, despite the architectural variance between them, all models tested in FIG. **13**S are capable of achieving relatively accurate predictions of peptide presentation.
[0811] FIG. **13**T is a line graph that compares the performance of two example best-in-class prior art models given two different criteria, and two example presentation models given two different sets of allele interacting and allele non-interacting variables, at predicting the likelihood that peptides in a testing dataset of peptides will be presented by a MHC class II molecule. Specifically, FIG. **13**T is a line graph that compares the performance of an example best-in-class prior art model that utilizes minimum NetMHCII 2.3 predicted binding affinity as a criterion to generate predictions (example model 1), an example best-in-class prior art model that utilizes minimum NetMHCII 2.3 predicted binding rank as a criterion to generate predictions (example model 2), an example presentation model that generates predictions of peptide presentation likelihood based on MHC class II molecule type and peptide sequence (example model 4), and an example presentation model that generates predictions of peptide presentation likelihood based on MHC class II molecule type, peptide sequence, RNA expression, gene identifier, and flanking sequence (example model 3).
[0812] The best-in-class prior art model used as example model 1 and example model 2 in FIG. **13**T is the NetMHCII 2.3 model. The NetMHCII 2.3 model generates predictions of peptide presentation likelihood based on MHC class II molecule type and peptide sequence. The NetMHCII 2.3 model was tested using the NetMHCII 2.3 website (www.cbs.dtu.dk/services/NetMHCII/, PMID 29315598).sup.76.
[0813] As noted above, the NetMHCII 2.3 model was tested according to two different criteria. Specifically, example model 1 model generated predictions of peptide presentation likelihood according to minimum NetMHCII 2.3 predicted binding affinity, and example model 2 generated predictions of peptide presentation likelihood according to minimum NetMHCII 2.3 predicted binding rank.
[0814] The presentation model used as example model 3 and example model 4 is an embodiment of the presentation model disclosed herein that is trained using data obtained via mass spectrometry. As noted above, the presentation model generated predictions of peptide presentation likelihood based on two different sets of allele interacting and allele non-interacting variables. Specifically, example model 4 generated predictions of peptide presentation likelihood based on MHC class II molecule type and peptide sequence (the same variable used by the NetMHCII 2.3 model), and example model 3 generated predictions of peptide presentation likelihood based on MHC class II molecule type, peptide sequence, RNA expression, gene identifier, and flanking sequence.
[0815] Prior using the example models of FIG. **13**T to predict the likelihood that the peptides in the testing dataset of peptides will be presented by a MHC class II molecule, the models were trained and validated. The NetMHCII 2.3 model (example model 1 and example model 2) was trained and validated using its own training and validation datasets based on HLA-peptide binding affinity assays deposited in the immune epitope database (IEDB, www.iedb.org). The training dataset used to train the NetMHCII 2.3 model is known to comprise almost exclusively 15-mer peptides. On the other hand, example models 3 and 4 were trained using the training dataset described above with regard to FIG. **13**R and validated and using the validation dataset described above with regard to FIG. **13**R.
[0816] Following the training and validation of the models, each of the models was tested using a testing dataset. As noted above, the NetMHCII 2.3 model is trained on a dataset comprising almost exclusively 15-mer peptides, meaning that NetMHCII 3.2 does not have the ability to give different priority to peptides of different weights, thereby reducing the predictive performance for NetMHCII 3.2 on HLA class II presentation mass spectrometry data containing peptides of all lengths. Therefore, to provide a fair comparison between the models not affected by variable peptide length, the testing dataset included exclusively 15-mer peptides. Specifically, the testing dataset comprised 933 15-mer peptides. 40 of the 933 peptides in the testing dataset were presented by MHC class II molecules-specifically by HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB4*01:03, and HLA-DRB5*01:01 molecules. The peptides included in the testing dataset were held out of the training datasets described above.
[0817] To test the example models using the testing dataset, for each of the example models, for each peptide of

the 933 peptides in the testing dataset, the model generated a prediction of presentation likelihood for the peptide. Specifically, for each peptide in the testing dataset, the example 1 model generated a presentation score for the peptide by the MHC class II molecules using MHC class II molecule types and peptide sequence, by ranking the peptide by the minimum NetMHCII 2.3 predicted binding affinity across the four HLA class II DR alleles in the testing dataset. Similarly, for each peptide in the testing dataset, the example 2 model generated a presentation score for the peptide by the MHC class II molecules using MHC class II molecule types and peptide sequence, by ranking the peptide by the minimum NetMHCII 2.3 predicted binding rank (i.e., quantile normalized binding affinity) across the four HLA class II DR alleles in the testing dataset. For each peptide in the testing dataset, the example 4 model generated a presentation likelihood for the peptide by the MHC class II molecules based on MHC class II molecule type and peptide sequence. Similarly, for each peptide in the testing dataset, the example model 3 generated a presentation likelihood for the peptide by the MHC class II molecules based on MHC class II molecule types, peptide sequence, RNA expression, gene identifier, and flanking sequence.

[0818] The performance of each of the four example models is depicted in the line graph in FIG. **13**T. Specifically, each of the four example models is associated with a ROC curve that depicts a ratio of a true positive rate to a false positive rate for each prediction made by the model. For instance, FIG. **13**T depicts a ROC curve for the example 1 model that utilized minimum NetMHCII 2.3 predicted binding affinity to generate predictions, a ROC curve for the example 2 model that utilized minimum NetMHCII 2.3 predicted binding rank to generate predictions, a ROC curve for the example 4 model that generated peptide presentation likelihoods based on MHC class II molecule type and peptide sequence, and a ROC curve for the example 3 model that generated peptide presentation likelihoods based on MHC class II molecule type, peptide sequence, RNA expression, gene identifier, and flanking sequence.

[0819] As noted above, the performance of a model at predicting the likelihood that a peptide will be presented by a MHC class II molecule is quantified by identifying an AUC for a ROC curve that depicts a ratio of a true positive rate to a false positive rate for each prediction made by the model. A model with a greater AUC has a higher performance (i.e., greater accuracy) relative to a model with a lesser AUC. As shown in FIG. **13**T, the curve for the example 3 model that generated peptide presentation likelihoods based on MHC class II molecule type, peptide sequence, RNA expression, gene identifier, and flanking sequence, achieved the highest AUC of 0.95. Therefore the example 3 model that generated peptide presentation likelihoods based on MHC class II molecule type, peptide sequence, RNA expression, gene identifier, and flanking sequence achieved the best performance. The curve for the example 4 model that generated peptide presentation likelihoods based on MHC class II molecule type and peptide sequence achieved the second highest AUC of 0.91. Therefore, the example 4 model that generated peptide presentation likelihoods based on MHC class II molecule type and peptide sequence achieved the second best performance. The curve for the example 1 model that utilized minimum NetMHCII 2.3 predicted binding affinity to generate predictions achieved the lowest AUC of 0.75. Therefore the example 1 model that utilized minimum NetMHCII 2.3 predicted binding affinity to generate predictions achieved the worst performance. The curve for the example 2 model that utilized minimum NetMHCII 2.3 predicted binding rank to generate predictions achieved the second lowest AUC of 0.76. Therefore, the example 2 model that utilized minimum NetMHCII 2.3 predicted binding rank to generate predictions achieved the second worst performance.

[0820] As shown in FIG. **13**T, the discrepancy in performance between the example models 1 and 2 and the example models 3 and 4 is large. Specifically, the performance of the NetMHCII 2.3 model (that utilizes either criterion of minimum NetMHCII 2.3 predicted binding affinity or minimum NetMHCII 2.3 predicted binding rank) is almost 25% lower than the performance of the presentation model disclosed herein (that generates peptide presentation likelihoods based on either MHC class II molecule type and peptide sequence, or on MHC class II molecule type, peptide sequence, RNA expression, gene identifier, and flanking sequence). Therefore, FIG. **13**T demonstrates that the presentation models disclosed herein are capable of achieving significantly more accurate presentation predictions than the current best-in-class prior art model, the NetMHCII 2.3 model.

[0821] Even further, as discussed above, the NetMHCII 2.3 model is trained on a training dataset that comprises almost exclusively 15-mer peptides. As a result, the NetMHCII 2.3 model is not trained to learn which peptides lengths are more likely to be presented by MHC class II molecules. Therefore, the NetMHCII 2.3 model does not weight its predictions of likelihood of peptide presentation by MHC class II molecules according to the length of the peptide. In other words, the NetMHCII 2.3 model does not modify its predictions of likelihood of peptide presentation by MHC class II molecules for peptides that have lengths outside of the modal peptide length of 15 amino acids. As a result, the NetMHCII 2.3 model overpredicts the likelihood of presentation of peptides with lengths greater or less than 15 amino acids.

[0822] On the other hand, the presentation models disclosed herein are trained using peptide data obtained via mass spectrometry, and therefore can be trained on training dataset that comprise peptides of all different lengths. As a result, the presentation models disclosed herein are able to learn which peptides lengths are more likely to be

presented by MHC class II molecules. Therefore, the presentation models disclosed herein can weight predictions of likelihood of peptide presentation by MHC class II molecules according to the length of the peptide. In other words, the presentation models disclosed herein are able to modify their predictions of likelihood of peptide presentation by MHC class II molecules for peptides that have lengths outside of the modal peptide length of 15 amino acids. As a result, the presentation models disclosed herein are capable of achieving significantly more accurate presentation predictions for peptides of lengths greater than or less than 15 amino acids, than the current best-in-class prior art model, the NetMHCII 2.3 model. This is one advantage of using the presentation models disclosed herein to predict likelihood of peptide presentation by MHC class II molecules.

XII.N. Example of Parameters Determined for MHC II Alleles

[0823] The following shows a set of parameters determined for a variation of the multi-allele presentation model (equation (16)) generating implicit per-allele presentation likelihoods for class II MHC alleles HLA-DRB1*12:01 and HLA-DRB1*10:01:

[00056] u = expit(relu(X .Math. $W^1 + b^1$) .Math. $W^2 + b^2$),

where relu(·) is the rectified linear unit (RELU) function, W.sup.1, b, W.sup.2, and b.sup.2 are the set of parameters θ determined for the model. The allele-interacting variables X are contained in a 1×399) matrix consisting of 1 row of one-hot encoded and middle-padded peptide sequences per input peptide. The dimensions of W.sup.1 are (399×256), the dimensions of b.sup.1 (1×256), the dimensions of W.sup.2 are (256×2), and b.sup.2 are (1×2). The first column of the output indicates the implicit per-allele probability of presentation for the peptide sequence by the allele HLA-DRB1*12:01, and the second column of the output indicates the implicit per-allele for the peptide sequence by the allele HLA-DRB1*10:01. For demonstration purposes, values for W.sup.1, b.sup.1, W.sup.2, and b.sup.2 are described in detail in international application PCT/US2018/028438, herein incorporated by reference for all that it teaches.

XIII. Example Computer

[0824] FIG. **14** illustrates an example computer **1400** for implementing the entities shown in FIGS. **1** and **3**. The computer **1400** includes at least one processor **1402** coupled to a chipset **1404**. The chipset **1404** includes a memory controller hub **1420** and an input/output (I/O) controller hub **1422**. A memory **1406** and a graphics adapter **1412** are coupled to the memory controller hub **1420**, and a display **1418** is coupled to the graphics adapter **1412**. A storage device **1408**, an input device **1414**, and network adapter **1416** are coupled to the I/O controller hub **1422**. Other embodiments of the computer **1400** have different architectures.

[0825] The storage device **1408** is a non-transitory computer-readable storage medium such as a hard drive, compact disk read-only memory (CD-ROM), DVD, or a solid-state memory device. The memory **1406** holds instructions and data used by the processor **1402**. The input interface **1414** is a touch-screen interface, a mouse, track ball, or other type of pointing device, a keyboard, or some combination thereof, and is used to input data into the computer **1400**. In some embodiments, the computer **1400** may be configured to receive input (e.g., commands) from the input interface **1414** via gestures from the user. The graphics adapter **1412** displays images and other information on the display **1418**. The network adapter **1416** couples the computer **1400** to one or more computer networks.

[0826] The computer **1400** is adapted to execute computer program modules for providing functionality described herein. As used herein, the term "module" refers to computer program logic used to provide the specified functionality. Thus, a module can be implemented in hardware, firmware, and/or software. In one embodiment, program modules are stored on the storage device **1408**, loaded into the memory **1406**, and executed by the processor **1402**.

[0827] The types of computers **1400** used by the entities of FIG. **1** can vary depending upon the embodiment and the processing power required by the entity. For example, the presentation identification system **160** can run in a single computer **1400** or multiple computers **1400** communicating with each other through a network such as in a server farm. The computers **1400** can lack some of the components described above, such as graphics adapters **1412**, and displays **1418**.

XIV. Neoantigen Delivery Vector Example

[0828] 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.

[0829] 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* 3.sup.rd Ed. (Plenum Press) Vols A and B(1992).

XIV.A. Neoantigen Cassette Design

[0830] Through vaccination, multiple class I MHC restricted tumor-specific neoantigens (TSNAs) that stimulate the corresponding cellular immune response(s) can be delivered. In one example, a vaccine cassette was engineered to encode multiple epitopes as a single gene product where the epitopes were either embedded within their natural, surrounding peptide sequence or spaced by non-natural linker sequences. Several design parameters were identified that could potentially impact antigen processing and presentation and therefore the magnitude and breadth of the TSNA specific CD8 T cell responses. In the present example, several model cassettes were designed and constructed to evaluate: (1) whether robust T cell responses could be generated to multiple epitopes incorporated in a single expression cassette; (2) what makes an optimal linker placed between the TSNAs within the expression cassette—that leads to optimal processing and presentation of all epitopes; (3) if the relative position of the epitopes within the cassette impact T cell responses; (4) whether the number of epitopes within a cassette influences the magnitude or quality of the T cell responses to individual epitopes; (5) if the addition of cellular targeting sequences improves T cell responses.

[0831] Two readouts were developed to evaluate antigen presentation and T cell responses specific for marker epitopes within the model cassettes: (1) an in vitro cell-based screen which allowed assessment of antigen presentation as gauged by the activation of specially engineered reporter T cells (Aarnoudse et al., 2002; Nagai et al., 2012); and (2) an in vivo assay that used HLA-A2 transgenic mice (Vitiello et al., 1991) to assess post-vaccination immunogenicity of cassette-derived epitopes of human origin by their corresponding epitope-specific T cell responses (Cornet et al., 2006; Depla et al., 2008; Ishioka et al., 1999).

XIV.B. Neoantigen Cassette Design Evaluation

XIV.B.1. Methods and Materials

TCR and Cassette Design and Cloning

[0832] The selected TCRs recognize peptides NLVPMVATV (SEQ ID NO: 132) (PDB #5D2N), CLGGLLTMV (SEQ ID NO: 133) (PDB #3REV), GILGFVFTL (SEQ ID NO: 134) (PDB #1OGA) LLFGYPVYV (SEQ ID NO: 135) (PDB #1A07) when presented by A*0201. Transfer vectors were constructed that contain 2A peptide-linked TCR subunits (beta followed by alpha), the EMCV IRES, and 2A-linked CD8 subunits (beta followed by alpha and by the puromycin resistance gene). Open reading frame sequences were codon-optimized and synthesized by GeneArt.

Cell Line Generation for In Vitro Epitope Processing and Presentation Studies

[0833] Peptides were purchased from ProImmune or Genscript diluted to 10 mg/mL with 10 mM tris(2-carboxyethyl)phosphine (TCEP) in water/DMSO (2:8, v/v). Cell culture medium and supplements, unless otherwise noted, were from Gibco. Heat inactivated fetal bovine serum (FBShi) was from Seradigm. QUANTI-Luc Substrate, Zeocin, and Puromycin were from InvivoGen. Jurkat-Lucia NFAT Cells (InvivoGen) were maintained in RPMI 1640 supplemented with 10% FBShi, Sodium Pyruvate, and 100 μg/mL Zeocin. Once transduced, these cells additionally received 0.3 μg/mL Puromycin. T2 cells (ATCC CRL-1992) were cultured in Iscove's Medium (IMDM) plus 20% FBShi. U-87 MG (ATCC HTB-14) cells were maintained in MEM Eagles Medium supplemented with 10% FBShi.

[0834] Jurkat-Lucia NFAT cells contain an NFAT-inducible Lucia reporter construct. The Lucia gene, when activated by the engagement of the T cell receptor (TCR), causes secretion of a coelenterazine-utilizing luciferase into the culture medium. This luciferase can be measured using the QUANTI-Luc luciferase detection reagent. Jurkat-Lucia cells were transduced with lentivirus to express antigen-specific TCRs. The HIV-derived lentivirus transfer vector was obtained from GeneCopoeia, and lentivirus support plasmids expressing VSV-G (pCMV-VsvG), Rev (pRSV-Rev) and Gag-pol (pCgpV) were obtained from Cell Design Labs.

[0835] Lentivirus was prepared by transfection of 50-80% confluent T75 flasks of HEK293 cells with Lipofectamine 2000 (Thermo Fisher), using 40 μl of lipofectamine and 20 μg of the DNA mixture (4:2:1:1 by weight of the transfer plasmid:pCgpV:pRSV-Rev:pCMV-VsvG). 8-10 mL of the virus-containing media were concentrated using the Lenti-X system (Clontech), and the virus resuspended in 100-200 μl of fresh medium. This volume was used to overlay an equal volume of Jurkat-Lucia cells (5×10E4-1×10E6 cells were used in different experiments). Following culture in 0.3 μg/ml puromycin-containing medium, cells were sorted to obtain clonality. These Jurkat-Lucia TCR clones were tested for activity and selectivity using peptide loaded T2 cells.

In Vitro Epitope Processing and Presentation Assay

[0836] T2 cells are routinely used to examine antigen recognition by TCRs. T2 cells lack a peptide transporter for antigen processing (TAP deficient) and cannot load endogenous peptides in the endoplasmic reticulum for presentation on the MHC. However, the T2 cells can easily be loaded with exogenous peptides. The five marker

peptides (NLVPMVATV (SEQ ID NO: 132), CLGGLLTMV (SEQ ID NO: 133), GLCTLVAML (SEQ ID NO: 136), LLFGYPVYV (SEQ ID NO: 135), GILGFVFTL (SEQ ID NO: 134)) and two irrelevant peptides (WLSLLVPFV (SEQ ID NO: 137), FLLTRICT (SEQ ID NO: 138)) were loaded onto T2 cells. Briefly, T2 cells were counted and diluted to 1×106 cells/mL with IMDM plus 1% FBShi. Peptides were added to result in 10 μg peptide/1×106 cells. Cells were then incubated at 37° C. for 90 minutes. Cells were washed twice with IMDM plus 20% FBShi, diluted to 5×10E5 cells/mL and 100 μL plated into a 96-well Costar tissue culture plate. Jurkat-Lucia TCR clones were counted and diluted to 5×10E5 cells/mL in RPMI 1640 plus 10% FBShi and 100 μL added to the T2 cells. Plates were incubated overnight at 37° C., 5% CO2. Plates were then centrifuged at 400g for 3 minutes and 20 μL supernatant removed to a white flat bottom Greiner plate. QUANTI-Luc substrate was prepared according to instructions and 50 μL/well added. Luciferase expression was read on a Molecular Devices SpectraMax iE3x.

[0837] To test marker epitope presentation by the adenoviral cassettes, U-87 MG cells were used as surrogate antigen presenting cells (APCs) and were transduced with the adenoviral vectors. U-87 MG cells were harvested and plated in culture media as 5×10E5 cells/100 μl in a 96-well Costar tissue culture plate. Plates were incubated for approximately 2 hours at 37° C. Adenoviral cassettes were diluted with MEM plus 10% FBShi to an MOI of 100, 50, 10, 5, 1 and 0 and added to the U-87 MG cells as 5 μl/well. Plates were again incubated for approximately 2 hours at 37° C. Jurkat-Lucia TCR clones were counted and diluted to 5×10E5 cells/mL in RPMI plus 10% FBShi and added to the U-87 MG cells as 100 μL/well. Plates were then incubated for approximately 24 hours at 37° C., 5% CO2. Plates were centrifuged at 400g for 3 minutes and 20 μL supernatant removed to a white flat bottom Greiner plate. QUANTI-Luc substrate was prepared according to instructions and 50 μL/well added. Luciferase expression was read on a Molecular Devices SpectraMax iE3x.

Mouse Strains for Immunogenicity Studies

[0838] Transgenic HLA-A2.1 (HLA-A2 Tg) mice were obtained from Taconic Labs, Inc. These mice carry a transgene consisting of a chimeric class I molecule comprised of the human HLA-A2.1 leader, α1, and α2 domains and the murine H2-Kb α3, transmembrane, and cytoplasmic domains (Vitiello et al., 1991). Mice used for these studies were the first generation offspring (F1) of wild type BALB/cAnNTac females and homozygous HLA-A2.1 Tg males on the C57Bl/6 background.

Adenovirus Vector (Ad5v) Immunizations

[0839] HLA-A2 Tg mice were immunized with 1×10.sup.10 to 1×10.sup.6 viral particles of adenoviral vectors via bilateral intramuscular injection into the tibialis anterior. Immune responses were measured at 12 days post-immunization.

Lymphocyte Isolation

[0840] Lymphocytes were isolated from freshly harvested spleens and lymph nodes of immunized mice. Tissues were dissociated in RPMI containing 10% fetal bovine serum with penicillin and streptomycin (complete RPMI) using the GentleMACS tissue dissociator according to the manufacturer's instructions.

Ex Vivo Enzyme-Linked Immunospot (ELISPOT) Analysis

[0841] ELISPOT analysis was performed according to ELISPOT harmonization guidelines (Janetzki et al., 2015) with the mouse IFNg ELISpotPLUS kit (MABTECH). 1×10.sup.5 splenocytes were incubated with 10 uM of the indicated peptides for 16 hours in 96-well IFNg antibody coated plates. Spots were developed using alkaline phosphatase. The reaction was timed for 10 minutes and was quenched by running the 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×(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.

Ex Vivo Intracellular Cytokine Staining (ICS) and Flow Cytometry Analysis

[0842] Freshly isolated lymphocytes at a density of 2-5×10.sup.6 cells/mL were incubated with 10 uM of the indicated peptides for 2 hours. After two hours, brefeldin A was added to a concentration of 5 ug/ml and cells were incubated with stimulant for an additional 4 hours. Following stimulation, viable cells were labeled with fixable viability dye eFluor780 according to manufacturer's protocol and stained with anti-CD8 APC (clone 53-6.7, BioLegend) at 1:400 dilution. Anti-IFNg PE (clone XMG1.2, BioLegend) was used at 1:100 for intracellular staining. Samples were collected on an Attune NxT Flow Cytometer (Thermo Scientific). Flow cytometry data was plotted and analyzed using FlowJo. To assess degree of antigen-specific response, both the percent IFNg+ of CD8+ cells and the total IFNg+ cell number/1×10.sup.6 live cells were calculated in response to each peptide stimulant.

XIV.B.2. In Vitro Evaluation of Neoantigen Cassette Designs

[0843] As an example of neoantigen cassette design evaluation, an in vitro cell-based assay was developed to assess whether selected human epitopes within model vaccine cassettes were being expressed, processed, and presented by antigen-presenting cells (FIG. **15**). Upon recognition, Jurkat-Lucia reporter T cells that were engineered to express one of five TCRs specific for well-characterized peptide-HLA combinations become activated and translocate the nuclear factor of activated T cells (NFAT) into the nucleus which leads to transcriptional activation of a luciferase reporter gene. Antigenic stimulation of the individual reporter CD8 T cell lines was quantified by bioluminescence.

[0844] Individual Jurkat-Lucia reporter lines were modified by lentiviral transduction with an expression construct that includes an antigen-specific TCR beta and TCR alpha chain separated by a P2A ribosomal skip sequence to ensure equimolar amounts of translated product (Banu et al., 2014). The addition of a second CD8 beta-P2A-CD8 alpha element to the lentiviral construct provided expression of the CD8 co-receptor, which the parent reporter cell line lacks, as CD8 on the cell surface is crucial for the binding affinity to target pMHC molecules and enhances signaling through engagement of its cytoplasmic tail (Lyons et al., 2006; Yachi et al., 2006).

[0845] After lentiviral transduction, the Jurkat-Lucia reporters were expanded under puromycin selection, subjected to single cell fluorescence assisted cell sorting (FACS), and the monoclonal populations tested for luciferase expression. This yielded stably transduced reporter cell lines for specific peptide antigens 1, 2, 4, and 5 with functional cell responses. (Table 2).

TABLE-US-00009 TABLE 2 Development of an in vitro T cell activation assay. Peptide-specific T cell recognition as measured by induction of luciferase indicates effective processing and presentation of the vaccine cassette antigens. Short Cassette Design Epitope AAY 1 24.5 ± 0.5 2 11.3 ± 0.4 3* n/a 4 26.1 ± 3.1 5 46.3 ± 1.9 *Reporter T cell for epitope 3 not yet generated

[0846] In another example, a series of short cassettes, all marker epitopes were incorporated in the same position (FIG. **16**A) and only the linkers separating the HLA-A*0201 restricted epitopes (FIG. **16**B) were varied. Reporter T cells were individually mixed with U-87 antigen-presenting cells (APCs) that were infected with adenoviral constructs expressing these short cassettes, and luciferase expression was measured relative to uninfected controls. All four antigens in the model cassettes were recognized by matching reporter T cells, demonstrating efficient processing and presentation of multiple antigens. The magnitude of T cell responses follow largely similar trends for the natural and AAY-linkers. The antigens released from the RR-linker based cassette show lower luciferase inductions (Table 3). The DPP-linker, designed to disrupt antigen processing, produced a vaccine cassette that led to low epitope presentation (Table 3).

TABLE-US-00010 TABLE 3 Evaluation of linker sequences in short cassettes. Luciferase induction in the in vitro T cell activation assay indicated that, apart from the DPP-based cassette, all linkers facilitated efficient release of the cassette antigens. T cell epitope only (no linker) = 9AA, natural linker one side = 17AA, natural linker both sides = 25AA, non-natural linkers = AAY, RR, DPP Short Cassette Designs Epitope 9AA 17AA 25AA AAY RR DPP 1 33.6 ± 0.9 42.8 ± 2.1 42.3 ± 2.3 24.5 ± 0.5 21.7 ± 0.9 0.9 ± 0.1 2 12.0 ± 0.9 10.3 ± 0.6 14.6 ± 04 11.3 ± 0.4 8.5 ± 0.3 1.1 ± 0.2 3* n/a n/a n/a n/a n/a n/a 4 26.6 ± 2.5 16.1 ± 0.6 16.6 ± 0.8 26.1 ± 3.1 12.5 ± 0.8 1.3 ± 0.2 5 29.7 ± 0.6 21.2 ± 0.7 24.3 ± 1.4 46.3 ± 1.9 19.7 ± 0.4 1.3 ± 0.1 *Reporter T cell for epitope 3 not yet generated

[0847] In another example, an additional series of short cassettes were constructed that, besides human and mouse epitopes, contained targeting sequences such as ubiquitin (Ub), MHC and Ig-kappa signal peptides (SP), and/or MHC transmembrane (TM) motifs positioned on either the N- or C-terminus of the cassette. (FIG. **17**). When delivered to U-87 APCs by adenoviral vector, the reporter T cells again demonstrated efficient processing and presentation of multiple cassette-derived antigens. However, the magnitude of T cell responses were not substantially impacted by the various targeting features (Table 4).

TABLE-US-00011 TABLE 4 Evaluation of cellular targeting sequences added to model vaccine cassettes. Employing the in vitro T cell activation assay demonstrated that the four HLA-A*0201 restricted marker epitopes are liberated efficiently from the model cassettes and targeting sequences did not substantially improve T cell recognition and activation. Short Cassette Designs Epitope A B C D E F G H I J 1 32.5 ± 1.5 31.8 ± 0.8 29.1 ± 1.2 29.1 ± 1.1 28.4 ± 0.7 20.4 ± 0.5 35.0 ± 1.3 30.3 ± 2.0 22.5 ± 0.9 38.1 ± 1.6 2 6.1 ± 0.2 6.3 ± 0.2 7.6 ± 0.4 7.0 ± 0.5 5.9 ± 0.2 3.7 ± 0.2 7.6 ± 0.4 5.4 ± 0.3 6.2 ± 0.4 6.4 ± 0.3 3* n/a n/a n/a n/a n/a n/a n/a n/a n/a n/a 4 12.3 ± 1.1 14.1 ± 0.7 12.2 ± 0.8 13.7 ± 1.0 11.7 ± 0.8 10.6 ± 0.4 11.0 ± 0.6 7.6 ± 0.6 16.1 ± 0.5 8.7 ± 0.5 5 44.4 ± 2.8 53.6 ± 1.6 49.9 ± 3.3 50.5 ± 2.8 41.7 ± 2.8 36.1 ± 1.1 46.5 ± 2.1 31.4 ± 0.6 75.4 ± 1.6 35.7 ± 2.2 *Reporter T cell for epitope 3 not yet generated

XIV.B.3. In Vivo Evaluation of Neoantigen Cassette Designs

[0848] As another example of neoantigen cassette design evaluation, vaccine cassettes were designed to contain 5 well-characterized human class I MHC epitopes known to stimulate CD8 T cells in an HLA-A*02:01 restricted

fashion (FIG. **16**A, **17**, **19**A). For the evaluation of their in vivo immunogenicity, vaccine cassettes containing these marker epitopes were incorporated in adenoviral vectors and used to infect HLA-A2 transgenic mice (FIG. **18**). This mouse model carries a transgene consisting partly of human HLA-A*0201 and mouse H2-Kb thus encoding a chimeric class I MHC molecule consisting of the human HLA-A2.1 leader, α1 and α2 domains ligated to the murine α3, transmembrane and cytoplasmic H2-Kb domain (Vitiello et al., 1991). The chimeric molecule allows HLA-A*02:01-restricted antigen presentation whilst maintaining the species-matched interaction of the CD8 co-receptor with the α3 domain on the MHC.

[0849] For the short cassettes, all marker epitopes generated a T cell response, as determined by IFN-gamma ELISPOT, that was approximately 10-50× stronger of what has been commonly reported (Comet et al., 2006; Depla et al., 2008; Ishioka et al., 1999). Of all the linkers evaluated, the concatamer of 25mer sequences, each containing a minimal epitope flanked by their natural amino acids sequences, generated the largest and broadest T cell response (Table 5). Intracellular cytokine staining (ICS) and flow cytometry analysis revealed that the antigen-specific T cell responses are derived from CD8 T cells.

TABLE-US-00012 TABLE 5 In vivo evaluation of linker sequences in short cassettes. ELISPOT data indicated that HLA-A2 transgenic mice, 17 days post-infection with 1e11 adenovirus viral particles, generated a T cell response to all class I MHC restricted epitopes in the cassette. Short Cassette Designs Epitope 9AA 17AA 25AA AAY RR DPP 1 2020 +/− 583 2505 +/− 1281 6844 +/− 956 1489 +/− 762 1675 +/− 690 1781 +/− 774 2 4472 +/− 755 3792 +/− 1319 7629 +/− 996 3851 +/− 1748 4726 +/− 1715 5868 +/− 1427 3 5830 +/− 315 3629 +/− 862 7253 +/− 491 4813 +/− 1761 6779 +/− 1033 7328 +/− 1700 4 5536 +/− 375 2446 +/− 955 2961 +/− 1487 4230 +/− 1759 6518 +/− 909 7222 +/− 1824 5 8800 +/− 0 7943 +/− 821 8423 +/− 442 8312 +/− 696 8800 +/− 0 1836 +/− 328

[0850] In another example, a series of long vaccine cassettes was constructed and incorporated in adenoviral vectors that, next to the original 5 marker epitopes, contained an additional 16 HLA-A*02:01, A*03:01 and B*44:05 epitopes with known CD8 T cell reactivity (FIG. **19**A, B). The size of these long cassettes closely mimicked the final clinical cassette design, and only the position of the epitopes relative to each other was varied. The CD8 T cell responses were comparable in magnitude and breadth for both long and short vaccine cassettes, demonstrating that (a) the addition of more epitopes did not substantially impact the magnitude of immune response to the original set of epitopes, and (b) the position of an epitope in a cassette did not substantially influence the ensuing T cell response to it (Table 6).

TABLE-US-00013 TABLE 6 In vivo evaluation of the impact of epitope position in long cassettes. ELISPOT data indicated that HLA-A2 transgenic mice, 17 days post-infection with 5e10 adenovirus viral particles, generated a T cell response comparable in magnitude for both long and short vaccine cassettes. Long Cassette Designs Epitope Standard Scrambled Short 1 863 +/− 1080 804 +/− 1113 1871 +/− 2859 2 6425 +/− 1594 28 +/− 62 5390 +/− 1357 3* 23 +/− 30 36 +/− 18 0 +/− 48 4 2224 +/− 1074 2727 +/− 644 2637 +/− 1673 5 7952 +/− 297 8100 +/− 0 8100 +/− 0 8100 +/− 0 *Suspected technical error caused an absence of a T cell response.

XIV.B.4. Neoantigen Cassette Design for Immunogenicity and Toxicology Studies

[0851] In summary, the findings of the model cassette evaluations (FIG. **16**-**19**, Tables 2-6) demonstrated that, for model vaccine cassettes, robustimmunogenicity was achieved when a “string of beads” approach was employed that encodes around 20 epitopes in the context of an adenovirus-based vector. The epitopes were assembled by concatenating 25mer sequences, each embedding a minimal CD8 T cell epitope (e.g. 9 amino acid residues) that were flanked on both sides by its natural, surrounding peptide sequence (e.g. 8 amino acid residues on each side). As used herein, a “natural” or “native” flanking sequence refers to the N- and/or C-terminal flanking sequence of a given epitope in the naturally occurring context of that epitope within its source protein. For example, the HCMV pp65 MHC I epitope NLVPMVATV (SEQ ID NO: 132) is flanked on its 5′ end by the native 5′ sequence WQAGILAR (SEQ ID NO: 139) and on its 3′ end by the native 3′ sequence QGQNLKYQ (SEQ ID NO: 140), thus generating the WQAGILARNLVPMVATVQGQNLKYQ (SEQ ID NO: 141) 25mer peptide found within the HCMV pp65 source protein. The natural or native sequence can also refer to a nucleotide sequence that encodes an epitope flanked by native flanking sequence(s). Each 25mer sequence is directly connected to the following 25mer sequence. In instances where the minimal CD8 T cell epitope is greater than or less than 9 amino acids, the flanking peptide length can be adjusted such that the total length is still a 25mer peptide sequence. For example, a 10 amino acid CD8 T cell epitope can be flanked by an 8 amino acid sequence and a 7 amino acid. The concatamer was followed by two universal class II MHC epitopes that were included to stimulate CD4 T helper cells and improve overall in vivo immunogenicity of the vaccine cassette antigens. (Alexander et al., 1994; Panina-Bordignon et al., 1989) The class II epitopes were linked to the final class I epitope by a GPGPG amino acid linker (SEQ ID NO:56). The two class II epitopes were also linked to each other by a GPGPG amino acid linker (SEQ ID NO: 56), as a well as flanked on the C-terminus by a GPGPG amino acid linker (SEQ ID NO: 56). Neither the position nor the number of epitopes appeared to substantially impact T cell recognition or

response. Targeting sequences also did not appear to substantially impact the immunogenicity of cassette-derived antigens.

[0852] As a further example, based on the in vitro and in vivo data obtained with model cassettes (FIG. **16-19**, Tables 2-6), a cassette design was generated that alternates well-characterized T cell epitopes known to be immunogenic in nonhuman primates (NHPs), mice and humans. The 20 epitopes, all embedded in their natural 25mer sequences, are followed by the two universal class II MHC epitopes that were present in all model cassettes evaluated (FIG. **20**). This cassette design was used to study immunogenicity as well as pharmacology and toxicology studies in multiple species.

XV. ChAd Neoantigen Cassette Delivery Vector

XV.A. ChAd Neoantigen Cassette Delivery Vector Construction

[0853] In one example, Chimpanzee adenovirus (ChAd) was engineered to be a delivery vector for neoantigen cassettes. In a further example, a full-length ChAdV68 vector was synthesized based on AC_000011.1 (sequence 2 from U.S. Pat. No. 6,083,716) with E1 (nt 457 to 3014) and E3 (nt 27,816-31,332) sequences deleted. Reporter genes under the control of the CMV promoter/enhancer were inserted in place of the deleted E1 sequences. Transfection of this clone into HEK293 cells did not yield infectious virus. To confirm the sequence of the wild-type C68 virus, isolate VR-594 was obtained from the ATCC, passaged, and then independently sequenced (SEQ ID NO:10). When comparing the AC_000011.1 sequence to the ATCC VR-594 sequence (SEQ ID NO:10) of wild-type ChAdV68 virus, 6 nucleotide differences were identified. In one example, a modified ChAdV68 vector was generated based on AC_000011.1, with the corresponding ATCC VR-594 nucleotides substituted at five positions (ChAdV68.5WTnt SEQ ID NO:1).

[0854] In another example, a modified ChAdV68 vector was generated based on AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,816-31,332) sequences deleted and the corresponding ATCC VR-594 nucleotides substituted at four positions. A GFP reporter (ChAdV68.4WTnt.GFP; SEQ ID NO:11) or model neoantigen cassette (ChAdV68.4WTnt.MAG25mer; SEQ ID NO:12) under the control of the CMV promoter/enhancer was inserted in place of deleted E1 sequences.

[0855] In another example, a modified ChAdV68 vector was generated based on AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,125-31,825) sequences deleted and the corresponding ATCC VR-594 nucleotides substituted at five positions. A GFP reporter (ChAdV68.5WTnt.GFP; SEQ ID NO:13) or model neoantigen cassette (ChAdV68.5WTnt.MAG25mer; SEQ ID NO:2) under the control of the CMV promoter/enhancer was inserted in place of deleted E1 sequences.

TABLE-US-00014 Full-Length ChAdVC68 sequence “ChAdV68.5WTnt” (SEQ ID NO: 1); AC_000011.1 sequence with corresponding ATCC VR-594 nucleotides substituted at five positions.

CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTT
GGGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGTTTT
GATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGG
TGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTTTCTGG
GCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAATCTGAGTA
ATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGATTACGTGG
GGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGTGTTTTTAC
GTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCCACTCTTGAG
TGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAAGATGAGGCAC
CTGAGAGACCTGCCCGATGAGAAAATCATCATCGCTTCCGGGAACGAGATTCTGGAACTGGTGGT
AAATGCCATGATGGGCGACGACCCTCCGGAGCCCCCCACCCCATTTGAGACACCTTCGCTGCACG
ATTTGTATGATCTGGAGGTGGATGTGCCCGAGGACGATCCCAATGAGGAGGCGGTAAATGATTTT
TTTAGCGATGCCGCGCTGCTAGCTGCCGAGGAGGCTTCGAGCTCTAGCTCAGACAGCGACTCTTC
ACTGCATACCCCTAGACCCGGCAGAGGTGAGAAAAAGATCCCCGAGCTTAAAGGGGAAGAGATGG
ACTTGCGCTGCTATGAGGAATGCTTGCCCCCGAGCGATGATGAGGACGAGCAGGCGATCCAGAAC
GCAGCGAGCCAGGGAGTGCAAGCCGCCAGCGAGAGCTTTGCGCTGGACTGCCCGCCTCTGCCCGG
ACACGGCTGTAAGTCTTGTGAATTTCATCGCATGAATACTGGAGATAAAGCTGTGTTGTGTGCAC
TTTGCTATATGAGAGCTTACAACCATTGTGTTTACAGTAAGTGTGATTAAGTTGAACTTTAGAGG
GAGGCAGAGAGCAGGGTGACTGGGCGATGACTGGTTTATTTATGTATATATGTTCTTTATATAGG
TCCCGTCTCTGACGCAGATGATGAGACCCCCACTACAAAGTCCACTTCGTCACCCCCAGAAATTG
GCACATCTCCACCTGAGAATATTGTTAGACCAGTTCCTGTTAGAGCCACTGGGAGGAGAGCAGCT
GTGGAATGTTTGGATGACTTGCTACAGGGTGGGGTTGAACCTTTGGACTTGTGTACCCGGAAACG
CCCCAGGCACTAAGTGCCACACATGTGTGTTTACTTGAGGTGATGTCAGTATTTATAGGGTGTGG
AGTGCAATAAAAAATGTGTTGACTTTAAGTGCGTGGTTTATGACTCAGGGGTGGGGACTGTGAGT

ATATAAGCAGGTGCAGACCTGTGTGGTTAGCTCAGAGCGGCATGGAGATTTGGACGGTCTTGGAA
GACTTTCACAAGACTAGACAGCTGCTAGAGAACGCCTCGAACGGAGTCTCTTACCTGTGGAGATT
CTGCTTCGGTGGCGACCTAGCTAGGCTAGTCTACAGGGCCAAACAGGATTATAGTGAACAATTTG
AGGTTATTTTGAGAGAGTGTTCTGGTCTTTTTGACGCTCTTAACTTGGGCCATCAGTCTCACTTT
AACCAGAGGATTTCGAGAGCCCTTGATTTTACTACTCCTGGCAGAACCACTGCAGCAGTAGCCTT
TTTTGCTTTTATTCTTGACAAATGGAGTCAAGAAACCCATTTCAGCAGGGATTACCAGCTGGATT
TCTTAGCAGTAGCTTTGTGGAGAACATGGAAGTGCCAGCGCCTGAATGCAATCTCCGGCTACTTG
CCGGTACAGCCGCTAGACACTCTGAGGATCCTGAATCTCCAGGAGAGTCCCAGGGCACGCCAACG
TCGCCAGCAGCAGCAGCAGGAGGAGGATCAAGAAGAGAACCCGAGAGCCGGCCTGGACCCTCCGG
CGGAGGAGGAGGAGTAGCTGACCTGTTTCCTGAACTGCGCCGGGTGCTGACTAGGTCTTCGAGTG
GTCGGGAGAGGGGGATTAAGCGGGAGAGGCATGATGAGACTAATCACAGAACTGAACTGACTGTG
GGTCTGATGAGTCGCAAGCGCCCAGAAACAGTGTGGTGGCATGAGGTGCAGTCGACTGGCACAGA
TGAGGTGTCGGTGATGCATGAGAGGTTTTCTCTAGAACAAGTCAAGACTTGTTGGTTAGAGCCTG
AGGATGATTGGGAGGTAGCCATCAGGAATTATGCCAAGCTGGCTCTGAGGCCAGACAAGAAGTAC
AAGATTACTAAGCTGATAAATATCAGAAATGCCTGCTACATCTCAGGGAATGGGGCTGAAGTGGA
GATCTGTCTCCAGGAAAGGGTGGCTTTCAGATGCTGCATGATGAATATGTACCCGGGAGTGGTGG
GCATGGATGGGGTTACCTTTATGAACATGAGGTTCAGGGGAGATGGGTATAATGGCACGGTCTTT
ATGGCCAATACCAAGCTGACAGTCCATGGCTGCTCCTTCTTTGGGTTTAATAACACCTGCATCGA
GGCCTGGGGTCAGGTCGGTGTGAGGGGCTGCAGTTTTTCAGCCAACTGGATGGGGGTCGTGGGCA
GGACCAAGAGTATGCTGTCCGTGAAGAAATGCTTGTTTGAGAGGTGCCACCTGGGGGTGATGAGC
GAGGGCGAAGCCAGAATCCGCCACTGCGCCTCTACCGAGACGGGCTGCTTTGTGCTGTGCAAGGG
CAATGCTAAGATCAAGCATAATATGATCTGTGGAGCCTCGGACGAGCGCGGCTACCAGATGCTGA
CCTGCGCCGGCGGGAACAGCCATATGCTGGCCACCGTACATGTGGCTTCCCATGCTCGCAAGCCC
TGGCCCGAGTTCGAGCACAATGTCATGACCAGGTGCAATATGCATCTGGGGTCCCGCCGAGGCAT
GTTCATGCCCTACCAGTGCAACCTGAATTATGTGAAGGTGCTGCTGGAGCCCGATGCCATGTCCA
GAGTGAGCCTGACGGGGGTGTTTGACATGAATGTGGAGGTGTGGAAGATTCTGAGATATGATGAA
TCCAAGACCAGGTGCCGAGCCTGCGAGTGCGGAGGGAAGCATGCCAGGTTCCAGCCCGTGTGTGT
GGATGTGACGGAGGACCTGCGACCCGATCATTTGGTGTTGCCCTGCACCGGGACGGAGTTCGGTT
CCAGCGGGGAAGAATCTGACTAGAGTGAGTAGTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGC
CAGAATAACTGAAATCTGTGCTTTTCTGTGTGTTGCAGCAGCATGAGCGGAAGCGGCTCCTTTGA
GGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCCTCCTGGGCGGGAGTGCGTCAGAATG
TGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCGCGAACTCTTCAACCCTGACCTATGCA
ACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGCAGCTGCTGCATCTGCCGCCAGCGCCGT
GCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACTCGAGTTCCACCAATA
ATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCCAGCTCGAGGCCTTGACCCAG
CGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCAC
GGTGAAATCCAAATAAAAAATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAACACAGAG
TCTGAATCTTTATTTGATTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCA
CCCGGTGGATCTTTTCCAGGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGC
CCGTCCCGGGGGGTGGAGGTAGCTCCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCAC
CCAGTCATAGCAGGGGCGCAGGGCATGGTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCA
CGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAG
ATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATGTTACCGCCCAGATCCCGCCTGGGGTT
CATGTTGTGCAGGACCACCAGCACGGTGTATCCGGTGCACTTGGGGAATTTATCATGCAACTTGG
AAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCGCCCAGGTTTTCCATGCACTCATCC
ATGATGATGGCGATGGGCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATC
ATAGTTGTGGTCCTGGGTGAGGTCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGG
ACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAG
GCTTTGAGCTCGGAGGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGC
GGGGGAGATGAGCTGGGCCGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGC
CGTAGATGACCCCGATGACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGG
AGGAGGGGGGCCACCTCGTTCATCATCTCGCGCACGTGCATGTTCTCGCGCACCAGTTCCGCCAG
GAGGCGCTCTCCCCCCAGGGATAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTC
CGTCGGCCATGGGCATTTTGGAGAGGGTTTGTTGCAAGAGTTCCAGGCGGTCCCAGAGCTCGGTG
ATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGACGGCTGCGGGAG
TAGGGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCG
CGTCAGGGTGGTCTCCGTCACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTGCGCT

TCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAGGTAGCAA
TTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGA
AGTCTGCCCGCAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGG
ACTCGGGGGCGTAGGCGTCCGCGCCGCAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTG
AGGTCGGGCTGGTCGGGGTCAAAAACCAGTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTT
GGTCTCCATGAGCTCGTGTCCCCGCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACT
TTATGGGCCGGTCCTCGAGCGGTGTGCCGCGGTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAG
ACGAAAGCCCGGGTCCAGGCCAGCACGAAGGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCAC
CAGCGGGTCCACCTTTTCCAGGGTATGCAAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGA
TTGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCCGGCCGGGGGGGGTATAAAAGGGTGCGGGT
CCCTGCTCGTCCTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTC
CCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGA
TATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAGACGATC
TTTTTGTTGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGA
GCGCATGGTCTGGTTTTTTTCCTTGTCGGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACT
CGCGCGCCACGCACTTCCATTCGGGGAAGACGGTGGTCAGCTCGTCGGGCACGATTCTGACCTGC
CAGCCCCGATTATGCAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATT
AGTCCAGCAGAGGCGTCCGCCCTTGCGCGAGCAGAAGGGGGGCAGGGGGTCCAGCATGACCTCGT
CGGGGGGGTCGGCATCGATGGTGAAGATGCCGGGCAGGAGGTCGGGGTCAAAGTAGCTGATGGAA
GTGGCCAGATCGTCCAGGGCAGCTTGCCATTCGCGCACGGCCAGCGCGCGCTCGTAGGGACTGAG
GGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGAGGCGTACATGCCGCAGATGTCGTAGACGT
AGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGC
ACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTT
TTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGA
AGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGC
AGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGAT
GATGTCATACTTGAGCTGTCCCTTTTGTTTCCACAGCTCGCGGTTGAGAAGGAACTCTTCGCGGT
CCTTCCAGTACTCTTCGAGGGGGAACCCGTCCTGATCTGCACGGTAAGAGCCTAGCATGTAGAAC
TGGTTGACGGCCTTGTAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTT
GCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGTCCCTGACCATGACCTTGAGGAACTGGTGCTTGA
AGTCGATATCGTCGCAGCCCCCCTGCTCCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGG
TTGGGCAAAGCGAAAGTAACATCGTTGAAGAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGT
GATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTGATGACCTGGGCGGCGAGCACGATCTCGT
CGAAGCCGTTGATGTTGTGGCCCACGATGTAGAGTTCCACGAATCGCGGACGGCCCTTGACGTGG
GGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGGGGTCGCTGAGCCCGTGCTGCTCGAGCGC
CCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCCAGGGCGGTTT
GCAGACGGTCCCGGTACTGACGGAACTGCTGCCCGACGGCCATTTTTTCGGGGGTGACGCAGTAG
AAGGTGCGGGGGTCCCCGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTC
GACGAGCCGGTCGTCCCCGGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGG
ACCCCATCCAGGTGTAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAG
CCGATGGGGAAGAACTGGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTA
GAAATGCCGACGGCGCGCCGAACACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAAC
GCTGCACGGGATGCACGTGCTGCACGAGCTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGGG
AAGTGGAGTCGTGGCGCCTGCATCTCGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTCTTC
TGCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGGCGCGAGCGG
GTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTC
AGGTCAGTGGGCAGCGGCGGCGCGCGGTTGACTTGCAGGAGTTTTTCCAGGGCGCGCGGGAGGTC
CAGATGGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTCGATGGCTTGCAGGGTCCCGTGCC
CCTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGCGACGGGGGCGGTGCCTCT
TCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCCGGGCGGCAGGGGCGGCTCGGGGCCCGGAGGC
AGGGGCGGCAGGGGCACGTCGGCGCCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACT
GGCGTGAGCGACGACGCGACGGTTGACGTCCTGGATCTGACGCCTCTGGGTGAAGGCCACGGGAC
CCGTGAGTTTGAACCTGAAAGAGAGTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGC
CGCAGGATCTCTTGCACGTCGCCCGAGTTGTCCTGGTAGGCGATCTCGGTCATGAACTGCTCGAT
CTCCTCCTCTTGAAGGTCTCCGCGGCCGGCGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGC
GGCCCATGAGCTGCGAGAAGGCGTTCATGCCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACG
CCCTCGGGATCGCgGGCGCGCATGACCACCTGGGCGAGGTTGAGCTCCACGTGGCGCGTGAAGAC

CGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGCGATGTGCTCGGTGACGAAGA
AATACATGATCCAGCGGCGGAGCGGCATCTCGCTGACGTCGCCCAGCGCCTCCAAACGTTCCATG
GCCTCGTAAAAGTCCACGGCGAAGTTGAAAAACTGGGAGTTGCGCGCCGAGACGGTCAACTCCTC
CTCCAGAAGACGGATGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCCCGGGAGTT
CCTCCACTTCCTCTTCTTCCTCCTCCACTAACATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGT
GGCGGGGGAGGGGGCCTGCGTCGCCGGCGGCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGT
CTCGCCGCGCCGGCGTCGCATGGTCTCGGTGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGA
AGACGCCGCCGCGCATCTCCAGGTGGCCGGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACG
ATGCATCTTATCAATTGCCCCGTAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGG
ATCTGAAAACCGCTGAACGAAGGCTTCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTT
CTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGCGGGCGATGCTGCTGGTGATGAAGTTGAAATAG
GCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCTTGCTGGATGCG
CAGACGGTCGGCCATGCCCCAGGCGTGGTCCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGCA
TGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCGGCCGTGCATGCGCGTGAGCCCGAAGCCG
CGCTGGGGCTGGACGAGCGCCAGGTCGGCGACGACGCGCTCGGCGAGGATGGCTTGCTGGATCTG
GGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGTGT
AGGAGCAGTTGGCCATGACGGACCAGTTGACGGTCTGGTGGCCCGGACGCACGAGCTCGTGGTAC
TTGAGGCGCGAGTAGGCGCGCGTGTCGAAGATGTAGTCGTTGCAGGTGCGCACCAGGTACTGGTA
GCCGATGAGGAAGTGCGGCGGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGG
GCGCGAGGTCCTCGAGCATGGTGCGGTGGTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCG
GCGGCGGTGGTGGAGGCGCGCGGGAACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAA
GTAGTTCATGGTGGGCACGGTCTGGCCCGTGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCA
AAAACGAAAGCGGTCAGCGGCTCGACTCCGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGC
GTGTACCCCGGTTCGAATCTCGAATCAGGCTGGAGCCGCAGCTAACGTGGTATTGGCACTCCCGT
CTCGACCCAAGCCTGCACCAACCCTCCAGGATACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGG
AGGCCGGATGAGACTAGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTGGA
GAAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGGTTCGAGGCCGGCCGGATTCCGCGGCTAAC
GAGGGCGTGGCTGCCCCGTCGTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCG
AGCCCCTCTTTTGTTTTGTTTGTTTTTGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCAC
CACCCTCCACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACT
TCCAGCCACGACCGCCGCGGCCGCCGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCT
TGGAAGAGGGCGAGGGGCTGGCGCGCCTGGGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAG
ATGAAAAGGGACGCTCGCGAGGCCTACGTGCCCAAGCAGAACCTGTTCAGAGACAGGAGCGGCGA
GGAGCCCGAGGAGATGCGCGCGGCCCGGTTCCACGCGGGGGGGGAGCTGCGGCGCGGCCTGGACC
GAAAGAGGGTGCTGAGGGACGAGGATTTCGAGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGC
GCGCACGTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTT
CCAAAAATCCTTCAACAACCACGTGCGCACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGA
TGCACCTGTGGGACCTGCTGGAGGCCATCGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGCAG
CTGTTCCTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTCAGGGAGGCGCTGCTGAATATCAC
CGAGCCCGAGGGCCGCTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGC
GCGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTAC
TACGCTAGGAAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTT
TTACATGCGCATGACCCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACA
GGATGCACCGTGCGGTGAGCGCCAGCAGGCGGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGT
CTGCAGCGGGCCCTGACCGGGGCCGGGACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCT
GCACTGGCAGCCCAGCCGCCGGGCCTTGGAGGCGGCGGCAGGACCCTACGTAGAAGAGGTGGACG
ATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGC
AACAACAACAGCCACCTCCTGATCCCGCGATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATT
AACTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGACGACCCGCAACCCCGA
AGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGC
GCTCCAACCCCACGCACGAGAAGGTCCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATC
CGCGGCGACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCAC
CAACGTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGC
GGTTCCACCGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCC
GCCAACGTGCCCCGGGGCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGAC
CGAGGTGCCCCAGAGCGAGGTGTACCAGTCCGGGCCGGACTACTTCTTCCAGACCAGTCGCCAGG
GCTTGCAGACCGTGAACCTGAGCCAGGCTTTCAAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCC

CCGGTCGGGGACCGCGCGACGGTGTCGAGCCTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCT
GGTGGCCCCCTTCACGGACAGCGGCAGCATCAACCGCAACTCGTACCTGGGCTACCTGATTAACC
TGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACCTACCAGGAGATCACCCACGTG
AGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAA
CCGGTCGCAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTGCGTTACG
TGCAGCAGAGCGTGGGCCTGTTCCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATG
ACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAACTGATGGA
CTACTTGCATCGGGCGGCCGCCATGAACTCTGACTATTTCACCAACGCCATCCTGAATCCCCACT
GGCTCCCGCCGCCGGGGTTCTACACGGGCGAGTACGACATGCCCGACCCCAATGACGGGTTCCTG
TGGGACGATGTGGACAGCAGCGTGTTCTCCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAA
GAAGGAAGGCAGCGACCGACGCCCGTCCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGG
TGCCCGAGGCCGCCAGTCCTTTCCCGAGCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAG
CTGGGCAGGATCACGCGCCCGCGCTTGCTGGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAG
ACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAAAGCCTGGTGGACAAGATGAGCCGCT
GGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTCGCAGGGGGCCACGAGCCGGGGC
AGCGCCGCCCGTAAACGCCGGTGGCACGACAGGCAGCGGGGACAGATGTGGGACGATGAGGACTC
CGCCGACGACAGCAGCGTGTTGGACTTGGGTGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCC
GTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACCAG
CGTGCGTTCGTTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGT
CCTCCTCCCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCT
GGAGGCTCCTTACGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACT
CGGAGCTGGCACCCTTGTACGATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATC
GCCTCGCTGAACTACCAGAACGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTT
CACCCCCACGGAGGCCAGCACCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGC
TGAAAACCATCATGCACACCAACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCAAG
GCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGACAGTGACAGAGGATTATGATGGTAG
TCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCA
TGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGGCGGCAGAACGGG
GTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACTAGGAACTTCAGGCTGGGCTGGGACCCCGT
GACCGAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGC
CCGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCAG
CCCTTCCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCT
GGATGTCGACGCCTATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTA
CCGCCTCTACCGAGGTCAGGGGCGATAATTTTGCAAGCGCCGCAGCAGTGGCAGCGGCCGAGGCG
GCTGAAACCGAAAGTAAGATAGTCATTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAA
CGTACTACCGGACAAGATAAACACCGCCTACCGCAGCTGGTACCTAGCCTACAACTATGGCGACC
CCGAGAAGGGCGTGCGCTCCTGGACGCTGCTCACCACCTCGGACGTCACCTGCGGCGTGGAGCAA
GTCTACTGGTCGCTGCCCGACATGATGCAAGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAG
CAACTACCCGGTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGG
CCGTCTACTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAG
AACCAGATCCTCGTCCGCCCGCCCGCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCT
CACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTG
ACGCCAGACGCCGCACCTGCCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTC
TCGAGCCGCACCTTCTAAATGTCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCG
CGCGCCCAGCAAGATGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCGCGTGCGCG
GGCACTTCCGCGCTCCCTGGGGCGCCCTCAAGGGCCGCGTGCGGTCGCGCACCACCGTCGACGAC
GTGATCGACCAGGTGGTGGCCGACGCGCGCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGT
GGACGCCGTCATCGACAGCGTGGTGGCcGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGC
GGCGCATCGCCCGGCGGCACCGGAGCACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGG
GCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGC
CGGCAGGACCCGGAGACGCGCGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGC
GGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGC
CCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCC
AAGCGCAAATTCAAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGT
GAAGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTG
ATGTGGACGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCCGGCGGCGCGTGCAGTGGCGCGGG
CGGAAGGTGCAACCGGTGCTGAGACCCGGCACCACCGTGGTCTTCACGCCCGGCGAGCGCTCCGG

CACCGCTTCCAAGCGCTCCTACGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCCG
AGCGCCTGGGCGAGTTTGCTTACGGCAAGCGCAGCCGTTCCGCACCGAAGGAAGAGGCGGTGTCC
ATCCCGCTGGACCACGGCAACCCCACGCCGAGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCC
GACCGCGGCGCCGCGCCGGGGGTTCAAGCGCGAGGGCGAGGATCTGTACCCCACCATGCAGCTGA
TGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGAAGGTGGACCCGGACGTGCAG
CCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACAT
CAAGATTCCCACGGAGCCCATGGAAACGCAGACCGAGCCCATGATCAAGCCCAGCACCAGCACCA
TGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACGGC
GCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCG
CGGCACGCGCTTCTACCGCGGTCATACCAGCAGCCGCCGCCGCAAGACCACCACTCGCCGCCGCC
GTCGCCGCACCGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGC
GCACCTCTGACCCTGCCGCGCGCGCGCTACCACCCGAGCATCGCCATTTAAACTTTCGCCtGCTT
TGCAGATCAATGGCCCTCACATGCCGCCTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACC
GCGCCGTAGAAGGCTGGCGGGGAACGGGATGCGTCGCCACCACCACCGGCGGCGGCGCGCCATCA
GCAAGCGGTTGGGGGGAGGCTTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGATCGGGGCG
ATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACTTGGAAACAT
CTTGTAATAAACCaATGGACTCTGACGCTCCTGGTCCTGTGATGTGTTTTCGTAGACAGATGGAA
GACATCAATTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTCATGGGCACCTGGAGCGA
CATCGGCACCAGCCAACTGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGA
ATTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACCACAGGGCAGGCGCTG
AGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAA
CGGGGTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGC
CGCCCGCCGGCTCCGTGGAGATGCCGCAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGC
GAGAAGCGACCCCGCCCCGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCCCGTA
CGAGGAGGCGGTGAAACTGGGTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGC
TGAAACCCGAAAAGCCCGCGACCCTGGACTTGCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTG
GCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCGCGCGCGACCCGGGGGCACCGCCCGCCCTCATGC
GAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCT
GCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTGTATGTATTATGTCGCCGCCGC
CGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGATGGCCACCCCATCGA
TGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGTCCGGGT
CTGGTGCAGTTTGCCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCAC
GGTGGCGCCCACGCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCG
TGGACCGCGAGGACAACACCTACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCGC
GTGCTGGACATGGCCAGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAA
ACCCTACTCCGGCACCGCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGA
CATATAAAGCCGATGGTGAAACTGCCACAGAAAAAACCTATACATATGGAAATGCACCCGTGCAG
GGCATTAACATCACAAAAGATGGTATTCAACTTGGAACTGACACCGATGATCAGCCAATCTACGC
AGATAAAACCTATCAGCCTGAACCTCAAGTGGGTGATGCTGAATGGCATGACATCACTGGTACTG
ATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTTTT
GCCAAGCCTACTAATAAAGAAGGAGGTCAGGCAAATGTGAAAACAGGAACAGGCACTACTAAAGA
ATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAA
TTGTTTTGTATACTGAAAATGTGGATTTGGAAACTCCAGATACCCATATTGTATACAAAGCAGGC
ACAGATGACAGCAGCTCTTCTATTAATTTGGGTCAGCAAGCCATGCCCAACAGACCTAACTACAT
TGGTTTCAGAGACAACTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGG
CCGGTCAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCCTAC
CAGCTCTTGCTTGACTCTCTGGGTGACAGAACCCGGTATTTCAGTATGTGGAATCAGGCGGTGGA
CAGCTATGATCCTGATGTGCGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATT
GTTTCCCTCTGGATGCTGTTGGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACTGAT
CAAACCACATGGACCAAAGATGACAGTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCATT
CGCCATGGAAATCAACATCCAAGCCAACCTGTGGAGGAACTTCCTCTACGCCAACGTGGCCCTGT
ACCTGCCCGACTCTTACAAGTACACGCCGGCCAATGTTACCCTGCCCACCAACACCAACACCTAC
GATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTACATCAACATCGGGGCGCG
CTGGTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATGCGGGGCTGCGCT
ACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTT
TTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGA
CGTCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCT

TCACCAGCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAG
GCCATGCTGCGCAACGACACCAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCT
CTACCCCATCCCGGCCAACGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCT
TCCGCGGCTGGTCCTTCACGCGTCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGAC
CCCTACTTCGTCTACTCGGGCTCCATCCCCTACCTCGACGGCACCTTCTACCTCAACCACACCTT
CAAGAAGGTCTCCATCACCTTCGACTCCTCCGTCAGCTGGCCCGGCAACGACCGGCTCCTGACGC
CCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAGGGCTACAACGTGGCCCAGTGCAACATG
ACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCACTACAACATCGGCTACCAGGGCTTCTACGT
GCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCATGAGCCGCCAGG
TGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTCACCCTGGCCTACCAGCACAACAACTCG
GGCTTCGTCGGCTACCTCGCGCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTA
CCCGCTCATCGGCAAGAGCGCCGTCACCAGCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCA
TGTGGCGCATCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAAC
ATGCTCTATGCCAACTCCGCCCACGCGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTC
CACCCTTCTCTATGTTGTCTTCGAAGTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCG
TCATCGAGGCCGTCTACCTGCGCACCCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCT
TCTTGCAAGCCATGGCCGCGGGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGC
TGCGGGCCCTACTTCCTGGGCACCTTCGATAAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCT
GGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCT
GGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAG
CAGATCTACCAGTTCGAGTACGAGGGCCTGCTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTG
CGTCACCCTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCT
GCTGCATGTTCCTGCACGCCTTCGTGCACTGGCCCGACCGCCCCATGGACAAGAACCCCACCATG
AACTTGCTGACGGGGGTGCCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCG
CAACCAGGAGGCGCTCTACCGCTTCCTCAACTCCCACTCCGCCTACTTTCGCTCCCACCGCGCGC
GCATCGAGAAGGCCACCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAA
ATGTCTTTAATAAACAGCACTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAA
AGGGTTCTGCCGGGTCTCGGCATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTACTTGGCCA
GCCACTTGAACTCGGGGATCAGCAGTTTGGGCAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGC
TTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACC
CGCGTTCTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAGCACTGGAACACCATCAGGGCCGGGT
GCTTCACGCTCGCCAGCACCGTCGCGTCGGTGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCC
ATCCCGAAGGGGGTCATCTTGCAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCTTGTGGTT
GCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACATGG
CCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACC
CCGCAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTCGTGCACGCAGCAGCGCGCGTC
GTTGTTGGCCAGCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTCGGGGT
TCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGG
ATCATGGTGGTCCCGTGCAGGCACCGCAGCTTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAG
CGCGCACCCGGTGCACTCCCAGTTCTTGTGGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCA
GGAAGCGGCCCATCATGGTGGTCAGGGTCTTGTTGCTAGTGAAGGTCAGCGGAATGCCGCGGTGC
TCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACCTCGCCCTGCTCGGGCATCAGCTGGAA
GTTGGCTTTCAGGTCGGTCTCCACGCGGTAGCGGTCCATCAGCATAGTCATGATTTCCATACCCT
TCTCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTTCTTCACCATCATCTTAGCGCTAGCAGCC
GCGGCCAGGGGGTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCG
CACCGGGGGGTAGCTGAAGCCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGT
CCTGGCTGACGTCCTGCAGGACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGC
GGCGGAGATGTTGGAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTATCTCTTCCTC
TTCTTGGTCCGAGGCCACGCGGCGGTAGGTATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGC
TCTCGCCGCCGCGACTTGGCGGATGGCTGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGG
CGGCGCTCTGACTGACTTCCTCCGCGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCAT
GGAGACTCAGCCATCGCCAACCTCGCCATCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGA
ATGAAAGCTTAACCGCCCCGCCGCCCAGCCCCGCCACCTCCGACGCGGCCGTCCCAGACATGCAA
GAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGACGCCCGCGGAGCACGAGGAGGAGCT
GGCAGTGCGCTTTTCACAAGAAGAGATACACCAAGAACAGCCAGAGCAGGAAGCAGAGAATGAGC
AGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGGGAGGACGCGCTC
ATCAAGCATCTGGCCCGGCAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCGAGGTGCC

CCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCA
AGCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTG
CCCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGC
CAACCGCACCCGCGCCGACGCCCTTTTCAACCTGGGTCCCGGCGCCCGCCTACCTGATATCGCCT
CCTTGGAAGAGGTTCCCAAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCT
CTGCAAGGAGAAGGAGGAGAGCATGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGC
GCGGCTGGCGGTGCTCAAACGCACGGTCGAGCTGACCCATTTCGCCTACCCGGCTCTGAACCTGC
CCCCCAAAGTCATGAGCGCGGTCATGGACCAGGTGCTCATCAAGCGCGCGTCGCCCATCTCCGAG
GACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTG
GCTGGGTCCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGCGCAAACTCATGATGGCCGTGGTCC
TGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGCGGAGACCCTGCGCAAGGTC
GAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCAACGT
GGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGTGC
TGCACACCACCCTGCGCGGGGAGGCCCGGCGCGACTACATCCGCGACTGCGTCTACCTCTACCTC
TGCCACACCTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGA
GCTCTGCAAGCTCCTGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCG
CCTCGGACCTGGCCGACCTCATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCC
GACTTTATGAGCCAAAGCATGTTGCAAAACTTTCGCTCTTTCATCCTCGAACGCTCCGGAATCCT
GCCCGCCACCTGCTCCGCGCTGCCCTCGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGC
CGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATC
GAGGACGTCAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACCG
CTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGCAAG
GGCCCAGCGAAGGCGAGGGTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTGTGGACC
TCGGCCTACTTGCGCAAGTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGAGGA
CCAATCCCATCCGCCCAAGGCCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCC
AATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTC
GACCCCCAGACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGC
TGAAAGTGGAGCTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAG
GAGGAGGAGATGGAGGAAGACTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCT
GGAGGAAGACGAGGAGGAGGCAGAGGAGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCT
CGGCGGGGGAGAAAGCAAGCAGCACGGATACCATCTCCGCTCCGGGTCGGGGTCCCGCTCGACCA
CACAGTAGATGGGACGAGACCGGACGATTCCCGAACCCCACCACCCAGACCGGTAAGAAGGAGCG
GCAGGGATACAAGTCCTGGCGGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGG
GCAACATCTCCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAACTTTCCCCGCAACATC
TTGCATTACTACCGTCACCTCCACAGCCCCTACTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAA
AGACCAGCAGAAAACCAGCAGCTAGAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGATCGCG
GCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCCACCCTCTATGCCATCTT
CCAGCAGAGTCGGGGGCAGGAGCAGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCC
GCAGTTGTCTGTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTC
TTCAACAAGTACTGCGCGCTCACTCTTAAAGAGTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGC
GGGAATTACGTCACCTGTGCCCTTCGCCCTAGCCGCCTCCACCCATCATCATGAGCAAAGAGATT
CCCACGCCTTACATGTGGAGCTACCAGCCCCAGATGGGCCTGGCCGCCGGTGCCGCCCAGGACTA
CTCCACCCGCATGAATTGGCTCAGCGCCGGGCCCGCGATGATCTCACGGGTGAATGACATCCGCG
CCCACCGAAACCAGATACTCCTAGAACAGTCAGCGCTCACCGCCACGCCCCGCAATCACCTCAAT
CCGCGTAATTGGCCCGCCGCCCTGGTGTACCAGGAAATTCCCCAGCCCACGACCGTACTACTTCC
GCGAGACGCCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCC
TGTGTCGTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTC
AACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATC
GGGGAGATCTTCCTTCACGCCTCGTCAGGCCGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCC
GCTCGGGTGGCATCGGCACTCTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCC
TTCTCCGGCTCCCCCGGCCACTACCCGGACGAGTTCATCCCGAACTTCGACGCCATCAGCGAGTC
GGTGGACGGCTACGATTGAATGTCCCATGGTGGCGCAGCTGACCTAGCTCGGCTTCGACACCTGG
ACCACTGCCGCCGCTTCCGCTGCTTCGCTCGGGATCTCGCCGAGTTTGCCTACTTTGAGCTGCCC
GAGGAGCACCCTCAGGGCCCGGCCCACGGAGTGCGGATCGTCGTCGAAGGGGGCCTCGACTCCCA
CCTGCTTCGGATCTTCAGCCAGCGTCCGATCCTGGTCGAGCGCGAGCAAGGACAGACCCTTCTGA
CTCTGTACTGCATCTGCAACCACCCCGGCCTGCATGAAAGTCTTTGTTGTCTGCTGTGTACTGAG
TATAATAAAAGCTGAGATCAGCGACTACTCCGGACTTCCGTGTGTTCCTGAATCCATCAACCAGT

CTTTGTTCTTCACCGGGAACGAGACCGAGCTCCAGCTCCAGTGTAAGCCCCACAAGAAGTACCTC
ACCTGGCTGTTCCAGGGCTCCCCGATCGCCGTTGTCAACCACTGCGACAACGACGGAGTCCTGCT
GAGCGGCCCTGCCAACCTTACTTTTTCCACCCGCAGAAGCAAGCTCCAGCTCTTCCAACCCTTCC
TCCCCGGGACCTATCAGTGCGTCTCGGGACCCTGCCATCACACCTTCCACCTGATCCCGAATACC
ACAGCGTCGCTCCCCGCTACTAACAACCAAACTAACCTCCACCAACGCCACCGTCGCGACCTTTC
TGAATCTAATACTACCACCCACACCGGAGGTGAGCTCCGAGGTCAACCAACCTCTGGGATTTACT
ACGGCCCCTGGGAGGTGGTTGGGTTAATAGCGCTAGGCCTAGTTGCGGGTGGGCTTTTGGTTCTC
TGCTACCTATACCTCCCTTGCTGTTCGTACTTAGTGGTGCTGTGTTGCTGGTTTAAGAAATGGGG
AAGATCACCCTAGTGAGCTGCGGTGCGCTGGTGGCGGTGTTGCTTTCGATTGTGGGACTGGGCGG
TGCGGCTGTAGTGAAGGAGAAGGCCGATCCCTGCTTGCATTTCAATCCCAACAAATGCCAGCTGA
GTTTTCAGCCCGATGGCAATCGGTGCGCGGTACTGATCAAGTGCGGATGGGAATGCGAGAACGTG
AGAATCGAGTACAATAACAAGACTCGGAACAATACTCTCGCGTCCGTGTGGCAGCCCGGGGACCC
CGAGTGGTACACCGTCTCTGTCCCCGGTGCTGACGGCTCCCCGCGCACCGTGAATAATACTTTCA
TTTTTGCGCACATGTGCGACACGGTCATGTGGATGAGCAAGCAGTACGATATGTGGCCCCCCACG
AAGGAGAACATCGTGGTCTTCTCCATCGCTTACAGCCTGTGCACGGCGCTAATCACCGCTATCGT
GTGCCTGAGCATTCACATGCTCATCGCTATTCGCCCCAGAAATAATGCCGAAAAAGAAAAACAGC
CATAACGTTTTTTTTCACACCTTTTTCAGACCATGGCCTCTGTTAAATTTTTGCTTTTATTTGCC
AGTCTCATTGCCGTCATTCATGGAATGAGTAATGAGAAAATTACTATTTACACTGGCACTAATCA
CACATTGAAAGGTCCAGAAAAAGCCACAGAAGTTTCATGGTATTGTTATTTTAATGAATCAGATG
TATCTACTGAACTCTGTGGAAACAATAACAAAAAAAATGAGAGCATTACTCTCATCAAGTTTCAA
TGTGGATCTGACTTAACCCTAATTAACATCACTAGAGACTATGTAGGTATGTATTATGGAACTAC
AGCAGGCATTTCGGACATGGAATTTTATCAAGTTTCTGTGTCTGAACCCACCACGCCTAGAATGA
CCACAACCACAAAAACTACACCTGTTACCACTATGCAGCTCACTACCAATAACATTTTTGCCATG
CGTCAAATGGTCAACAATAGCACTCAACCCACCCCACCCAGTGAGGAAATTCCCAAATCCATGAT
TGGCATTATTGTTGCTGTAGTGGTGTGCATGTTGATCATCGCCTTGTGCATGGTGTACTATGCCT
TCTGCTACAGAAAGCACAGACTGAACGACAAGCTGGAACACTTACTAAGTGTTGAATTTTAATTT
TTTAGAACCATGAAGATCCTAGGCCTTTTAATTTTTTCTATCATTACCTCTGCTCTATGCAATTC
TGACAATGAGGACGTTACTGTCGTTGTCGGATCAAATTATACACTGAAAGGTCCAGCGAAGGGTA
TGCTTTCGTGGTATTGCTATTTTGGATCTGACACTACAGAAACTGAATTATGCAATCTTAAGAAT
GGCAAAATTCAAAATTCTAAAATTAACAATTATATATGCAATGGTACTGATCTGATACTCCTCAA
TATCACGAAATCATATGCTGGCAGTTACACCTGCCCTGGAGATGATGCTGACAGTATGATTTTTT
ACAAAGTAACTGTTGTTGATCCCACTACTCCACCTCCACCCACCACAACTACTCACACCACACAC
ACAGATCAAACCGCAGCAGAGGAGGCAGCAAAGTTAGCCTTGCAGGTCCAAGACAGTTCATTTGT
TGGCATTACCCCTACACCTGATCAGCGGTGTCCGGGGCTGCTAGTCAGCGGCATTGTCGGTGTGC
TTTCGGGATTAGCAGTCATAATCATCTGCATGTTCATTTTTGCTTGCTGCTATAGAAGGCTTTAC
CGACAAAAATCAGACCCACTGCTGAACCTCTATGTTTAATTTTTTCCAGAGTCATGAAGGCAGTT
AGCGCTCTAGTTTTTTGTTCTTTGATTGGCATTGTTTTTTGCAATCCTATTCCTAAAGTTAGCTT
TATTAAAGATGTGAATGTTACTGAGGGGGGCAATGTGACACTGGTAGGTGTAGAGGGTGCTGAAA
ACACCACCTGGACAAAATACCACCTCAATGGGTGGAAAGATATTTGCAATTGGAGTGTATTAGTT
TATACATGTGAGGGAGTTAATCTTACCATTGTCAATGCCACCTCAGCTCAAAATGGTAGAATTCA
AGGACAAAGTGTCAGTGTATCTAATGGGTATTTTACCCAACATACTTTTATCTATGACGTTAAAG
TCATACCACTGCCTACGCCTAGCCCACCTAGCACTACCACACAGACAACCCACACTACACAGACA
ACCACATACAGTACATTAAATCAGCCTACCACCACTACAGCAGCAGAGGTTGCCAGCTCGTCTGG
GGTCCGAGTGGCATTTTTGATGTGGGCCCCATCTAGCAGTCCCACTGCTAGTACCAATGAGCAGA
CTACTGAATTTTTGTCCACTGTCGAGAGCCACACCACAGCTACCTCCAGTGCCTTCTCTAGCACC
GCCAATCTCTCCTCGCTTTCCTCTACACCAATCAGTCCCGCTACTACTCCTAGCCCCGCTCCTCT
TCCCACTCCCCTGAAGCAAACAGACGGCGGCATGCAATGGCAGATCACCCTGCTCATTGTGATCG
GGTTGGTCATCCTGGCCGTGTTGCTCTACTACATCTTCTGCCGCCGCATTCCCAACGCGCACCGC
AAGCCGGTCTACAAGCCCATCATTGTCGGGCAGCCGGAGCCGCTTCAGGTGGAAGGGGGTCTAAG
GAATCTTCTCTTCTCTTTTACAGTATGGTGATTGAACTATGATTCCTAGACAATTCTTGATCACT
ATTCTTATCTGCCTCCTCCAAGTCTGTGCCACCCTCGCTCTGGTGGCCAACGCCAGTCCAGACTG
TATTGGGCCCTTCGCCTCCTACGTGCTCTTTGCCTTCACCACCTGCATCTGCTGCTGTAGCATAG
TCTGCCTGCTTATCACCTTCTTCCAGTTCATTGACTGGATCTTTGTGCGCATCGCCTACCTGCGC
CACCACCCCCAGTACCGCGACCAGCGAGTGGCGCGGCTGCTCAGGCTCCTCTGATAAGCATGCGG
GCTCTGCTACTTCTCGCGCTTCTGCTGTTAGTGCTCCCCCGTCCCGTCGACCCCCGGTCCCCCAC
CCAGTCCCCCGAGGAGGTCCGCAAATGCAAATTCCAAGAACCCTGGAAATTCCTCAAATGCTACC
GCCAAAAATCAGACATGCATCCCAGCTGGATCATGATCATTGGGATCGTGAACATTCTGGCCTGC

ACCCTCATCTCCTTTGTGATTTACCCCTGCTTTGACTTTGGTTGGAACTCGCCAGAGGCGCTCTA
TCTCCCGCCTGAACCTGACACACCACCACAGCAACCTCAGGCACACGCACTACCACCACTACAGC
CTAGGCCACAATACATGCCCATATTAGACTATGAGGCCGAGCCACAGCGACCCATGCTCCCCGCT
ATTAGTTACTTCAATCTAACCGGCGGAGATGACTGACCCACTGGCCAACAACAACGTCAACGACC
TTCTCCTGGACATGGACGGCCGCGCCTCGGAGCAGCGACTCGCCCAACTTCGCATTCGCCAGCAG
CAGGAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTGGCCATCCACCAGTGCAAGAGAGGCATCTT
CTGCCTGGTGAAACAGGCCAAGATCTCCTACGAGGTCACTCCAAACGACCATCGCCTCTCCTACG
AGCTCCTGCAGCAGCGCCAGAAGTTCACCTGCCTGGTCGGAGTCAACCCCATCGTCATCACCCAG
CAGTCTGGCGATACCAAGGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGTCCACACTCT
GATCAAGACCCTCTGCGGCCTCCGCGACCTCCTCCCCATGAACTAATCACCCCCTTATCCAGTGA
AATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAATAAAGA
TACAATCATATTGATGATTTGAGTTTAACAAAAAAATAAAGAATCACTTACTTGAAATCTGATAC
CAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGGTACTGCA
GGCCCCGGCGGGCTGCAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTGTCCCTCA
ATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTCGACCCC
GTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTCGTCTCTTC
AGATGGATTCCAAGAGAAGCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCACCACCA
AGAACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCTCGGGAAAACTCATC
TCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTAACAT
GGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACCATTAAATATAC
TGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACTCCGTGGCTCT
GCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATAAAGCTTACCTT
AGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTGGGCTAAAGGTT
TAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTTGGAAGCAGTAGT
ACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGGCCTTAGCTTTGA
CAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTGGACAACACCTG
ATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAACTAACACTTTGCTTGACTAAA
TGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAGGAAGTGGAAACCTAAACCCCAT
TACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACGGTGTTCTTTTAACAG
AACATTCTACACTAAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATGGCACTCCATAT
ACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCTTATCCAAAGTCACAAAGTTCTACTACTAA
AAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTATGCTTCTCACTATAA
CCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATACACCTGGACTAAT
GGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACATCGCCCAAGAATG
AACACTGTATCCCACCCTGCATGCCAACCCTTCCCACCCCACTCTGTGGAACAAACTCTGAAACA
CAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGATTCGAGCAGTTATTT
TTCCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCCGCACAGCCTTGAACATCTGA
ATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCAGAGCGAGCCAGTCT
CGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCT
GAGGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAAT
CATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCCCCGCAGCAGTCGCTGCCGCCGC
CGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCCACGGC
CCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATGCGGATCTCGCTCAGGTCGCTGC
AGTACGTGCAACACAGAACCACCAGGTTGTTCAACAGTCCATAGTTCAACACGCTCCAGCCGAAA
CTCATCGCGGGAAGGATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTAAATCAAGTGGTG
CCCCCTCCAGAACACGCTGCCCACGTACATGATCTCCTTGGGCATGTGGCGGTTCACCACCTCCC
GGTACCACATCACCCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCGGAACCACAGGGCCAGC
ACCGCCCCGCCCGCCATGCAGCGAAGAGACCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCG
CTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCA
TGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCATATCCCAGGGCACGGGGAACTCT
TGCAGGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAG
GGTATCGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCAC
AGCGTGGTAAGGGGGCCGGCCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTCGCGACCGT
GTCATGATGCAGTTGCTTTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGC
ACACCGATCGCCGGCGGCGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAAAACAGCCAC
TCTCTCAGACCGTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGC
TCTGATCACATCGACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTT

CGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCAAACGGTCTCGG
AGTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAAC
AGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGC
GCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATCATGTTA
CACTCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGAACTAGTTCcTG
AGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGC
ACACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAGATTGACAAGCGGAATAT
CAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCC
TCTCCGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAA
CCGAAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGG
TGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGAA
AAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCGTTC
CAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCTGGC
GAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGGCGCGACCCTCG
TAAAAATTGTCGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGTGAATGAT
TCGACAAGATGAATACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAAGCGCCCGAGGAAGC
AATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAA
TTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGCCCCCGATCCCTC
CAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAGA
GTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCCAGATCTAC
ACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCAGCACACGCCCA
GAAACCGGTGACACACTCAAAAAAATACGCGCACTTCCTCAAACGCCCAAAACTGCCGTCATTTC
CGGGTTCCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTAAAAACGTCACC
CGCCCCGCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTCAAACA
CCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATTGATGATGG ATCC VR-594
C68 (SEQ ID NO: 10); Independently sequenced; Full-Length C68
CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTT
GGGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGTTTT
GATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGG
TGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTTTCTGG
GCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAATCTGAGTA
ATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGATTACGTGG
GGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGTGTTTTTAC
GTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCCACTCTTGAG
TGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAAGATGAGGCAC
CTGAGAGACCTGCCCGATGAGAAAATCATCATCGCTTCCGGGAACGAGATTCTGGAACTGGTGGT
AAATGCCATGATGGGCGACGACCCTCCGGAGCCCCCCACCCCATTTGAGACACCTTCGCTGCACG
ATTTGTATGATCTGGAGGTGGATGTGCCCGAGGACGATCCCAATGAGGAGGCGGTAAATGATTTT
TTTAGCGATGCCGCGCTGCTAGCTGCCGAGGAGGCTTCGAGCTCTAGCTCAGACAGCGACTCTTC
ACTGCATACCCCTAGACCCGGCAGAGGTGAGAAAAAGATCCCCGAGCTTAAAGGGGAAGAGATGG
ACTTGCGCTGCTATGAGGAATGCTTGCCCCCGAGCGATGATGAGGACGAGCAGGCGATCCAGAAC
GCAGCGAGCCAGGGAGTGCAAGCCGCCAGCGAGAGCTTTGCGCTGGACTGCCCGCCTCTGCCCGG
ACACGGCTGTAAGTCTTGTGAATTTCATCGCATGAATACTGGAGATAAAGCTGTGTTGTGTGCAC
TTTGCTATATGAGAGCTTACAACCATTGTGTTTACAGTAAGTGTGATTAAGTTGAACTTTAGAGG
GAGGCAGAGAGCAGGGTGACTGGGCGATGACTGGTTTATTTATGTATATATGTTCTTTATATAGG
TCCCGTCTCTGACGCAGATGATGAGACCCCCACTACAAAGTCCACTTCGTCACCCCCAGAAATTG
GCACATCTCCACCTGAGAATATTGTTAGACCAGTTCCTGTTAGAGCCACTGGGAGGAGAGCAGCT
GTGGAATGTTTGGATGACTTGCTACAGGGTGGGGTTGAACCTTTGGACTTGTGTACCCGGAAACG
CCCCAGGCACTAAGTGCCACACATGTGTGTTTACTTGAGGTGATGTCAGTATTTATAGGGTGTGG
AGTGCAATAAAAAATGTGTTGACTTTAAGTGCGTGGTTTATGACTCAGGGGTGGGGACTGTGAGT
ATATAAGCAGGTGCAGACCTGTGTGGTTAGCTCAGAGCGGCATGGAGATTTGGACGGTCTTGGAA
GACTTTCACAAGACTAGACAGCTGCTAGAGAACGCCTCGAACGGAGTCTCTTACCTGTGGAGATT
CTGCTTCGGTGGCGACCTAGCTAGGCTAGTCTACAGGGCCAAACAGGATTATAGTGAACAATTTG
AGGTTATTTTGAGAGAGTGTTCTGGTCTTTTTGACGCTCTTAACTTGGGCCATCAGTCTCACTTT
AACCAGAGGATTTCGAGAGCCCTTGATTTTACTACTCCTGGCAGAACCACTGCAGCAGTAGCCTT
TTTTGCTTTTATTCTTGACAAATGGAGTCAAGAAACCCATTTCAGCAGGGATTACCAGCTGGATT
TCTTAGCAGTAGCTTTGTGGAGAACATGGAAGTGCCAGCGCCTGAATGCAATCTCCGGCTACTTG

CCGGTACAGCCGCTAGACACTCTGAGGATCCTGAATCTCCAGGAGAGTCCCAGGGCACGCCAACG
TCGCCAGCAGCAGCAGCAGGAGGAGGATCAAGAAGAGAACCCGAGAGCCGGCCTGGACCCTCCGG
CGGAGGAGGAGGAGTAGCTGACCTGTTTCCTGAACTGCGCCGGGTGCTGACTAGGTCTTCGAGTG
GTCGGGAGAGGGGGATTAAGCGGGAGAGGCATGATGAGACTAATCACAGAACTGAACTGACTGTG
GGTCTGATGAGTCGCAAGCGCCCAGAAACAGTGTGGTGGCATGAGGTGCAGTCGACTGGCACAGA
TGAGGTGTCGGTGATGCATGAGAGGTTTTCTCTAGAACAAGTCAAGACTTGTTGGTTAGAGCCTG
AGGATGATTGGGAGGTAGCCATCAGGAATTATGCCAAGCTGGCTCTGAGGCCAGACAAGAAGTAC
AAGATTACTAAGCTGATAAATATCAGAAATGCCTGCTACATCTCAGGGAATGGGGCTGAAGTGGA
GATCTGTCTCCAGGAAAGGGTGGCTTTCAGATGCTGCATGATGAATATGTACCCGGGAGTGGTGG
GCATGGATGGGGTTACCTTTATGAACATGAGGTTCAGGGGAGATGGGTATAATGGCACGGTCTTT
ATGGCCAATACCAAGCTGACAGTCCATGGCTGCTCCTTCTTTGGGTTTAATAACACCTGCATCGA
GGCCTGGGGTCAGGTCGGTGTGAGGGGCTGCAGTTTTTCAGCCAACTGGATGGGGGTCGTGGGCA
GGACCAAGAGTATGCTGTCCGTGAAGAAATGCTTGTTTGAGAGGTGCCACCTGGGGGTGATGAGC
GAGGGCGAAGCCAGAATCCGCCACTGCGCCTCTACCGAGACGGGCTGCTTTGTGCTGTGCAAGGG
CAATGCTAAGATCAAGCATAATATGATCTGTGGAGCCTCGGACGAGCGCGGCTACCAGATGCTGA
CCTGCGCCGGCGGGAACAGCCATATGCTGGCCACCGTACATGTGGCTTCCCATGCTCGCAAGCCC
TGGCCCGAGTTCGAGCACAATGTCATGACCAGGTGCAATATGCATCTGGGGTCCCGCCGAGGCAT
GTTCATGCCCTACCAGTGCAACCTGAATTATGTGAAGGTGCTGCTGGAGCCCGATGCCATGTCCA
GAGTGAGCCTGACGGGGGTGTTTGACATGAATGTGGAGGTGTGGAAGATTCTGAGATATGATGAA
TCCAAGACCAGGTGCCGAGCCTGCGAGTGCGGAGGGAAGCATGCCAGGTTCCAGCCCGTGTGTGT
GGATGTGACGGAGGACCTGCGACCCGATCATTTGGTGTTGCCCTGCACCGGGACGGAGTTCGGTT
CCAGCGGGGAAGAATCTGACTAGAGTGAGTAGTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGC
CAGAATAACTGAAATCTGTGCTTTTCTGTGTGTTGCAGCAGCATGAGCGGAAGCGGCTCCTTTGA
GGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCCTCCTGGGCGGGAGTGCGTCAGAATG
TGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCGCGAACTCTTCAACCCTGACCTATGCA
ACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGCAGCTGCTGCATCTGCCGCCAGCGCCGT
GCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACTCGAGTTCCACCAATA
ATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCCAGCTCGAGGCCTTGACCCAG
CGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCAC
GGTGAAATCCAAATAAAAAATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAACACAGAG
TCTGAATCTTTATTTGATTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCA
CCCGGTGGATCTTTTCCAGGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGC
CCGTCCCGGGGGTGGAGGTAGCTCCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCAC
CCAGTCATAGCAGGGGCGCAGGGCATGGTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCA
CGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAG
ATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATGTTACCGCCCAGATCCCGCCTGGGGTT
CATGTTGTGCAGGACCACCAGCACGGTGTATCCGGTGCACTTGGGGAATTTATCATGCAACTTGG
AAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCGCCCAGGTTTTCCATGCACTCATCC
ATGATGATGGCGATGGGCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATC
ATAGTTGTGGTCCTGGGTGAGGTCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGG
ACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAG
GCTTTGAGCTCGGAGGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGC
GGGGGAGATGAGCTGGGCCGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGC
CGTAGATGACCCCGATGACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGG
AGGAGGGGGGCCACCTCGTTCATCATCTCGCGCACGTGCATGTTCTCGCGCACCAGTTCCGCCAG
GAGGCGCTCTCCCCCCAGGGATAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTC
CGTCGGCCATGGGCATTTTGGAGAGGGTTTGTTGCAAGAGTTCCAGGCGGTCCCAGAGCTCGGTG
ATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGACGGCTGCGGGAG
TAGGGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCG
CGTCAGGGTGGTCTCCGTCACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTGCGCT
TCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAGGTAGCAA
TTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGA
AGTCTGCCCGCAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGG
ACTCGGGGGCGTAGGCGTCCGCGCCGCAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTG
AGGTCGGGCTGGTCGGGGTCAAAAACCAGTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTT
GGTCTCCATGAGCTCGTGTCCCCGCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACT
TTATGGGCCGGTCCTCGAGCGGTGTGCCGCGGTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAG

ACGAAAGCCCGGGTCCAGGCCAGCACGAAGGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCAC
CAGCGGGTCCACCTTTTCCAGGGTATGCAAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGA
TTGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCCGGCCGGGGGGGGTATAAAAGGGTGCGGGT
CCCTGCTCGTCCTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTC
CCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGA
TATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAGACGATC
TTTTTGTTGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGA
GCGCATGGTCTGGTTTTTTTCCTTGTCGGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACT
CGCGCGCCACGCACTTCCATTCGGGGAAGACGGTGGTCAGCTCGTCGGGCACGATTCTGACCTGC
CAGCCCCGATTATGCAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATT
AGTCCAGCAGAGGCGTCCGCCCTTGCGCGAGCAGAAGGGGGGCAGGGGGTCCAGCATGACCTCGT
CGGGGGGGTCGGCATCGATGGTGAAGATGCCGGGCAGGAGGTCGGGGTCAAAGTAGCTGATGGAA
GTGGCCAGATCGTCCAGGGCAGCTTGCCATTCGCGCACGGCCAGCGCGCGCTCGTAGGGACTGAG
GGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGAGGCGTACATGCCGCAGATGTCGTAGACGT
AGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGC
ACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTT
TTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGA
AGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGC
AGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGAT
GATGTCATACTTGAGCTGTCCCTTTTGTTTCCACAGCTCGCGGTTGAGAAGGAACTCTTCGCGGT
CCTTCCAGTACTCTTCGAGGGGGAACCCGTCCTGATCTGCACGGTAAGAGCCTAGCATGTAGAAC
TGGTTGACGGCCTTGTAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTT
GCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGTCCCTGACCATGACCTTGAGGAACTGGTGCTTGA
AGTCGATATCGTCGCAGCCCCCCTGCTCCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGG
TTGGGCAAAGCGAAAGTAACATCGTTGAAGAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGT
GATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTGATGACCTGGGCGGCGAGCACGATCTCGT
CGAAGCCGTTGATGTTGTGGCCCACGATGTAGAGTTCCACGAATCGCGGACGGCCCTTGACGTGG
GGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGGGGTCGCTGAGCCCGTGCTGCTCGAGCGC
CCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCCAGGGCGGTTT
GCAGACGGTCCCGGTACTGACGGAACTGCTGCCCGACGGCCATTTTTTCGGGGGTGACGCAGTAG
AAGGTGCGGGGGTCCCCGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTC
GACGAGCCGGTCGTCCCCGGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGG
ACCCCATCCAGGTGTAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAG
CCGATGGGGAAGAACTGGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTA
GAAATGCCGACGGCGCGCCGAACACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAAC
GCTGCACGGGATGCACGTGCTGCACGAGCTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGGG
AAGTGGAGTCGTGGCGCCTGCATCTCGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTCTTC
TGCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGGCGCGAGCGG
GTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTC
AGGTCAGTGGGCAGCGGCGGCGCGCGGTTGACTTGCAGGAGTTTTTCCAGGGCGCGCGGGAGGTC
CAGATGGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTCGATGGCTTGCAGGGTCCCGTGCC
CCTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGCGACGGGGGCGGTGCCTCT
TCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCCGGGCGGCAGGGGCGGCTCGGGGCCCGGAGGC
AGGGGCGGCAGGGGCACGTCGGCGCCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACT
GGCGTGAGCGACGACGCGACGGTTGACGTCCTGGATCTGACGCCTCTGGGTGAAGGCCACGGGAC
CCGTGAGTTTGAACCTGAAAGAGAGTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGC
CGCAGGATCTCTTGCACGTCGCCCGAGTTGTCCTGGTAGGCGATCTCGGTCATGAACTGCTCGAT
CTCCTCCTCTTGAAGGTCTCCGCGGCCGGCGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGC
GGCCCATGAGCTGCGAGAAGGCGTTCATGCCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACG
CCCTCGGGATCGCgGGCGCGCATGACCACCTGGGCGAGGTTGAGCTCCACGTGGCGCGTGAAGAC
CGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGCGATGTGCTCGGTGACGAAGA
AATACATGATCCAGCGGCGGAGCGGCATCTCGCTGACGTCGCCCAGCGCCTCCAAACGTTCCATG
GCCTCGTAAAAGTCCACGGCGAAGTTGAAAAACTGGGAGTTGCGCGCCGAGACGGTCAACTCCTC
CTCCAGAAGACGGATGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCCCGGGAGTT
CCTCCACTTCCTCTTCTTCCTCCTCCACTAACATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGT
GGCGGGGGAGGGGGCCTGCGTCGCCGGCGGCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGT
CTCGCCGCGCCGGCGTCGCATGGTCTCGGTGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGA

AGACGCCGCCGCGCATCTCCAGGTGGCCGGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACG
ATGCATCTTATCAATTGCCCCGTAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGG
ATCTGAAAACCGCTGAACGAAGGCTTCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTT
CTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGGGGGCGATGCTGCTGGTGATGAAGTTGAAATAG
GCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCTTGCTGGATGCG
CAGACGGTCGGCCATGCCCCAGGCGTGGTCCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGCA
TGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCGGCCGTGCATGCGCGTGAGCCCGAAGCCG
CGCTGGGGCTGGACGAGCGCCAGGTCGGCGACGACGCGCTCGGCGAGGATGGCTTGCTGGATCTG
GGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGTGT
AGGAGCAGTTGGCCATGACGGACCAGTTGACGGTCTGGTGGCCCGGACGCACGAGCTCGTGGTAC
TTGAGGCGCGAGTAGGCGCGCGTGTCGAAGATGTAGTCGTTGCAGGTGCGCACCAGGTACTGGTA
GCCGATGAGGAAGTGCGGCGGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGG
GCGCGAGGTCCTCGAGCATGGTGCGGTGGTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCG
GCGGCGGTGGTGGAGGCGCGCGGGAACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAA
GTAGTTCATGGTGGGCACGGTCTGGCCCGTGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCA
AAAACGAAAGCGGTCAGCGGCTCGACTCCGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGC
GTGTACCCCGGTTCGAATCTCGAATCAGGCTGGAGCCGCAGCTAACGTGGTATTGGCACTCCCGT
CTCGACCCAAGCCTGCACCAACCCTCCAGGATACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGG
AGGCCGGATGAGACTAGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTGGA
GAAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGGTTCGAGGCCGGCCGGATTCCGCGGCTAAC
GAGGGCGTGGCTGCCCCGTCGTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCG
AGCCCCTCTTTTGTTTTGTTTGTTTTTGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCAC
CACCCTCCACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACT
TCCAGCCACGACCGCCGCGGCCGCCGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCT
TGGAAGAGGGCGAGGGGCTGGCGCGCCTGGGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAG
ATGAAAAGGGACGCTCGCGAGGCCTACGTGCCCAAGCAGAACCTGTTCAGAGACAGGAGCGGCGA
GGAGCCCGAGGAGATGCGCGCGGCCCGGTTCCACGCGGGGGGGGAGCTGCGGCGCGGCCTGGACC
GAAAGAGGGTGCTGAGGGACGAGGATTTCGAGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGC
GCGCACGTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTT
CCAAAAATCCTTCAACAACCACGTGCGCACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGA
TGCACCTGTGGGACCTGCTGGAGGCCATCGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGCAG
CTGTTCCTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTCAGGGAGGCGCTGCTGAATATCAC
CGAGCCCGAGGGCCGCTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGC
GCGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTAC
TACGCTAGGAAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTT
TTACATGCGCATGACCCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACA
GGATGCACCGTGCGGTGAGCGCCAGCAGGCGGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGT
CTGCAGCGGGCCCTGACCGGGGCCGGGACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCT
GCACTGGCAGCCCAGCCGCCGGGCCTTGGAGGCGGCGGCAGGACCCTACGTAGAAGAGGTGGACG
ATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGC
AACAACAACAGCCACCTCCTGATCCCGCGATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATT
AACTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGACGACCCGCAACCCCGA
AGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGC
GCTCCAACCCCACGCACGAGAAGGTCCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATC
CGCGGCGACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCAC
CAACGTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGC
GGTTCCACCGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCC
GCCAACGTGCCCCGGGGCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGAC
CGAGGTGCCCCAGAGCGAGGTGTACCAGTCCGGGCCGGACTACTTCTTCCAGACCAGTCGCCAGG
GCTTGCAGACCGTGAACCTGAGCCAGGCTTTCAAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCC
CCGGTCGGGGACCGCGCGACGGTGTCGAGCCTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCT
GGTGGCCCCCTTCACGGACAGCGGCAGCATCAACCGCAACTCGTACCTGGGCTACCTGATTAACC
TGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACCTACCAGGAGATCACCCACGTG
AGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAA
CCGGTCGCAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTGCGTTACG
TGCAGCAGAGCGTGGGCCTGTTCCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATG
ACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAACTGATGGA

CTACTTGCATCGGGCGGCCGCCATGAACTCTGACTATTTCACCAACGCCATCCTGAATCCCCACT
GGCTCCCGCCGCCGGGGTTCTACACGGGCGAGTACGACATGCCCGACCCCAATGACGGGTTCCTG
TGGGACGATGTGGACAGCAGCGTGTTCTCCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAA
GAAGGAAGGCAGCGACCGACGCCCGTCCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGG
TGCCCGAGGCCGCCAGTCCTTTCCCGAGCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAG
CTGGGCAGGATCACGCGCCCGCGCTTGCTGGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAG
ACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAAAGCCTGGTGGACAAGATGAGCCGCT
GGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTCGCAGGGGGCCACGAGCCGGGGC
AGCGCCGCCCGTAAACGCCGGTGGCACGACAGGCAGCGGGGACAGATGTGGGACGATGAGGACTC
CGCCGACGACAGCAGCGTGTTGGACTTGGGTGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCC
GTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACCAG
CGTGCGTTCGTTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGT
CCTCCTCCCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCT
GGAGGCTCCTTACGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACT
CGGAGCTGGCACCCTTGTACGATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATC
GCCTCGCTGAACTACCAGAACGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTT
CACCCCCACGGAGGCCAGCACCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGC
TGAAAACCATCATGCACACCAACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCAAG
GCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGACAGTGACAGAGGATTATGATGGTAG
TCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCA
TGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGGCGGCAGAACGGG
GTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACTAGGAACTTCAGGCTGGGCTGGGACCCCGT
GACCGAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGC
CCGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCAG
CCCTTCCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCT
GGATGTCGACGCCTATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTA
CCGCCTCTACCGAGGTCAGGGGCGATAATTTTGCAAGCGCCGCAGCAGTGGCAGCGGCCGAGGCG
GCTGAAACCGAAAGTAAGATAGTCATTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAA
CGTACTACCGGACAAGATAAACACCGCCTACCGCAGCTGGTACCTAGCCTACAACTATGGCGACC
CCGAGAAGGGCGTGCGCTCCTGGACGCTGCTCACCACCTCGGACGTCACCTGCGGCGTGGAGCAA
GTCTACTGGTCGCTGCCCGACATGATGCAAGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAG
CAACTACCCGGTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGG
CCGTCTACTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAG
AACCAGATCCTCGTCCGCCCGCCCGCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCT
CACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTG
ACGCCAGACGCCGCACCTGCCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTC
TCGAGCCGCACCTTCTAAATGTCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCG
CGCGCCCAGCAAGATGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCGCGTGCGCG
GGCACTTCCGCGCTCCCTGGGGCGCCCTCAAGGGCCGCGTGCGGTCGCGCACCACCGTCGACGAC
GTGATCGACCAGGTGGTGGCCGACGCGCGCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGT
GGACGCCGTCATCGACAGCGTGGTGGCcGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGC
GGCGCATCGCCCGGCGGCACCGGAGCACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGG
GCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGC
CGGCAGGACCCGGAGACGCGCGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGC
GGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGC
CCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCC
AAGCGCAAATTCAAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGT
GAAGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTG
ATGTGGACGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCCGGCGGCGCGTGCAGTGGCGCGGG
CGGAAGGTGCAACCGGTGCTGAGACCCGGCACCACCGTGGTCTTCACGCCCGGCGAGCGCTCCGG
CACCGCTTCCAAGCGCTCCTACGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCCG
AGCGCCTGGGCGAGTTTGCTTACGGCAAGCGCAGCCGTTCCGCACCGAAGGAAGAGGCGGTGTCC
ATCCCGCTGGACCACGGCAACCCCACGCCGAGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCC
GACCGCGGCGCCGCGCCGGGGGTTCAAGCGCGAGGGCGAGGATCTGTACCCCACCATGCAGCTGA
TGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGAAGGTGGACCCGGACGTGCAG
CCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACAT
CAAGATTCCCACGGAGCCCATGGAAACGCAGACCGAGCCCATGATCAAGCCCAGCACCAGCACCA

TGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACGGC
GCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCG
CGGCACGCGCTTCTACCGCGGTCATACCAGCAGCCGCCGCCGCAAGACCACCACTCGCCGCCGCC
GTCGCCGCACCGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGC
GCACCTCTGACCCTGCCGCGCGCGCGCTACCACCCGAGCATCGCCATTTAAACTTTCGCCtGCTT
TGCAGATCAATGGCCCTCACATGCCGCCTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACC
GCGCCGTAGAAGGCTGGCGGGGAACGGGATGCGTCGCCACCACCACCGGCGGCGGCGCGCCATCA
GCAAGCGGTTGGGGGGAGGCTTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGATCGGGGCG
ATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACTTGGAAACAT
CTTGTAATAAACCaATGGACTCTGACGCTCCTGGTCCTGTGATGTGTTTTCGTAGACAGATGGAA
GACATCAATTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTCATGGGCACCTGGAGCGA
CATCGGCACCAGCCAACTGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGA
ATTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACCACAGGGCAGGCGCTG
AGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAA
CGGGGTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGC
CGCCCGCCGGCTCCGTGGAGATGCCGCAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGC
GAGAAGCGACCCCGCCCCGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCCCGTA
CGAGGAGGCGGTGAAACTGGGTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGC
TGAAACCCGAAAAGCCCGCGACCCTGGACTTGCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTG
GCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCGCGCGCGACCCGGGGGCACCGCCCGCCCTCATGC
GAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCT
GCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTGTATGTATTATGTCGCCGCCGC
CGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGATGGCCACCCCATCGA
TGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGTCCGGGT
CTGGTGCAGTTTGCCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCAC
GGTGGCGCCCACGCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCG
TGGACCGCGAGGACAACACCTACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCGC
GTGCTGGACATGGCCAGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAA
ACCCTACTCCGGCACCGCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGA
CATATAAAGCCGATGGTGAAACTGCCACAGAAAAAACCTATACATATGGAAATGCACCCGTGCAG
GGCATTAACATCACAAAAGATGGTATTCAACTTGGAACTGACACCGATGATCAGCCAATCTACGC
AGATAAAACCTATCAGCCTGAACCTCAAGTGGGTGATGCTGAATGGCATGACATCACTGGTACTG
ATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTTTT
GCCAAGCCTACTAATAAAGAAGGAGGTCAGGCAAATGTGAAAACAGGAACAGGCACTACTAAAGA
ATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAA
TTGTTTTGTATACTGAAAATGTGGATTTGGAAACTCCAGATACCCATATTGTATACAAAGCAGGC
ACAGATGACAGCAGCTCTTCTATTAATTTGGGTCAGCAAGCCATGCCCAACAGACCTAACTACAT
TGGTTTCAGAGACAACTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGG
CCGGTCAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCCTAC
CAGCTCTTGCTTGACTCTCTGGGTGACAGAACCCGGTATTTCAGTATGTGGAATCAGGCGGTGGA
CAGCTATGATCCTGATGTGCGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATT
GTTTCCCTCTGGATGCTGTTGGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACTGAT
CAAACCACATGGACCAAAGATGACAGTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCATT
CGCCATGGAAATCAACATCCAAGCCAACCTGTGGAGGAACTTCCTCTACGCCAACGTGGCCCTGT
ACCTGCCCGACTCTTACAAGTACACGCCGGCCAATGTTACCCTGCCCACCAACACCAACACCTAC
GATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTACATCAACATCGGGGCGCG
CTGGTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATGCGGGGCTGCGCT
ACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTT
TTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGA
CGTCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCT
TCACCAGCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAG
GCCATGCTGCGCAACGACACCAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCT
CTACCCCATCCCGGCCAACGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCT
TCCGCGGCTGGTCCTTCACGCGTCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGAC
CCCTACTTCGTCTACTCGGGCTCCATCCCCTACCTCGACGGCACCTTCTACCTCAACCACACCTT
CAAGAAGGTCTCCATCACCTTCGACTCCTCCGTCAGCTGGCCCGGCAACGACCGGCTCCTGACGC
CCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAGGGCTACAACGTGGCCCAGTGCAACATG

ACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCACTACAACATCGGCTACCAGGGCTTCTACGT
GCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCATGAGCCGCCAGG
TGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTCACCCTGGCCTACCAGCACAACAACTCG
GGCTTCGTCGGCTACCTCGCGCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTA
CCCGCTCATCGGCAAGAGCGCCGTCACCAGCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCA
TGTGGCGCATCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAAC
ATGCTCTATGCCAACTCCGCCCACGCGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTC
CACCCTTCTCTATGTTGTCTTCGAAGTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCG
TCATCGAGGCCGTCTACCTGCGCACCCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCT
TCTTGCAAGCCATGGCCGCGGGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGC
TGCGGGCCCTACTTCCTGGGCACCTTCGATAAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCT
GGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCT
GGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAG
CAGATCTACCAGTTCGAGTACGAGGGCCTGCTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTG
CGTCACCCTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCT
GCTGCATGTTCCTGCACGCCTTCGTGCACTGGCCCGACCGCCCCATGGACAAGAACCCCACCATG
AACTTGCTGACGGGGGTGCCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCG
CAACCAGGAGGCGCTCTACCGCTTCCTCAACTCCCACTCCGCCTACTTTCGCTCCCACCGCGCGC
GCATCGAGAAGGCCACCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAA
ATGTCTTTAATAAACAGCACTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAA
AGGGTTCTGCCGGGTCTCGGCATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTACTTGGCCA
GCCACTTGAACTCGGGGATCAGCAGTTTGGGCAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGC
TTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACC
CGCGTTCTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAGCACTGGAACACCATCAGGGCCGGGT
GCTTCACGCTCGCCAGCACCGTCGCGTCGGTGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCC
ATCCCGAAGGGGGTCATCTTGCAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCTTGTGGTT
GCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACATGG
CCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACC
CCGCAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTCGTGCACGCAGCAGCGCGCGTC
GTTGTTGGCCAGCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTCGGGGT
TCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGG
ATCATGGTGGTCCCGTGCAGGCACCGCAGCTTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAG
CGCGCACCCGGTGCACTCCCAGTTCTTGTGGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCA
GGAAGCGGCCCATCATGGTGGTCAGGGTCTTGTTGCTAGTGAAGGTCAGCGGAATGCCGCGGTGC
TCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACCTCGCCCTGCTCGGGCATCAGCTGGAA
GTTGGCTTTCAGGTCGGTCTCCACGCGGTAGCGGTCCATCAGCATAGTCATGATTTCCATACCCT
TCTCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTTCTTCACCATCATCTTAGCGCTAGCAGCC
GCGGCCAGGGGGTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCG
CACCGGGGGGTAGCTGAAGCCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGT
CCTGGCTGACGTCCTGCAGGACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGC
GGCGGAGATGTTGGAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTATCTCTTCCTC
TTCTTGGTCCGAGGCCACGCGGCGGTAGGTATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGC
TCTCGCCGCCGCGACTTGGCGGATGGCTGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGG
CGGCGCTCTGACTGACTTCCTCCGCGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCAT
GGAGACTCAGCCATCGCCAACCTCGCCATCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGA
ATGAAAGCTTAACCGCCCCGCCGCCCAGCCCCGCCACCTCCGACGCGGCCGTCCCAGACATGCAA
GAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGACGCCCGCGGAGCACGAGGAGGAGCT
GGCAGTGCGCTTTTCACAAGAAGAGATACACCAAGAACAGCCAGAGCAGGAAGCAGAGAATGAGC
AGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGGGAGGACGCGCTC
ATCAAGCATCTGGCCCGGCAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCGAGGTGCC
CCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCA
AGCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTG
CCCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGC
CAACCGCACCCGCGCCGACGCCCTTTTCAACCTGGGTCCCGGCGCCCGCCTACCTGATATCGCCT
CCTTGGAAGAGGTTCCCAAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCT
CTGCAAGGAGAAGGAGGAGAGCATGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGC
GCGGCTGGCGGTGCTCAAACGCACGGTCGAGCTGACCCATTTCGCCTACCCGGCTCTGAACCTGC

CCCCCAAAGTCATGAGCGCGGTCATGGACCAGGTGCTCATCAAGCGCGCGTCGCCCATCTCCGAG
GACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTG
GCTGGGTCCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGCGCAAACTCATGATGGCCGTGGTCC
TGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGCGGAGACCCTGCGCAAGGTC
GAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCAACGT
GGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGTGC
TGCACACCACCCTGCGCGGGGAGGCCCGGCGCGACTACATCCGCGACTGCGTCTACCTCTACCTC
TGCCACACCTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGA
GCTCTGCAAGCTCCTGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCG
CCTCGGACCTGGCCGACCTCATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCC
GACTTTATGAGCCAAAGCATGTTGCAAAACTTTCGCTCTTTCATCCTCGAACGCTCCGGAATCCT
GCCCGCCACCTGCTCCGCGCTGCCCTCGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGC
CGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATC
GAGGACGTCAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACCG
CTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGCAAG
GGCCCAGCGAAGGCGAGGGTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTGTGGACC
TCGGCCTACTTGCGCAAGTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGAGGA
CCAATCCCATCCGCCCAAGGCCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCC
AATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTC
GACCCCCAGACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGC
TGAAAGTGGAGCTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAG
GAGGAGGAGATGGAGGAAGACTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCT
GGAGGAAGACGAGGAGGAGGCAGAGGAGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCT
CGGCGGGGGAGAAAGCAAGCAGCACGGATACCATCTCCGCTCCGGGTCGGGGTCCCGCTCGACCA
CACAGTAGATGGGACGAGACCGGACGATTCCCGAACCCCACCACCCAGACCGGTAAGAAGGAGCG
GCAGGGATACAAGTCCTGGCGGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGG
GCAACATCTCCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAACTTTCCCCGCAACATC
TTGCATTACTACCGTCACCTCCACAGCCCCTACTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAA
AGACCAGCAGAAAACCAGCAGCTAGAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGATCGCG
GCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCCACCCTCTATGCCATCTT
CCAGCAGAGTCGGGGGCAGGAGCAGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCC
GCAGTTGTCTGTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTC
TTCAACAAGTACTGCGCGCTCACTCTTAAAGAGTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGC
GGGAATTACGTCACCTGTGCCCTTCGCCCTAGCCGCCTCCACCCATCATCATGAGCAAAGAGATT
CCCACGCCTTACATGTGGAGCTACCAGCCCCAGATGGGCCTGGCCGCCGGTGCCGCCCAGGACTA
CTCCACCCGCATGAATTGGCTCAGCGCCGGGCCCGCGATGATCTCACGGGTGAATGACATCCGCG
CCCACCGAAACCAGATACTCCTAGAACAGTCAGCGCTCACCGCCACGCCCCGCAATCACCTCAAT
CCGCGTAATTGGCCCGCCGCCCTGGTGTACCAGGAAATTCCCCAGCCCACGACCGTACTACTTCC
GCGAGACGCCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCC
TGTGTCGTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTC
AACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATC
GGGGAGATCTTCCTTCACGCCTCGTCAGGCCGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCC
GCTCGGGTGGCATCGGCACTCTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCC
TTCTCCGGCTCCCCCGGCCACTACCCGGACGAGTTCATCCCGAACTTCGACGCCATCAGCGAGTC
GGTGGACGGCTACGATTGAATGTCCCATGGTGGCGCAGCTGACCTAGCTCGGCTTCGACACCTGG
ACCACTGCCGCCGCTTCCGCTGCTTCGCTCGGGATCTCGCCGAGTTTGCCTACTTTGAGCTGCCC
GAGGAGCACCCTCAGGGCCCGGCCCACGGAGTGCGGATCGTCGTCGAAGGGGGCCTCGACTCCCA
CCTGCTTCGGATCTTCAGCCAGCGTCCGATCCTGGTCGAGCGCGAGCAAGGACAGACCCTTCTGA
CTCTGTACTGCATCTGCAACCACCCCGGCCTGCATGAAAGTCTTTGTTGTCTGCTGTGTACTGAG
TATAATAAAAGCTGAGATCAGCGACTACTCCGGACTTCCGTGTGTTCCTGAATCCATCAACCAGT
CTTTGTTCTTCACCGGGAACGAGACCGAGCTCCAGCTCCAGTGTAAGCCCCACAAGAAGTACCTC
ACCTGGCTGTTCCAGGGCTCCCCGATCGCCGTTGTCAACCACTGCGACAACGACGGAGTCCTGCT
GAGCGGCCCTGCCAACCTTACTTTTTCCACCCGCAGAAGCAAGCTCCAGCTCTTCCAACCCTTCC
TCCCCGGGACCTATCAGTGCGTCTCGGGACCCTGCCATCACACCTTCCACCTGATCCCGAATACC
ACAGCGTCGCTCCCCGCTACTAACAACCAAACTAACCTCCACCAACGCCACCGTCGCGACCTTTC
TGAATCTAATACTACCACCCACACCGGAGGTGAGCTCCGAGGTCAACCAACCTCTGGGATTTACT
ACGGCCCCTGGGAGGTGGTTGGGTTAATAGCGCTAGGCCTAGTTGCGGGTGGGCTTTTGGTTCTC

TGCTACCTATACCTCCCTTGCTGTTCGTACTTAGTGGTGCTGTGTTGCTGGTTTAAGAAATGGGG
AAGATCACCCTAGTGAGCTGCGGTGCGCTGGTGGCGGTGTTGCTTTCGATTGTGGGACTGGGCGG
TGCGGCTGTAGTGAAGGAGAAGGCCGATCCCTGCTTGCATTTCAATCCCAACAAATGCCAGCTGA
GTTTTCAGCCCGATGGCAATCGGTGCGCGGTACTGATCAAGTGCGGATGGGAATGCGAGAACGTG
AGAATCGAGTACAATAACAAGACTCGGAACAATACTCTCGCGTCCGTGTGGCAGCCCGGGGACCC
CGAGTGGTACACCGTCTCTGTCCCCGGTGCTGACGGCTCCCCGCGCACCGTGAATAATACTTTCA
TTTTTGCGCACATGTGCGACACGGTCATGTGGATGAGCAAGCAGTACGATATGTGGCCCCCCACG
AAGGAGAACATCGTGGTCTTCTCCATCGCTTACAGCCTGTGCACGGCGCTAATCACCGCTATCGT
GTGCCTGAGCATTCACATGCTCATCGCTATTCGCCCCAGAAATAATGCCGAAAAAGAAAAACAGC
CATAACGTTTTTTTTCACACCTTTTTCAGACCATGGCCTCTGTTAAATTTTTGCTTTTATTTGCC
AGTCTCATTGCCGTCATTCATGGAATGAGTAATGAGAAAATTACTATTTACACTGGCACTAATCA
CACATTGAAAGGTCCAGAAAAAGCCACAGAAGTTTCATGGTATTGTTATTTTAATGAATCAGATG
TATCTACTGAACTCTGTGGAAACAATAACAAAAAAAATGAGAGCATTACTCTCATCAAGTTTCAA
TGTGGATCTGACTTAACCCTAATTAACATCACTAGAGACTATGTAGGTATGTATTATGGAACTAC
AGCAGGCATTTCGGACATGGAATTTTATCAAGTTTCTGTGTCTGAACCCACCACGCCTAGAATGA
CCACAACCACAAAAACTACACCTGTTACCACTATGCAGCTCACTACCAATAACATTTTTGCCATG
CGTCAAATGGTCAACAATAGCACTCAACCCACCCCACCCAGTGAGGAAATTCCCAAATCCATGAT
TGGCATTATTGTTGCTGTAGTGGTGTGCATGTTGATCATCGCCTTGTGCATGGTGTACTATGCCT
TCTGCTACAGAAAGCACAGACTGAACGACAAGCTGGAACACTTACTAAGTGTTGAATTTTAATTT
TTTAGAACCATGAAGATCCTAGGCCTTTTAATTTTTTCTATCATTACCTCTGCTCTATGCAATTC
TGACAATGAGGACGTTACTGTCGTTGTCGGATCAAATTATACACTGAAAGGTCCAGCGAAGGGTA
TGCTTTCGTGGTATTGCTATTTTGGATCTGACACTACAGAAACTGAATTATGCAATCTTAAGAAT
GGCAAAATTCAAAATTCTAAAATTAACAATTATATATGCAATGGTACTGATCTGATACTCCTCAA
TATCACGAAATCATATGCTGGCAGTTACACCTGCCCTGGAGATGATGCTGACAGTATGATTTTTT
ACAAAGTAACTGTTGTTGATCCCACTACTCCACCTCCACCCACCACAACTACTCACACCACACAC
ACAGATCAAACCGCAGCAGAGGAGGCAGCAAAGTTAGCCTTGCAGGTCCAAGACAGTTCATTTGT
TGGCATTACCCCTACACCTGATCAGCGGTGTCCGGGGCTGCTAGTCAGCGGCATTGTCGGTGTGC
TTTCGGGATTAGCAGTCATAATCATCTGCATGTTCATTTTTGCTTGCTGCTATAGAAGGCTTTAC
CGACAAAAATCAGACCCACTGCTGAACCTCTATGTTTAATTTTTTCCAGAGTCATGAAGGCAGTT
AGCGCTCTAGTTTTTTGTTCTTTGATTGGCATTGTTTTTTGCAATCCTATTCCTAAAGTTAGCTT
TATTAAAGATGTGAATGTTACTGAGGGGGGCAATGTGACACTGGTAGGTGTAGAGGGTGCTGAAA
ACACCACCTGGACAAAATACCACCTCAATGGGTGGAAAGATATTTGCAATTGGAGTGTATTAGTT
TATACATGTGAGGGAGTTAATCTTACCATTGTCAATGCCACCTCAGCTCAAAATGGTAGAATTCA
AGGACAAAGTGTCAGTGTATCTAATGGGTATTTTACCCAACATACTTTTATCTATGACGTTAAAG
TCATACCACTGCCTACGCCTAGCCCACCTAGCACTACCACACAGACAACCCACACTACACAGACA
ACCACATACAGTACATTAAATCAGCCTACCACCACTACAGCAGCAGAGGTTGCCAGCTCGTCTGG
GGTCCGAGTGGCATTTTTGATGTtGGCCCCATCTAGCAGTCCCACTGCTAGTACCAATGAGCAGA
CTACTGAATTTTTGTCCACTGTCGAGAGCCACACCACAGCTACCTCCAGTGCCTTCTCTAGCACC
GCCAATCTCTCCTCGCTTTCCTCTACACCAATCAGTCCCGCTACTACTCCTAGCCCCGCTCCTCT
TCCCACTCCCCTGAAGCAAACAGACGGCGGCATGCAATGGCAGATCACCCTGCTCATTGTGATCG
GGTTGGTCATCCTGGCCGTGTTGCTCTACTACATCTTCTGCCGCCGCATTCCCAACGCGCACCGC
AAGCCGGTCTACAAGCCCATCATTGTCGGGCAGCCGGAGCCGCTTCAGGTGGAAGGGGGTCTAAG
GAATCTTCTCTTCTCTTTTACAGTATGGTGATTGAACTATGATTCCTAGACAATTCTTGATCACT
ATTCTTATCTGCCTCCTCCAAGTCTGTGCCACCCTCGCTCTGGTGGCCAACGCCAGTCCAGACTG
TATTGGGCCCTTCGCCTCCTACGTGCTCTTTGCCTTCACCACCTGCATCTGCTGCTGTAGCATAG
TCTGCCTGCTTATCACCTTCTTCCAGTTCATTGACTGGATCTTTGTGCGCATCGCCTACCTGCGC
CACCACCCCCAGTACCGCGACCAGCGAGTGGCGCGGCTGCTCAGGCTCCTCTGATAAGCATGCGG
GCTCTGCTACTTCTCGCGCTTCTGCTGTTAGTGCTCCCCCGTCCCGTCGACCCCCGGTCCCCCAC
CCAGTCCCCCGAGGAGGTCCGCAAATGCAAATTCCAAGAACCCTGGAAATTCCTCAAATGCTACC
GCCAAAAATCAGACATGCATCCCAGCTGGATCATGATCATTGGGATCGTGAACATTCTGGCCTGC
ACCCTCATCTCCTTTGTGATTTACCCCTGCTTTGACTTTGGTTGGAACTCGCCAGAGGCGCTCTA
TCTCCCGCCTGAACCTGACACACCACCACAGCAACCTCAGGCACACGCACTACCACCACTACAGC
CTAGGCCACAATACATGCCCATATTAGACTATGAGGCCGAGCCACAGCGACCCATGCTCCCCGCT
ATTAGTTACTTCAATCTAACCGGCGGAGATGACTGACCCACTGGCCAACAACAACGTCAACGACC
TTCTCCTGGACATGGACGGCCGCGCCTCGGAGCAGCGACTCGCCCAACTTCGCATTCGCCAGCAG
CAGGAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTGGCCATCCACCAGTGCAAGAGAGGCATCTT
CTGCCTGGTGAAACAGGCCAAGATCTCCTACGAGGTCACTCCAAACGACCATCGCCTCTCCTACG

AGCTCCTGCAGCAGCGCCAGAAGTTCACCTGCCTGGTCGGAGTCAACCCCATCGTCATCACCCAG
CAGTCTGGCGATACCAAGGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGTCCACACTCT
GATCAAGACCCTCTGCGGCCTCCGCGACCTCCTCCCCATGAACTAATCACCCCCTTATCCAGTGA
AATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAATAAAGA
TACAATCATATTGATGATTTGAGTTTAACAAAAAAATAAAGAATCACTTACTTGAAATCTGATAC
CAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGGTACTGCA
GGCCCCGGCGGGCTGCAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTGTCCCTCA
ATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTCGACCCC
GTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTCGTCTCTTC
AGATGGATTCCAAGAGAAGCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCACCACCA
AGAACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCTCGGGAAAACTCATC
TCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTAACAT
GGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACCATTAAATATAC
TGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACTCCGTGGCTCT
GCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATAAAGCTTACCTT
AGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTGGGCTAAAGGTT
TAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTTGGAAGCAGTAGT
ACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGGCCTTAGCTTTGA
CAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTGGACAACACCTG
ATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAACTAACACTTTGCTTGACTAAA
TGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAGGAAGTGGAAACCTAAACCCCAT
TACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACGGTGTTCTTTTAACAG
AACATTCTACACTAAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATGGCACTCCATAT
ACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCTTATCCAAAGTCACAAAGTTCTACTACTAA
AAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTATGCTTCTCACTATAA
CCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATACACCTGGACTAAT
GGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACATCGCCCAAGAATG
AACACTGTATCCCACCCTGCATGCCAACCCTTCCCACCCCACTCTGTGGAACAAACTCTGAAACA
CAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGATTCGAGCAGTTATTT
TTCCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCCGCACAGCCTTGAACATCTGA
ATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCAGAGCGAGCCAGTCT
CGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCT
GAGGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAAT
CATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCCCCGCAGCAGTCGCTGCCGCCGC
CGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCCACGGC
CCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATGCGGATCTCGCTCAGGTCGCTGC
AGTACGTGCAACACAGAACCACCAGGTTGTTCAACAGTCCATAGTTCAACACGCTCCAGCCGAAA
CTCATCGCGGGAAGGATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTAAATCAAGTGGTG
CCCCCTCCAGAACACGCTGCCCACGTACATGATCTCCTTGGGCATGTGGCGGTTCACCACCTCCC
GGTACCACATCACCCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCGGAACCACAGGGCCAGC
ACCGCCCCGCCCGCCATGCAGCGAAGAGACCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCG
CTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCA
TGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCATATCCCAGGGCACGGGGAACTCT
TGCAGGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAG
GGTATCGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCAC
AGCGTGGTAAGGGGGCCGGCCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTCGCGACCGT
GTCATGATGCAGTTGCTTTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGC
ACACCGATCGCCGGCGGCGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAAAACAGCCAC
TCTCTCAGACCGTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGC
TCTGATCACATCGACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTT
CGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCAAACGGTCTCGG
AGTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAAC
AGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGC
GCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATCATGTTA
CACTCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGAACTAGTTCcTG
AGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGC
ACACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAGATTGACAAGCGGAATAT

CAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCC
TCTCCGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAA
CCGAAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGG
TGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGAA
AAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCGTTC
CAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCTGGC
GAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGGCGCGACCCTCG
TAAAAATTGTCGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGTGAATGAT
TCGACAAGATGAATACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAAGCGCCCGAGGAAGC
AATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAA
TTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGCCCCCGATCCCTC
CAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAGA
GTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCCAGATCTAC
ACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCAGCACACGCCCA
GAAACCGGTGACACACTCAAAAAAATACGCGCACTTCCTCAAACGCCCAAAACTGCCGTCATTTC
CGGGTTCCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTAAAAACGTCACC
CGCCCCGCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTCAAACA
CCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATTGATGATGG

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

CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTT
GGGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGGGGGGCGAGTGACGTTTT
GATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGG
TGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTTTCTGG
GCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAATCTGAGTA
ATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGATTACGTGG
GGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGTGTTTTTAC
GTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCCACTCTTGAG
TGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAAGTAGGGATAA
CAGGGTAATgacattgattattgactagttGttaaTAGTAATCAATTACGGGGTCATTAGTTCAT
AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTGCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCAT
ATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
CATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGg
TGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGT
CTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATG
TCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAA
GCAGAgcTCGTTTAGTGAACCGTCAGATCGCCTGGAACGCCATCCACGCTGTTTTGACCTCCATA
GAAGACAGCGATCGCGccacCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCAT
CCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGG
CCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAA
GCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCA
AGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAA
CTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCA
AGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAA
AGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTC
TCGGCATGGACGAGCTtTACAAGTAGtgaGTTTAAACTCCCATTTAAATGTGAGGGTTAATGCTT
CGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAA
ATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAAC
AAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTT
TAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATAACTATAACGGTCCTAAGGTAGCGAGTGAG

TAGTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGT
GTGTTGCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGG
GGCGTCTCCCCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCC
GTGCAGCCCGCGAACTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGC
TGCCGCCGCAGCTGCTGCATCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACT
ACGGCACTCTGGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTG
TTGCTGCTGATGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGC
TCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAAT
AAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTCGCGCGC
GGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAGGACCCGGTAG
AGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCTCCATTG
CAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATGGT
GTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACA
AATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAG
ATTGGCGATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACCACCAGCACGGTGT
ATCCGGTGCACTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACG
CCTTTGTGCCCGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCGATGGGCCCGTGGGCGGC
GGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATCATAGTTGTGGTCCTGGGTGAGGTCATCAT
AGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGGACTGGGGGACAAAGGTACCCTCGATCCCG
GGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTGAGCTCGGAGGGGGGGATCATGTC
CACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGCCGAAAGCAAGT
TCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGATGACCGGCTGCAGG
TGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCCACCTCGTTCATCATCTC
GCGCACGTGCATGTTCTCGCGCACCAGTTCCGCCAGGAGGCGCTCTCCCCCCAGGGATAGGAGCT
CCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGCCATGGGCATTTTGGAGAGGGTT
TGTTGCAAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAG
ACCTCCTCGTTTCGCGGGTTGGGACGGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGC
AGCCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTCACGGTGAAGG
GGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAAAACCGC
TCCCGATCGGCGCCCTGCGCGTCGGCCAGGTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCTC
GGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGGG
ACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGGGGCGTAGGCGTCCCCGCCGCAG
TGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCAG
TTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGGG
TGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCCG
CGGTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAA
GGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCA
AACACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACCG
GGGGTCCCGGCCGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCTCACTGTCTTCCGGATC
GCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCAC
TCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTC
AAGAGCCCCTCGTCCATCTGGTCAGAAAAGACGATCTTTTTGTTGTCGAGCTTGGTGGCGAAGGA
GCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTGGTTTTTTTCCTTGTCGG
CGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCCACGCACTTCCATTCGGGGAAG
ACGGTGGTCAGCTCGTCGGGCACGATTCTGACCTGCCAGCCCCGATTATGCAGGGTGATGAGGTC
CACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGCCCTTGCGCG
AGCAGAAGGGGGGCAGGGGGTCCAGCATGACCTCGTCGGGGGGGTCGGCATCGATGGTGAAGATG
CCGGGCAGGAGGTCGGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCA
TTCGCGCACGGCCAGCGCGCtCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAA
GCGCGGAGGCGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAG
GTGGGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGC
GAGGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAA
TGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCG
ACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGCAGCTTGGCGACGAGCTCGGCGGTGACTAG
GACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATGATGTCATACTTGAGCTGTCCCTTTTGTT
TCCACAGCTCGCGGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCGAGGGGGAACCCG
TCCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTGTAGGCGCAGCAGCC

CTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGCGTGAGGGCGAAAG
TGTCCCTGACCATGACCTTGAGGAACTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCC
CAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAA
GAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCC
GGTTGTTGATGACCTGGGCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCCACGATG
TAGAGTTCCACGAATCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAG
CTCGTCGGGGTCGCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGGGTTGGCGCGGA
GGAAGGAAGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGTACTGACGGAACTGC
TGCCCGACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCCCCGTGCCAGCGATC
CCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCGGTCGTCCCCGGAGAGTTTCA
TGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCAGGTGTAGGTTTCCACATCG
TAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACTGGATCTCCTGCCA
CCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAACACTCGT
GCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACGAGC
TGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGGGAAGTGGAGTCGTGGCGCCTGCATCTCGTG
CTGTACTACGTCGTGGTGGTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTCATGCTGACGAGCC
CGCGCGGGAGGCAGGTCCAGACCTCGGCGCGAGCGGGTCGGAGAGCGAGGACGAGGGCGCGCAGG
CCGGAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGTT
GACTTGCAGGAGTTTTTCCAGGGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCAT
TGGTGGCGACGTCGATGGCTTGCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTCCCCCGTTTC
TTCTTGGGCGGCTGGGGCGACGGGGGGGGTGCCTCTTCCATGGTTAGAAGCGGCGGCGAGGACGC
GCGCCGGGCGGCAGGGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTCGGCGCCGCG
CGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACGT
CCTGGATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTGAAAGAGAGTTCG
ACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCTTGCACGTCGCCCGAGTT
GTCCTGGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCCGG
CGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTCATG
CCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCgGGCGCGCATGACCAC
CTGGGCGAGGTTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGT
AGTTGAGCGTGGTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGGAGCGGCATC
TCGCTGACGTCGCCCAGCGCCTCCAAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAA
AAACTGGGAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGATGAGCTCGGCGATGG
TGGCGCGCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCCTCTTCTTCCTCCTCCACT
AACATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAGGGGGCCTGCGTCGCCGGCG
GCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCGGCGTCGCATGGTCTCGG
TGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAGACGCCGCCGCGCATCTCCAGGTGGCCG
GGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCCCGTAGGGAC
TCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGAAGGCTTCGA
GCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGGTCATGTTGGTTGGGAGCG
GGGGGGGCGATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAG
GAGCACCAGGTCTTTGGGCCCGGCTTGCTGGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGGT
CCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCGCTCCACGGGCACCTCCTCCTCG
CCCGCGCGGCCGTGCATGCGCGTGAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTCGGC
GACGACGCGCTCGGCGAGGATGGCTTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGT
CGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACGGACCAGTTG
ACGGTCTGGTGGCCCGGACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGAA
GATGTAGTCGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGGC
GGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTGG
TAGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGAACTC
GCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTGGCCCG
TGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGGCTCGACTCC
GTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAATCTCGAATCAGG
CTGGAGCCGCAGCTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAG
GATACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCGGATGAGACTAGTAAGCGCGGAAA
GCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGT
GCCCCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAG
ACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCTCTTTTGTTTTGTTTGTTTTTGC

CAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTCCACCGCAACAACAGCCCCCTCC
ACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGACCGCCGCGGCCGCCGTGA
GCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTGGCGCGCCTG
GGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGAGGCCTACGT
GCCCAAGCAGAACCTGTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGATGCGCGCGGCCCGGT
TCCACGCGGGGCGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGTGCTGAGGGACGAGGATTTC
GAGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCAC
GGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGCA
CCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTGGAGGCCATC
GTGCAGAACCCCACCAGCAAGCCGCTGACGGCGCAGCTGTTCCTGGTGGTGCAGCATAGTCGGGA
CAACGAAGCGTTCAGGGAGGCGCTGCTGAATATCACCGAGCCCGAGGGCCGCTGGCTCCTGGACC
TGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGCTGCCGCTGTCCGAGAAGCTGGCG
GCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGGAAGATCTACAAGACCCCGTA
CGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTGA
CCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAGCGCCAGCAGG
CGGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGGGAC
CGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCCTTGG
AGGCGGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTG
GAAGACTGATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGCG
ATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCAT
GCAACGCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACC
GGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCCCACGCACCAGAAGGTCCTG
GCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCGACGAGGCCGGCCTGGTGTACAA
CGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAACGTGCAGACCAACCTGGACCGCATGG
TGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCACCGCGAGTCCAACCTGGGATCC
ATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCGCCAACGTGCCCCGGGGCCAGGAGGACTA
CACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGGTGTACCAGT
CCGGGCCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCGTGAACCTGAGCCAGGCT
TTCAAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCCCCGGTCGGGGACCGCGCGACGGTGTCGAG
CCTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGGCAGCA
TCAACCGCAACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCAC
GTGGACGAGCAGACCTACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGG
CAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGATCCCGCCCCAGTACG
CGCTCAGCACCGAGGAGGAGCGCATCCTGCGTTACGTGCAGCAGAGCGTGGGCCTGTTCCTGATG
CAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGGAGCCCAGCATGTA
CGCCAGCAACCCCCCGTTCATCAATAAACTGATGGACTACTTGCATCGGGCGGCCGCCATGAACT
CTGACTATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCCCCGCCGGGGTTCTACACGGGC
GAGTACGACATGCCCGACCCCAATGACGGGTTCCTGTGGGACGATGTGGACAGCAGCGTGTTCTC
CCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCCCGTCCT
CGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCTTTCCCGAGC
TTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCT
GGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATA
ACGGGATAGAAAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGAC
GATCCCCGGGCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCGGTGGCACGA
CAGGCAGCGGGGACAGATGTGGGACGATGAGGACTCCGCCGACGACAGCAGCGTGTTGGACTTGG
GTGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGGCGCATGATGTAAGAGAAACCG
AAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCTGTTGTTGTT
GTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTCCCTCGTACGAGAGCGTGATGCA
GCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTACGTGCCCCCGCGGTACC
TGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTACGATACCACC
CGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACGACCACAG
CAACTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCACCCAGACCA
TCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCC
AACGTGAACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCC
CAATGGGGTGACAGTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGG
AATTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATC
ATCGACAATTACTTGGCGGTGGGGGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTC

GACACTAGGAACTTCAGGCTGGGCTGGGACCCCGTGACCGAGCTGGTCATGCCCGGGGTGTACAC
CAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTCACCGAGAGCC
GCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCAGCCCTTCCAGGAAGGCTTCCAGATCATGTAC
GAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTCGACGCCTATGAGAAAAGCAAGGA
GGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTCAGGGGCGATAATT
TTGCAAGCGCCGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCATTCAG
CCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTA
CCGCAGCTGGTACCTAGCCTACAACTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGC
TCACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATGCAA
GACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCT
GCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTCA
CCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCCGCGCCC
ACCATTACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGCCGCTGCGCAG
CAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTGCCCCTACGTCT
ACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTAAATGTCCATTCTC
ATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGATGTACGGAGGCGCTCG
CCAACGCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGGGGCGCCCTCA
AGGGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGCGCGC
AACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCGTGGTGGCcGA
CGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACCGGAGCACCC
CCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTC
AGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGC
GGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACG
CCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCCCCTCGCACTTGAAGATGTTCACTTCGC
GATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCCAAGCGCAAATTCAAGGAAGAGATGCTCCAG
GTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGAAGGAGGAAAGAAAGCCCCGCAAAATCAA
GCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTGATGTGGACGGATTGGTGGAGTTTGTGCGCG
AGTTCGCCCCCCGGCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCGGTGCTGAGACCCGGC
ACCACCGTGGTCTTCACGCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCCTACGACGAGGT
GTACGGGGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGCTTACGGCAAGC
GCAGCCGTTCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCCACGCCG
AGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGGTTCAAGCG
CGAGGGCGAGGATCTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACG
TGCTGGAGACCATGAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCCCATCAAGCAG
GTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACATCAAGATTCCCACGGAGCCCATGGAAACGCA
GACCGAGCCCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCAT
CGGCTCCTAGTCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTGCTGATGCCCAACTACGCG
CTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGCTTCTACCGCGGTCATACCAG
CAGCCGCCGCCGCAAGACCACCACTCGCCGCCGCCGTCGCCGCACCGCCGCTGCAACCACCCCTG
CCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCTGCCGCGCGCGCTAC
CACCCGAGCATCGCCATTTAAACTTTCGCCTGCTTTGCAGATCAATGGCCCTCACATGCCCCCTT
CGCGTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACGGGAT
GCGTCGCCACCACCACCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGAGGCTTCCTGCCCG
CGCTGATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAG
GCCTCTCAGCGCCACTGAGACACACTTGGAAACATCTTGTAATAAACCaATGGACTCTGACGCTC
CTGGTCCTGTGATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTTCGTCCCTGGCTCCGCGA
CACGGCACGCGGCCGTTCATGGGCACCTGGAGCGACATCGGCACCAGCCAACTGAACGGGGGCGC
CTTCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGAATTTCGGGTCCACGCTTAAAACCTATGGCA
GCAAGGCGTGGAACAGCACCACAGGGCAGGCGCTGAGGGATAAGCTGAAAGAGCAGAACTTCCAG
CAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGGGGTGGTGGACCTGGCCAACCAGGCCGT
GCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCCGCCCGCCGGCTCCGTGGAGATGCCGCAGG
TGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCCCGATGCGGAGGAG
ACGCTGCTGACGCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAACTGGGTCTGCCCAC
CACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCTGGACT
TGCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGCC
CGCGCGCGACCCGGGGGCACCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGT
GGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTG

CTTGTCTGTGTGTGTATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGG
CGCGTCGCCGAGTTGCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCG
CCGGACAGGACGCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTGCCCGCGCCACAGACACC
TACTTCAGTCTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCACGATGTGACCACCGA
CCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGACAACACCTACTCGTACA
AAGTGCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACATGGCCAGCACCTACTTTGAC
ATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTCCGGCACCGCCTACAACAGTCT
GGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAAAGCCGATGGTGAAACTGCCACAG
AAAAAACCTATACATATGGAAATGCACCCGTGCAGGGCATTAACATCACAAAAGATGGTATTCAA
CTTGGAACTGACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGT
GGGTGATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGC
CTGATACCAAAATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTCAG
GCAAATGTGAAAACAGGAACAGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTGACAA
CAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGATTTGG
AAACTCCAGATACCCATATTGTATACAAAGCAGGCACAGATGACAGCAGCTCTTCTATTAATTTG
GGTCAGCAAGCCATGCCCAACAGACCTAACTACATTGGTTTCAGAGACAACTTTATCGGGCTCAT
GTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCAGGCTTCTCAGCTGAATGCTGTGG
TTGACTTGCAAGACAGAAACACCGAGCTGTCCTACCAGCTCTTGCTTGACTCTCTGGGTGACAGA
ACCCGGTATTTCAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTGCGCATTATTGA
AAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGCAGAACAG
ATACTTATCAGGGAATTAAGGCTAATGGAACTGATCAAACCACATGGACCAAAGATGACAGTGTC
AATGATGCTAATGAGATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACATCCAAGCCAACCT
GTGGAGGAACTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAGTACACGCCGG
CCAATGTTACCCTGCCCACCAACACCAACACCTACGATTACATGAACGGCCGGGTGGTGGCCCCC
TCGCTGGTGGACTCCTACATCAACATCGGGGCGCGCTGGTCGCTGGATCCCATGGACAACGTGAA
CCCCTTCAACCACCACCGCAATGCGGGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCT
ACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCC
GGGTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACATGATCCTGCAGAGCTCCCTCGG
CAACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCAGCATCAACCTCTACGCCACCTTCT
TCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAACGACACCAACGACCAG
TCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAACGCCACCAACGT
GCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCTTCCGCGGCTGGTCCTTCACGCGTCTCAAGA
CCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTGGGGCTCCATCCCC
TACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTC
CGTCAGCTGGCCCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAAGCGCACCGTCG
ACGGCGAGGGCTACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTG
GCCCACTACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTC
CTTCTTCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACC
AGGCCGTCACCCTGGCCTACCAGCACAACAACTGGGGCTTCGTCGGCTACCTCGCGCCCACCATG
CGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGTCACCAG
CGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCAACTTCA
TGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGCCCACGCGCTA
GACATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTTCGAAGTCTT
CGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCGCACCCCCT
TCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGGCG
AGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCTGGGCACCTTCGAT
AAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCG
CGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGCTCGAACACCTGCTACCTCT
TCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCCTG
CTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTGCGTCACCCTGGAAAAGTCCACCCAGACCGT
GCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTGCATGTTCCTGCACGCCTTCGTGCACT
GGCCCGACCGCCCCATGGACAAGAACCCCACCATGAACTTGCTGACGGGGGTGCCCAACGGCATG
CTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTCTACCGCTTCCTCAA
CTCCCACTCCGCCTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCACCGCCTTCGACCGCA
TGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCACTTTCATGTTA
CACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTTCTGCCGGGTCTCGGCATGGCCCGC
GGGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACTCGGGGATCAGCAGTTTGG

GCAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGG
TCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACAC
GGGGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTCGCGTCGG
TGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCATCCCGAAGGGGGTCATCTTGCAGGTCTGC
CTTCCCATGGTGGGCACGCACCCGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCAT
CTGGGCCTGGTCGGCGTTCATCCCCGGGTACATGGCCTTCATGAAAGCCTCCAATTGCCTGAACG
CCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCCGCAGGACTTGCTAGAGAACTGGTTGGTG
GCGCACCCGGCGTCGTGCACGCAGCAGCGCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCC
CCAGCGGTTCTGGGTGATCTTGGCCCGGTCGGGGTTCTCCTTCAGCGCGCGCTGCCCGTTCTCGC
TCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGC
TTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCCAGTTCTTGTG
GGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGGTCT
TGTTGCTAGTGAAGGTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGG
CGGTACACCTCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCTCCACGCGGTA
GCGGTCCATCAGCATAGTCATGATTTCCATACCCTTCTGCCAGGCCGAGACGATGGGCAGGCTCA
TAGGGTTCTTCACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCTCGTCCAGGGTC
TCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGTAGCTGAAGCCCACGGCCGC
CAGCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCTGCAGGACCACATGCT
TGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGTTGGAGATGGCGAGGGGGAG
CGCGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCGAGGCCACGCGGCGGTAGGT
ATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCTCGCCGCCGCGACTTGGGGGATGGCTGG
CAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCCTCCGCGGCCG
GCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACCTCGCCATC
TGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCCGCCCAGCC
CCGCCACCTCCGACGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTG
GGCTATGTGACGCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTTCACAAGAAGAGATACA
CCAAGAACAGCCAGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCG
ACTACCTCCACCTGAGCGGGGGGGAGGACGCGCTCATCAAGCATCTGGCCCGGCAGGCCACCATC
GTCAAGGATGCGCTGCTCGACCGCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTA
CGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCAAGCGCCAGCCCAATGGCACCTGCGAGCCCA
ACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTACCACATCTTT
TTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCACCCGCGCCGACGCCCTTTTCAA
CCTGGGTCCCGGCGCCCCCCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAAGATCTTCGAGG
GTCTGGGCAGCGACGAGACTCGGGCCCCGAACGCTCTGCAAGGAGAAGGAGGAGAGCATGAGCAC
CACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGA
GCTGACCCATTTCGCCTACCCGGCTCTGAACCTGCCCCCCAAAGTCATGAGCGCGGTCATGGACC
AGGTGCTCATCAAGCGCGCGTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGC
AAGCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCCTAATGCTAGTCCCCAGAGTTT
GGAAGAGCGGCGCAAACTCATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCC
GCTTCTTCGCCGACGCGGAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCAC
GGGTTCGTGCGCCAGGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCTCCTACATGGG
CATCTTGCACGAGAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCGGGGAGGCCCGGC
GCGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGACGGGCATGGGCGTG
TGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCCTGCAGAAGAACCTCAA
GGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACCTCATTTTCCCCG
AGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGCATGTTGCAAAAC
TTTCGCTCTTTCATCCTGGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGGA
CTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGCGCC
TGGCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTCAGCGGCGAGGGCCTGCTCGAG
TGCCACTGCCGCTGCAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAG
CGAGACCCAGATCATCGGCACCTTCGAGTTGCAAGGGCCCAGCGAAGGCGAGGGTTCAGCCGCCA
AGGGGGGTCTGAAACTCACCCCGGGGCTGTGGACCTCGGCCTACTTGCGCAAGTTCGTGCCCGAG
GACTACCATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTC
GGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCCATCCAGAAATCCCGCCAAG
AATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAGACCGGTGAGGAGCTCAACCCC
GGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAGCTGCCCCCCGTGGAGGATT
TGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGACTGGGACAGC

ACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGGAGGA
GGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCTCGGCGGGGGAGAAAGCAAGCAGCACGGATA
CCATCTCCGCTCCGGGTCGGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTC
CCGAACCCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCCTGGCGGGGGCACAA
AAACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGC
TCTTCCACCGCGGGGTGAACTTTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCC
TACTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAAAGACCAGCAGAAAACCAGCAGCTAGAAAAT
CCACAGCGGCGGCAGCAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAAACCCGGGAGCTG
AGGAACCGGATCTTTCCCACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGGAGCAGGAACT
GAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAGAGCGAAGACC
AACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCTCACTCTTAAA
GAGTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGTGCCCTTCGCCCT
AGCCGCCTCCACCCATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTGGAGCTACCAGCCC
CAGATGGGCCTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGG
GCCCGCGATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGATACTCCTAGAACAGT
CAGCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTAC
CAGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGAC
TAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTCGTCACCGCCCCGCTCAGGGTATAA
AGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGT
CTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTCCTTCACGCCTCGTCAGGC
CGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGTGGCATCGGCACTCTCCAGTTCG
TGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGGCCACTACCCGGAC
GAGTTCATCCCGAACTTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATTGAATGTCCCATGG
TGGCGCAGCTGACCTAGCTCGGCTTCGACACCTGGACCACTGCCGCCGCTTCCGCTGCTTCGCTC
GGGATCTCGCCGAGTTTGCCTACTTTGAGCTGCCCGAGGAGCACCCTCAGGGCCCGGCCCACGGA
GTGCGGATCGTCGTCGAAGGGGGCCTCGACTCCCACCTGCTTCGGATCTTCAGCCAGCGTCCGAT
CCTGGTCGAGCGCGAGCAAGGACAGACCCTTCTGACTCTGTACTGCATCTGCAACCACCCCGGCC
TGCATGAAAGTCTTTGTTGTCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTACTC
CGGACTTCCGTGTGTTCCTGAATCCATCAACCAGTCTTTGTTCTTCACCGGGAACGAGACCGAGC
TCCAGCTCCAGTGTAAGCCCCACAAGAAGTACCTCACCTGGCTGTTCCAGGGCTCCCCGATCGCC
GTTGTCAACCACTGCGACAACGACGGAGTCCTGCTGAGCGGCCCTGCCAACCTTACTTTTTCCAC
CCGCAGAAGCAAGCTCCAGCTCTTCCAACCCTTCCTCCCCGGGACCTATCAGTGCGTCTCGGGAC
CCTGCCATCACACCTTCCACCTGATCCCGAATACCACAGCGTCGCTCCCCGCTACTAACAACCAA
ACTAACCTCCACCAACGCCACCGTCGCGACGGCCACAATACATGCCCATATTAGACTATGAGGCC
GAGCCACAGCGACCCATGCTCCCCGCTATTAGTTACTTCAATCTAACCGGCGGAGATGACTGACC
CACTGGCCAACAACAACGTCAACGACCTTCTCCTGGACATGGACGGCCGCGCCTCGGAGCAGCGA
CTCGCCCAACTTCGCATTCGCCAGCAGCAGGAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTGGC
CATCCACCAGTGCAAGAGAGGCATCTTCTGCCTGGTGAAACAGGCCAAGATCTCCTACGAGGTCA
CTCCAAACGACCATCGCCTCTCCTACGAGCTCCTGCAGCAGCGCCAGAAGTTCACCTGCCTGGTC
GGAGTCAACCCCATCGTCATCACCCAGCAGTCTGGCGATACCAAGGGGTGCATCCACTGCTCCTG
CGACTCCCCCGACTGCGTCCACACTCTGATCAAGACCCTCTGCGGCCTCCCCGACCTCCTCCCCA
TGAACTAATCACCCCCTTATCCAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAA
AAATAATCATTTGATTTGAAATAAAGATACAATCATATTGATGATTTGAGTTTAACAAAAAAATA
AAGAATCACTTACTTGAAATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCAC
TCCCCTCTTCCCAGCTCTGGTACTGCAGGCCCCGGCGGGCTGCAAACTTCCTCCACACGCTGAAG
GGGATGTCAAATTCCTCCTGTCCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGC
GTCCGGGTGGATGATGACTTCGACCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCC
CTTCATCAACCCCCCCTTCGTCTCTTCAGATGGATTCCAAGAGAAGCCCCTGGGGGTGTTGTCCC
TGCGACTGGCCGACCCCGTCACCACCAAGAACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGTG
GACCTCGATTCCTCGGGAAAACTCATCTCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTT
TTCCAACAACACCATTTCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCT
TACAAGTTTCTCCACCATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTT
GGATCAGGTTTAGGACTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGA
TACTGATGGAAACATAAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTG
AAAGCAACATAAGCTGGGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGA
AATGGGTTAGAGTTTGGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGT
TAAACTTGGATCTGGCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACG

ATAAACTCACTTTGTGGACAACACCTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGAT
GCAAAACTAACACTTTGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGT
TGTAGGAAGTGGAAACCTAAACCCCATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTT
TTGATGCAAACGGTGTTCTTTTAACAGAACATTCTACACTAAAAAAATACTGGGGGTATAGGCAG
GGAGATAGCATAGATGGCACTCCATATACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCTTA
TCCAAAGTCACAAAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATGGAGATG
TTTCAAAACCTATGCTTCTCACTATAACCCTCAATGGTACTGATGACAGCAACAGTACATATTCA
ATGTCATTTTCATACACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTA
TACCTTCTCATACATCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCCTTCCCACC
CCACTCTGTGGAACAAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAAC
AGTTTTACAGGATTCGAGCAGTTATTTTTCCTCCACCCTCCCAGGACATGGAATACACCACCCTC
TCCCCCCGCACAGCCTTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTT
CCACACAGTTTCAGAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCC
GCATCTGCACCTCACAGCTCAACAGCTGAGGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGG
AAGAAGCAGAAGAGCGGCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATC
AGGCCCCGCAGCAGTCGCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAG
GGACTCCCTCAGCATGATGCCCACGGCCCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGC
GCATGCGGATCTCGCTCAGGTCGCTGCAGTACGTGCAACACAGAACCACCAGGTTGTTCAACAGT
CCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGGAAGGATGCTACCCACGTGGCCGTCGTA
CCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACATGATCTCCT
TGGGCATGTGGCGGTTCACCACCTCCCGGTACCACATCACCCTCTGGTTGAACATGCAGCCCCGG
ATGATCCTGCGGAACCACAGGGCCAGCACCGCCCCGCCCGCCATGCAGCGAAGAGACCCCGGGTC
CCGGCAATGGCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTA
TGTTGGCACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAA
ACCATATCCCAGGGCACGGGGAACTCTTGCAGGACAGCGAACCCCGCAGAACAGGGCAATCCTCG
CACAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGATCCTCCACCA
GAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCGGCCGATACGGGTGATGGCGG
GACGCGGCTGATCGTGTTCGCGACCGTGTCATGATGCAGTTGCTTTCGGACATTTTCGTACTTGC
TGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCGGCGGCGGTCTCGGCGCTTGGAACGC
TCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACCGTGCAGCAGATCTAGGGCCTCAGGAGT
GATGAAGATCCCATCATGCCTGATGGCTCTGATCACATCGACCACCGTGGAATGGGCCAGACCCA
GCCAGATGATGCAATTTTGTTGGGTTTCGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCATG
ATTAACTTTTAATCCAAACGGTCTCGGAGTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTC
GCCCCCGCTGTGTTGGTGGAAAATAACAGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCA
CGGTGGCTTCCAGCAAAGCCTCCACGCGCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGG
TTCTCTAATTCCTCAATCATCATGTTACACTCCTGCACCATCCCCAGATAATTTTCATTTTTCCA
GCCTTGAATGATTCGAACTAGTTCGTGAGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCA
GAGCGCCCTCCACCGGCATTCTTAAGCACACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCA
CCTGCAGCAGATTGACAAGCGGAATATCAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGC
AATAACTGTAAGTACTCTTTCATATCCTCTCCGAAATTTTTAGCCATAGGACCACCAGGAATAAG
ATTAGGGCAAGCCACAGTACAGATAAACCGAAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGAC
TGCTATAAGCATGCTGGCTAGACCCGGTGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGG
CAATTTTTAAGAAAATCAACAAAAGAAAAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAAC
GATGAAGTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAA
ATGAACATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGCAG
GCCACGGGGTCTCCGGCGCGACCCTCGTAAAAATTGTCGCTATGATTGAAAACCATCACAGAGAG
ACGTTCCCGGTGGCCGGCGTGAATGATTCGACAAGATGAATACACCCCCGGAACATTGGCGTCCG
CGAGTGAAAAAAAGCGCCCGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAA
GCGATGCCATGCGGATGAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCCTGCTG
CACAGGCAGCAAAGCCCCCGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACC
GAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTC
TCTGCTCAATATATAGCCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAA
TAATCACACACGCCCAGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGCGCACTTCC
TCAAACGCCCAAAACTGCCGTCATTTCCGGGTTCCCACGCTACGTCATCAAAACACGACTTTCAA
ATTCCGTCGACCGTTAAAAACGTCACCCGCCCCGCCCCTAACGGTCGCCCGTCTCTCAGCCAATC
AGCGCCCCGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGT
ATATTATTGATGATGG ChAdV68.4WTnt.MAG25mer (SEQ ID NO: 12); AC_000011.1 with E1

(nt 577 to 3403) and E3 (nt 27,816-31,332) sequences deleted; corresponding ATCC VR-594 nucleotides substituted at four positions; model neoantigen cassette under the control of the CMV promoter/enhancer inserted in place of deleted E1

CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTT
GGGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGGGGGGCGAGTGACGTTTT
GATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGG
TGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTTTCTGG
GCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAATCTGAGTA
ATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGATTACGTGG
GGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGTGTTTTTAC
GTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCCACTCTTGAG
TGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAAGTAGGGATAA
CAGGGTAATgacattgattattgactagttGttaaTAGTAATCAATTACGGGGTCATTAGTTCAT
AGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAA
CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCAT
ATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
CATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGg
TGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGT
CTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATG
TCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAA
GCAGAgcTCGTTTAGTGAACCGTCAGATCGCCTGGAACGCCATCCACGCTGTTTTGACCTCCATA
GAAGACAGCGATCGCGccacCATGGCCGGGATGTTCCAGGCACTGTCCGAAGGCTGCACACCCTA
TGATATTAACCAGATGCTGAATGTCCTGGGAGACCACCAGGTCTCTGGCCTGGAGCAGCTGGAGA
GCATCATCAACTTCGAGAAGCTGACCGAGTGGACAAGCTCCAATGTGATGCCTATCCTGTCCCCA
CTGACCAAGGGCATCCTGGGCTTCGTGTTTACCCTGACAGTGCCTTCTGAGCGGGGCCTGTCTTG
CATCAGCGAGGCAGACGCAACCACACCAGAGTCCGCCAATCTGGGCGAGGAGATCCTGTCTCAGC
TGTACCTGTGGCCCCGGGTGACATATCACTCCCCTTCTTACGCCTATCACCAGTTCGAGCGGAGA
GCCAAGTACAAGAGACACTTCCCAGGCTTTGGCCAGTCTCTGCTGTTCGGCTACCCCGTGTACGT
GTTCGGCGATTGCGTGCAGGGCGACTGGGATGCCATCCGGTTTAGATACTGCGCACCACCTGGAT
ATGCACTGCTGAGGTGTAACGACACCAATTATTCCGCCCTGCTGGCAGTGGGCGCCCTGGAGGGC
CCTCGCAATCAGGATTGGCTGGGCGTGCCAAGGCAGCTGGTGACACGCATGCAGGCCATCCAGAA
CGCAGGCCTGTGCACCCTGGTGGCAATGCTGGAGGAGACAATCTTCTGGCTGCAGGCCTTTCTGA
TGGCCCTGACCGACAGCGGCCCCAAGACAAACATCATCGTGGATTCCCAGTACGTGATGGGCATC
TCCAAGCCTTCTTTCCAGGAGTTTGTGGACTGGGAGAACGTGAGCCCAGAGCTGAATTCCACCGA
TCAGCCATTCTGGCAGGCAGGAATCCTGGCAAGGAACCTGGTGCCTATGGTGGCCACAGTGCAGG
GCCAGAATCTGAAGTACCAGGGCCAGAGCCTGGTCATCAGCGCCTCCATCATCGTGTTTAACCTG
CTGGAGCTGGAGGGCGACTATCGGGACGATGGCAACGTGTGGGTGCACACCCCACTGAGCCCCAG
AACACTGAACGCCTGGGTGAAGGCCGTGGAGGAGAAGAAGGGCATCCCAGTGCACCTGGAGCTGG
CCTCCATGACCAATATGGAGCTGATGTCTAGCATCGTGCACCAGCAGGTGAGGACATACGGACCC
GTGTTCATGTGCCTGGGAGGCCTGCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAGTGCGGGT
GCTGGAGCTGTTCAGAGCCGCCCAGCTGGCCAACGATGTGGTGCTGCAGATCATGGAGCTGTGCG
GAGCAGCCTTTCGCCAGGTGTGCCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAG
TGGAACAATGAGACAACACAGCCTCAGATCGCCAACTGTAGCGTGTACGACTTCTTCGTGTGGCT
GCACTACTATAGCGTGAGGGATACCCTGTGGCCCCCCGTGACATACCACATGAATAAGTACGCCT
ATCACATGCTGGAGAGGCGCGCCAAGTATAAGAGAGGCCCTGGCCCAGGCGCAAAGTTTGTGGCA
GCATGGACCCTGAAGGCCGCCGCCGGCCCCGGCCCCGGCCAGTATATCAAGGCTAACAGTAAGTT
CATTGGAATCACAGAGCTGGGACCCGGACCTGGATAATGAGTTTAAACTCCCATTTAAATGTGAG
GGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATG
CAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAG
CTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGT
GGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATAACTATAACGGTCCTAAGG
TAGCGAGTGAGTAGTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTG
TGCTTTTCTGTGTGTTGCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCC
TTATCTGACGGGGCGTCTCCCCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGG
ACGGCCGGCCCGTGCAGCCCGCGAACTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCG
TTGGACGCAGCTGCCGCCGCAGCTGCTGCATCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGG

CGCCGGCTACTACGGCACTCTGGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCCTGAACG
AGGAGAAGCTGTTGCTGCTGATGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACC
CAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAA
AATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGAT
TTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCA
GGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGG
TAGCTCCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCG
CAGGGCATGGTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGT
AGGTGTTTACAAATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCC
TGGATCTTGAGATTGGCGATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACCAC
CAGCACGGTGTATCCGGTGCACTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGA
ATTTGGCGACGCCTTTGTGCCCGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCGATGGGC
CCGTGGGGGGGGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATCATAGTTGTGGTCCTGGGTG
AGGTCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGGACTGGGGGACAAAGGTACC
CTCGATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTGAGCTCGGAGGGGG
GGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGCC
GAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGATGAC
CGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCCACCTCGT
TCATCATCTCGCGCACGTGCATGTTCTCGCGCACCAGTTCCGCCAGGAGGCGCTCTCCCCCCAGG
GATAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGCCATGGGCATTTT
GGAGAGGGTTTGTTGCAAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTC
GATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGACGGCTGCGGGAGTAGGGCACCAGACGATGGG
CGTCCAGCGCAGCCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTC
ACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGT
CGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAGGTAGCAATTGACCATGAGTTCGTAGT
TGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCGCAGGCGGGA
CAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGGGGCGTAGGCGTC
CGCGCCGCAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGT
CAAAAACCAGTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGT
CCCCGCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAG
CGGTGTGCCGCGGTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGG
CCAGCACGAAGGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCC
AGGGTATGCAAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGC
CACGTGACCGGGGGTCCCGGCCGGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCTCACTGT
CTTCCGGATCGCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATG
ACCTCGGCACTCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCGGCGGA
GATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAGACGATCTTTTTGTTGTCGAGCTTGG
TGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTGGTTTTTT
TCCTTGTCGGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCCACGCACTTCCA
TTCGGGGAAGACGGTGGTCAGCTCGTCGGGCACGATTCTGACCTGCCAGCCCCGATTATGCAGGG
TGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCG
CCCTTGCGCGAGCAGAAGGGGGGCAGGGGGTCCAGCATGACCTCGTCGGGGGGGTCGGCATCGAT
GGTGAAGATGCCGGGCAGGAGGTCGGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGG
CAGCTTGCCATTCGCGCACGGCCAGCGCGCCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGG
GATGGGTAAGCGCGGAGGCGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCTCCTCGAGGATG
CCGATGTAGGTGGGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTG
CGAGGGGGCGAGGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCT
GGCGGAAAATGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGTGG
GGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGCAGCTTGGCGACGAGCTCGGC
GGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATGATGTCATACTTGAGCTGTC
CCTTTTGTTTCCACAGCTCGCGGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCGAGG
GGGAACCCGTCCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTGTAGGC
GCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGCGTGA
GGGCGAAAGTGTCCCTGACCATGACCTTGAGGAACTGGTGCTTGAAGTCGATATCGTCGCAGCCC
CCCTGCTCCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAAC
ATCGTTGAAGAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCA
CCTCGGCCCGGTTGTTGATGACCTGGGCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGG

CCCACGATGTAGAGTTCCACGAATCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTC
GTAGGTGAGCTCGTCGGGGTCGCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGGGT
TGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGTACTGA
CGGAACTGCTGCCCGACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCCCCGTG
CCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCGGTCGTCCCCGG
AGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCAGGTGTAGGTT
TCCACATCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACTGGAT
CTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCG
AACACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGC
TGCACGAGCTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGGGAAGTGGAGTCGTGGCGCCTG
CATCTCGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTCATGC
TGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGGCGCGAGCGGGTCGGAGAGCGAGGACGAGG
GCGCGCAGGCCGGAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGTGGGCAGCGGCGG
CGCGCGGTTGACTTGCAGGAGTTTTTCCAGGGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCA
CCGCGCCATTGGTGGCGACGTCGATGGCTTGCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTC
CCCCGTTTCTTCTTGGGCGGCTGGGGCGACGGGGGCGGTGCCTCTTCCATGGTTAGAAGCGGCGG
CGAGGACGCGCGCCGGGCGGCAGGGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTC
GGCGCCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGAC
GGTTGACGTCCTGGATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTGAAA
GAGAGTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCTTGCACGTC
GCCCGAGTTGTCCTGGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTCTTGAAGGTCTC
CGCGGCCGGCGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAG
GCGTTCATGCCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCgGGCGCG
CATGACCACCTGGGCGAGGTTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCT
GGTAGAGGTAGTTGAGCGTGGTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGG
AGCGGCATCTCGCTGACGTCGCCCAGCGCCTCCAAACGTTCCATGGCCTCGTAAAAGTCCACGGC
GAAGTTGAAAAACTGGGAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGATGAGCT
CGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCCTCTTCTTCC
TCCTCCACTAACATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAGGGGGCCTGCG
TCGCCGGCGGCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCGGCGTCGCA
TGGTCTCGGTGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAGACGCCGCCGCGCATCTCC
AGGTGGCCGGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCC
CGTAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGA
AGGCTTCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGGTCATGTTGG
TTGGGAGCGGGGGGGGCGATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGAT
GGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCTTGCTGGATGCGCAGACGGTCGGCCATGCCCC
AGGCGTGGTCCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCGCTCCACGGGCACC
TCCTCCTCGCCCGCGCGGCCGTGCATGCGCGTGAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGC
CAGGTCGGCGACGACGCGCTCGGCGAGGATGGCTTGCTGGATCTGGGTGAGGGTGGTCTGGAAGT
CATCAAAGTCGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACG
GACCAGTTGACGGTCTGGTGGCCCGGACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCG
CGTGTCGAAGATGTAGTCGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCG
GCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAGCATG
GTGCGGTGGTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCG
CGGGAACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGG
TCTGGCCCGTGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGG
CTCGACTCCGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAATCT
CGAATCAGGCTGGAGCCGCAGCTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCA
ACCCTCCAGGATACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCGGATGAGACTAGTAA
GCGCGGAAAGCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCAGGGTTGCG
TTGCGGTGTGCCCCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTGGCTGCCCCGTC
GTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCTCTTTTGTTTTGTT
TGTTTTTGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTCCACCGCAACAACA
GCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGACCGCCGCGG
CCGCCGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTG
GCGCGCCTGGGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGA
GGCCTACGTGCCCAAGCAGAACCTGTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGATGCGCG

CGGCCCGGTTCCACGCGGGGCGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGTGCTGAGGGAC
GAGGATTTCGAGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAA
CCTGGTCACGGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACC
ACGTGCGCACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTG
GAGGCCATCGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGCAGCTGTTCCTGGTGGTGCAGCA
TAGTCGGGACAACGAAGCGTTCAGGGAGGCGCTGCTGAATATCACCGAGCCCGAGGGCCGCTGGC
TCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGCTGCCGCTGTCCGAG
AAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGGAAGATCTACAA
GACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGACCCTGA
AAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAGC
GCCAGCAGGCGGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGG
GGCCGGGACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCC
GGGCCTTGGAGGCGGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGC
GAGTACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCT
GATCCCGCGATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGAC
CCAGGCCATGCAACGCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCC
AGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCCCACGCACGAG
AAGGTCCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCGACGAGGCCGGCCT
GGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAACGTGCAGACCAACCTGG
ACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCACCGCGAGTCCAAC
CTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCGCCAACGTGCCCCGGGGCCA
GGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGG
TGTACCAGTCCGGGCCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCGTGAACCTG
AGCCAGGCTTTCAAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCCCCGGTCGGGGACCGCGCGAC
GGTGTCGAGCCTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACA
GCGGCAGCATCAACCGCAACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGC
CAGGCGCACGTGGACGAGCAGACCTACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGA
CGACCCGGGCAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGATCCCGC
CCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTGCGTTACGTGCAGCAGAGCGTGGGCCTG
TTCCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGGAGCC
CAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAACTGATGGACTACTTGCATCGGGCGGCCG
CCATGAACTCTGACTATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCGCCGGGGTTC
TACACGGGCGAGTACGACATGCCCGACCCCAATGACGGGTTCCTGTGGGACGATGTGGACAGCAG
CGTGTTCTCCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGAC
GCCCGTCCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCT
TTCCCGAGCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCC
GCGCTTGCTGGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACT
TCCCCAATAACGGGATAGAAAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAG
CACAGGGACGATCCCCGGGCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCG
GTGGCACGACAGGCAGCGGGGACAGATGTGGGACGATGAGGACTCCCCCGACGACAGCAGCGTGT
TGGACTTGGGTGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGGCGCATGATGTAA
GAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCT
GTTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTCCCTCGTACGAGAG
CGTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTACGTGCCCC
CGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTAC
GATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAA
CGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCA
CCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACC
AACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCG
CAAGACCCCCAATGGGGTGACAGTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATG
AATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATCGACCTGATGAAC
AACGCCATCATCGACAATTACTTGGCGGTGGGGCGGCAGAACGGGGTGCTGGAGAGCGACATCGG
CGTGAAGTTCGACACTAGGAACTTCAGGCTGGGCTGGGACCCCGTGACCGAGCTGGTCATGCCCG
GGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTC
ACCGAGAGCCCCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCAGCCCTTCCAGGAAGGCTTCCA
GATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCCCGCTCCTGGATGTCGACGCCTATGAGA
AAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTCAGG

GGCGATAATTTTGCAAGCGCCGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGAT
AGTCATTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAA
ACACCGCCTACCGCAGCTGGTACCTAGCCTACAACTATGGCGACCCCGAGAAGGGCGTGCGCTCC
TGGACGCTGCTCACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGA
CATGATGCAAGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCG
CCGAGCTCCTGCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTG
CGGGCCTTCACCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCC
GCCCGCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGC
CGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTGC
CCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTAAAT
GTCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGATGTACG
GAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGG
GGCGCCCTCAAGGGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGC
CGACGCGCGCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCG
TGGTGGCCGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCAC
CGGAGCACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAG
GGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCG
CGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGGAACGTGTACTGG
GTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCCCCTCGCACTTGAAGATG
TTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCCAAGCGCAAATTCAAGGAAGA
GATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGAAGGAGGAAAGAAAGCCCC
GCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTGATGTGGACGGATTGGTGGAG
TTTGTGCGCGAGTTCGCCCCCCGGCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCGGTGCT
GAGACCCGGCACCACCGTGGTCTTCACGCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCCT
ACGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGCT
TACGGCAAGCGCAGCCGTTCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAA
CCCCACGCCGAGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGG
GGTTCAAGCGCGAGGGCGAGGATCTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAG
CTGGAAGACGTGCTGGAGACCATGAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCC
CATCAAGCAGGTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACATCAAGATTGCCACGGAGCCCA
TGGAAACGCAGACCGAGCCCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGACGGATCCC
TGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTGCTGATGCC
CAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGCTTCTACCGCG
GTCATACCAGCAGCCGCCGCCGCAAGACCACCACTCGCCGCCGCCGTCGCCGCACCGCCGCTGCA
ACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCTGCCGCG
CGCGCGCTACCACCCGAGCATCGCCATTTAAACTTTCGCCTGCTTTGCAGATCAATGGCCCTCAC
ATGCCGCCTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGG
GGAACGGGATGCGTCGCCACCACCACCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGAGGC
TTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGT
GGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACTTGGAAACATCTTGTAATAAACCaATGGAC
TCTGACGCTCCTGGTCCTGTGATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTTCGTCCCT
GGCTCCGCGACACGGCACGCGGCCGTTCATGGGCACCTGGAGCGACATCGGCACCAGCCAACTGA
ACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGAATTTCGGGTCCACGCTTAAA
ACCTATGGCACCAAGGCGTGGAACAGCACCACAGGGCAGGCGCTGAGGGATAAGCTGAAAGAGCA
GAACTTCCAGCAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGGGGTGGTGGACCTGGCCA
ACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCCGCCCGCCGGCTCCGTGGAG
ATGCCGCAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCCCGA
TGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCCCCCCCGTACGAGGAGGCGGTGAAACTGG
GTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCG
ACCCTGGACTTGCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCCGCCGGT
GGCCGTGGCCCCGCGCGCGACCCGGGGGCACCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGA
ACAGCATCGTGGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGC
GCTTAACTTGCTTGTCTGTGTGTGTATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGG
AGTGAAGAGGCGCGTCGCCGAGTTGCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTAC
ATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTGCCCGCGC
CACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCACGATG
TGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGACAACACC

TACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACATGGCCAGCAC
CTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTCCGGCACCGCCT
ACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAAAGCCGATGGTGAA
ACTGCCACAGAAAAAACCTATACATATGGAAATGCACCCGTGCAGGGCATTAACATCACAAAAGA
TGGTATTCAACTTGGAACTGACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTG
AACCTCAACTGGGTGATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGA
GCTCTTAAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGA
AGGAGGTCAGGCAAATGTGAAAACAGGAACAGGCACTACTAAAGAATATGACATAGACATGGCTT
TCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAATTGTTTTGTATACTGAAAAT
GTGGATTTGGAAACTCCAGATACCCATATTGTATACAAAGCAGGCACAGATGACAGCAGCTCTTC
TATTAATTTGGGTCAGCAAGCCATGCCCAACAGACCTAACTACATTGGTTTCAGAGACAACTTTA
TCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCAGGCTTCTCAGCTG
AATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCCTACCAGCTCTTGCTTGACTCTCT
GGGTGACAGAACCCGGTATTTCAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTGC
GCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTT
GGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACTGATCAAACCACATGGACCAAAGA
TGACAGTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACATCC
AAGCCAACCTGTGGAGGAACTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAG
TACACGCCGGCCAATGTTACCCTGCCCACCAACACCAACACCTACGATTACATGAACGGCCGGGT
GGTGGCGCCCTCGCTGGTGGACTCCTACATCAACATCGGGGCGCGCTGGTCGCTGGATCCCATGG
ACAACGTGAACCCCTTCAACCACCACCGCAATGCGGGGCTGCGCTACCGCTCCATGCTCCTGGGC
AACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCAAGAGCCTCCT
GCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACATGATCCTGCAGA
GCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCAGCATCAACCTCTAC
GCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAACGACAC
CAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAACG
CCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCTTCCGCGGCTGGTCCTTCACG
CGTCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGG
CTCCATCCCCTACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCT
TCGACTCCTCCGTCAGCTGGCCCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAAG
CGCACCGTCGACGGCGAGGGCTACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGT
CCAGATGCTGGCCCACTACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACC
GCATGTACTCCTTCTTCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTCAACTAC
AAGGACTACCAGGCCGTCACCCTGGCCTACCAGCACAACAACTCGGGCTTCGTCGGCTACCTCGC
GCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCG
CCGTCACCAGCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCC
AGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGC
CCACGCGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCT
TCGAAGTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTG
CGCACCCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCG
GGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCTGGG
CACCTTCGATAAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACA
CGGCCGGCCGCGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGCTCGAACACC
TGCTACCTCTTCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACCAGTTCGAGTA
CGAGGGCCTGCTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTGCGTCACCCTGGAAAAGTCCA
CCCAGACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTGCATGTTCCTGCACGCC
TTCGTGCACTGGCCCGACCGCCCCATGGACAAGAACCCCACCATGAACTTGCTGACGGGGGTGCC
CAACGGCATGCTCCAGTCCCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTCTACC
GCTTCCTCAACTCCCACTCCCCCTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCACCGCC
TTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCAC
TTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTTCTGCCGGGTCTCGG
CATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACTCGGGGATC
AGCAGTTTGGGCAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGC
GCCCAGCAGGTCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGT
TGCGGTACACGGGGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACC
GTCGCGTCGGTGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCATCCCGAAGGGGGTCATCTT
GCAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGA

TCAGCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACATGGCCTTCATGAAAGCCTCCAAT
TGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCCGCAGGACTTGCTAGAGAA
CTGGTTGGTGGCGCACCCGGCGTCGTGCACGCAGCAGCGCGCGTCGTTGTTGGCCAGCTGCACCA
CGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTCGGGGTTCTCCTTCAGCGCGCGCTGC
CCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAG
GCACCGCAGCTTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCC
AGTTCTTGTGGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTG
GTCAGGGTCTTGTTGCTAGTGAAGGTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTG
GCAGATGCGGCGGTACACCTCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCT
CCACGCGGTAGCGGTCCATCAGCATAGTCATGATTTCCATACCCTTCTCCCAGGCCGAGACGATG
GGCAGGCTCATAGGGTTCTTCACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCTC
GTCCAGGGTCTCAAAGCTGCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGTAGCTGAAGC
CCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCTGCAGG
ACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGTTGGAGATGG
CGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCGAGGCCACGC
GGCGGTAGGTATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCTCGCCGCCGCGACTTGGC
GGATGGCTGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCC
TCCGCGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAA
CCTCGCCATCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCG
CCGCCCAGCCCCGCCACCTCCGACGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGA
GATTGACCTGGGCTATGTGACGCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTTCACAAG
AAGAGATACACCAAGAACAGCCAGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAG
CATGACGGCGACTACCTCCACCTGAGCGGGGGGGAGGACGCGCTCATCAAGCATCTGGCCCGGCA
GGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCA
GCCGCGCCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCAAGCGCCAGCCCAATGGCACC
TGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTA
CCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCACCCGCGCCGACG
CCCTTTTCAACCTGGGTCCCGGCGCCCGCCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAAG
ATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCTCTGCAAGGAGAAGGAGGAGA
GCATGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAAC
GCACGGTCGAGCTGACCCATTTCGCCTACCCGGCTCTGAACCTGCCCCCCAAAGTCATGAGCGCG
GTCATGGACCAGGTGCTCATCAAGCGCGCGTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTC
CGAGGAGGGCAAGCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCCTAATGCTAGTC
CCCAGAGTTTGGAAGAGCGGCGCAAACTCATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAG
TGCCTGCGCCGCTTCTTCGCCGACGCGGAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCT
CTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCT
CCTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCGGG
GAGGCCCGGCGCGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGACGGG
CATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCCTGCAGA
AGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACCTC
ATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGCAT
GTTGCAAAACTTTCGCTCTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGC
TGCCCTCGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTGTGGAGCCACTGCTAC
CTGCTGCGCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTCAGCGGCGAGGG
CCTGCTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCC
AGCTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGCAAGGGCCCAGCGAAGGCGAGGGT
TCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTGTGGACCTCGGCCTACTTGCGCAAGTT
CGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCATCCGCCCAAGG
CCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCCATCCAGAAA
TCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAGACCGGTGAGGA
GCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAGCTGCCGCCC
GTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGA
CTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGG
CAGAGGAGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCTCGGCGGGGGAGAAAGCAAGC
AGCACGGATACCATCTCCGCTCCGGGTCGGGGTCCCGCTCGACCACACAGTAGATGGGACGAGAC
CGGACGATTCCCGAACCCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCCTGGG
GGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTCACCCGG

CGCTACCTGCTCTTCCACCCCGGGGTGAACTTTCCCCGCAACATCTTGCATTACTACCGTCACCT
CCACAGCCCCTACTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAAAGACCAGCAGAAAACCAGCA
GCTAGAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAAAC
CCGGGAGCTGAGGAACCGGATCTTTCCCACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGG
AGCAGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAG
AGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCT
CACTCTTAAAGAGTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGTGC
CCTTCGCCCTAGCCGCCTCCACCCATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTGGAG
CTACCAGCCCCAGATGGGCCTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGAATTGGC
TCAGCGCCGGGCCCGCGATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGATACTC
CTAGAACAGTCAGCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGC
CCTGGTGTACCAGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAG
TCCAGCTGACTAACTCAGGTGTCCAGCTGGCGGGGGGCGCCACCCTGTGTCGTCACCGCCCCGCT
CAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGAGCTC
TTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTCCTTCACGC
CTCGTCAGGCCGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGTGGCATCGGCACT
CTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGGCCA
CTACCCGGACGAGTTCATCCCGAACTTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATTGAA
TGTCCCATGGTGGCGCAGCTGACCTAGCTCGGCTTCGACACCTGGACCACTGCCGCCGCTTCCGC
TGCTTCGCTCGGGATCTCGCCGAGTTTGCCTACTTTGAGCTGCCCGAGGAGCACCCTCAGGGCCC
GGCCCACGGAGTGCGGATCGTCGTCGAAGGGGGCCTCGACTCCCACCTGCTTCGGATCTTCAGCC
AGCGTCCGATCCTGGTCGAGCGCGAGCAAGGACAGACCCTTCTGACTCTGTACTGCATCTGCAAC
CACCCCGGCCTGCATGAAAGTCTTTGTTGTCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCA
GCGACTACTCCGGACTTCCGTGTGTTCCTGAATCCATCAACCAGTCTTTGTTCTTCACCGGGAAC
GAGACCGAGCTCCAGCTCCAGTGTAAGCCCCACAAGAAGTACCTCACCTGGCTGTTCCAGGGCTC
CCCGATCGCCGTTGTCAACCACTGCGACAACGACGGAGTCCTGCTGAGCGGCCCTGCCAACCTTA
CTTTTTCCACCCGCAGAAGCAAGCTCCAGCTCTTCCAACCCTTCCTCCCCGGGACCTATCAGTGC
GTCTCGGGACCCTGCCATCACACCTTCCACCTGATCCCGAATACCACAGCGTCGCTCCCCGCTAC
TAACAACCAAACTAACCTCCACCAACGCCACCGTCGCGACGGCCACAATACATGCCCATATTAGA
CTATGAGGCCGAGCCACAGCGACCCATGCTCCCCGCTATTAGTTACTTCAATCTAACCGGCGGAG
ATGACTGACCCACTGGCCAACAACAACGTCAACGACCTTCTCCTGGACATGGACGGCCGCGCCTC
GGAGCAGCGACTGGCCCAACTTCGCATTCGCCAGCAGCAGGAGAGAGCCGTCAAGGAGCTGCAGG
ATGCGGTGGCCATCCACCAGTGCAAGAGAGGCATCTTCTGCCTGGTGAAACAGGCCAAGATCTCC
TACGAGGTCACTCCAAACGACCATCGCCTCTCCTACGAGCTCCTGCAGCAGCGCCAGAAGTTCAC
CTGCCTGGTCGGAGTCAACCCCATCGTCATCACCCAGCAGTCTGGCGATACCAAGGGGTGCATCC
ACTGCTCCTGCGACTCCCCCGACTGCGTCCACACTCTGATCAAGACCCTCTGCGGCCTCCGCGAC
CTCCTCCCCATGAACTAATCACCCCCTTATCCAGTGAAATAAAGATCATATTGATGATGATTTTA
CAGAAATAAAAAATAATCATTTGATTTGAAATAAAGATACAATCATATTGATGATTTGAGTTTAA
CAAAAAAATAAAGAATCACTTACTTGAAATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAAC
ACCACTTCACTCCCCTCTTCCCAGCTCTGGTACTGCAGGCCCCGGCGGGCTGCAAACTTCCTCCA
CACGCTGAAGGGGATGTCAAATTCCTCCTGTCCCTCAATCTTCATTTTATCTTCTATCAGATGTC
CAAAAAGCGCGTCCGGGTGGATGATGACTTCGACCCCGTCTACCCCTACGATGCAGACAACGCAC
CGACCGTGCCCTTCATCAACCCCCCCTTCGTCTCTTCAGATGGATTCCAAGAGAAGCCCCTGGGG
GTGTTGTCCCTGCGACTGGCCGACCCCGTCACCACCAAGAACGGGGAAATCACCCTCAAGCTGGG
AGAGGGGGTGGACCTCGATTCCTCGGGAAAACTCATCTCCAACACGGCCACCAAGGCCGCCGCCC
CTCTCAGTTTTTCCAACAACACCATTTCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGA
AAATTATCCTTACAAGTTTCTCCACCATTAAATATACTGAGAACAAGCATTCTAAACACACTAGC
TTTAGGTTTTGGATCAGGTTTAGGACTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCAC
TTACATTTGATACTGATGGAAACATAAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGA
GATGCAATTGAAAGCAACATAAGCTGGGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAAC
CAACATTGGAAATGGGTTAGAGTTTGGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACC
CAATCCAAGTTAAACTTGGATCTGGCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAAC
AAAGAAGACGATAAACTCACTTTGTGGACAACACCTGATCCATCACCAAACTGTCAAATACTCGC
AGAAAATGATGCAAAACTAACACTTTGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGT
CAGTCTTAGTTGTAGGAAGTGGAAACCTAAACCCCATTACTGGCACCGTAAGCAGTGCTCAGGTG
TTTCTACGTTTTGATGCAAACGGTGTTCTTTTAACAGAACATTCTACACTAAAAAAATACTGGGG
GTATAGGCAGGGAGATAGCATAGATGGCACTCCATATACCAATGCTGTAGGATTCATGCCCAATT

TAAAAGCTTATCCAAAGTCACAAAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATG
AATGGAGATGTTTCAAAACCTATGCTTCTCACTATAACCCTCAATGGTACTGATGACAGCAACAG
TACATATTCAATGTCATTTTCATACACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGG
CTAACTCTTATACCTTCTCATACATCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAAC
CCTTCCCACCCCACTCTGTGGAACAAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTA
TTGATTCAACAGTTTTACAGGATTCGAGCAGTTATTTTTCCTCCACCCTCCCAGGACATGGAATA
CACCACCCTCTCCCCCCGCACAGCCTTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGG
TCTCCACGTTCCACACAGTTTCAGAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCC
GGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCTGAGGATTGTCCTCGGTGGTCGGGATCAC
GGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCGGTG
GTGTCGCATCAGGCCCCGCAGCAGTCGCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCAGGGGGT
CCGGGTCCAGGGACTCCCTCAGCATGATGCCCACGGCCCTCAGCATCAGTCGTCTGGTGCGGGGG
CGCAGCAGCGCATGCGGATCTCGCTCAGGTCGCTGCAGTACGTGCAACACAGAACCACCAGGTTG
TTCAACAGTCCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGGAAGGATGCTACCCACGTG
GCCGTCGTACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACA
TGATCTCCTTGGGCATGTGGCGGTTCACCACCTCCCGGTACCACATCACCCTCTGGTTGAACATG
CAGCCCCGGATGATCCTGCGGAACCACAGGGCCAGCACCGCCCCGCCCGCCATGCAGCGAAGAGA
CCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGA
ACAAGTCTATGTTGGCACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCG
GGGGTCAAAACCATATGCCAGGGCACGGGGAACTCTTGCAGGACAGCGAACCCCGCAGAACAGGG
CAATCCTCGCACAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGAT
CCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCGGCCGATACGGG
TGATGGCGGGACGCGGCTGATCGTGTTCGCGACCGTGTCATGATGCAGTTGCTTTCGGACATTTT
CGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCGGCGGCGGTCTCGGCGC
TTGGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACCGTGCAGCAGATCTAGGGC
CTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACATCGACCACCGTGGAATGGG
CCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTTCGGTGACGGCGGGGGAGGGAAGAACAGGA
AGAACCATGATTAACTTTTAATCCAAACGGTCTCGGAGTACTTCAAAATGAAGATCGCGGAGATG
GCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCAGGTCAAAGGTGATACGGTTCTCGA
GATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGCGCACATCCAGAAACAAGACAATAGCGAAA
GCGGGAGGGTTCTCTAATTCCTCAATCATCATGTTACACTCCTGCACCATCCCCAGATAATTTTC
ATTTTTCCAGCCTTGAATGATTCGAACTAGTTCGTGAGGTAAATCCAAGCCAGCCATGATAAAGA
GCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCACACCCTCATAATTCCAAGATATTCTGCT
CCTGGTTCACCTGCAGCAGATTGACAAGCGGAATATCAAAATCTCTGCCGCGATCCCTGAGCTCC
TCCCTCAGCAATAACTGTAAGTACTCTTTCATATCCTCTCCGAAATTTTTAGCCATAGGACCACC
AGGAATAAGATTAGGGCAAGCCACAGTACAGATAAACCGAAGTCCTCCCCAGTGAGCATTGCCAA
ATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGGTGATATCTTCCAGATAACTGGACAGAAAA
TCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGAAAAATCCTCCAGGTGGACGTTTAGAGCCTC
GGGAACAACGATGAAGTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGAA
AAAACAAAAATGAACATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGC
ACCAGGCAGGCCACGGGGTCTCCGGCGCGACCCTCGTAAAAATTGTCGCTATGATTGAAAACCAT
CACAGAGAGACGTTCCCGGTGGCCGGCGTGAATGATTCGACAAGATGAATACACCCCCGGAACAT
TGGCGTCCGCGAGTGAAAAAAAGCGCCCGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAG
TCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATTACT
CCCCTCCTGCACAGGCAGCAAAGCCCCCGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCA
TAGCTTACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTC
CACCCGCTCTCTGCTCAATATATAGCCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATA
CCCGCCAAATAATCACACACGCCCAGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACG
CGCACTTCCTCAAACGCCCAAAACTGCCGTCATTTCCGGGTTCCCACGCTACGTCATCAAAACAC
GACTTTCAAATTCCGTCGACCGTTAAAAACGTCACCCGCCCCGCCCCTAACGGTCGCCCGTCTCT
CAGCCAATCAGCGCCCCGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAA
GTTTGAGGTATATTATTGATGATGG ChAdV68.5WTnt.GFP (SEQ ID NO: 13); AC_000011.1 with E1 (nt 577 to 3403) and E3 (nt 27,125-31,825) sequences deleted; corresponding ATCC VR-594 nucleotides substituted at five positions; GFP reporter under the control of the CMV promoter/enhancer inserted in place of deleted E1
CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTT
GGGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGGGGGCGAGTGACGTTTTG

ATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGGT
GTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTTTCTGGG
CGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAATCTGAGTAA
TTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGATTACGTGGG
GGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGTGTTTTTACG
TAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCCACTCTTGAGT
GCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAAGTAGGGATAAC
AGGGTAATgacattgattattgactagttGttaaTAGTAATCAATTACGGGGTCATTAGTTCATA
GCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAAC
CACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA
TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATA
TGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCCCTGGCATTATGCCCAGTAC
ATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGgT
GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTC
TCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT
CGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG
CAGAgcTCGTTTAGTGAACCGTCAGATCGCCTGGAACGCCATCCACGCTGTTTTGACCTCCATAG
AAGACAGCGATCGCGecacCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGGGTGCCCATCC
TGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCC
CACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGC
AGCACGACTTCTTCAAGTCCCCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAG
GACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACT
ACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAG
ATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCAT
CGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAG
ACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTTTACAAGTAGtgaGTTTAAACTCCCATTTAAATGTGAGGGTTAATGCTTCG
AGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAAT
GCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAA
GTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTA
AAGCAAGTAAAACCTCTACAAATGTGGTAAAATAACTATAACGGTCCTAAGGTAGCGAGTGAGTA
GTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGT
GTTGCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGG
CGTCTCCCCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGT
GCAGCCCGCGAACTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTG
CCGCCGCAGCTGCTGCATCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTAC
GGCACTCTGGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTT
GCTGCTGATGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTC
AGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAATAA
ATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTCGCGCGCGG
TAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAGGACCCGGTAGAG
GTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCTCCATTGCA
GGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATGGTGT
TGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAA
TCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGAT
TGGCGATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACCACCAGCACGGTGTAT
CCGGTGCACTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACGCC
TTTGTGCCCGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCGATGGGCCCGTGGGCGGCGG
CCTGGGCAAAGACGTTTCGGGGGTCGGACACATCATAGTTGTGGTCCTGGGTGAGGTCATCATAG
GCCATTTTAATGAATTTGGGGCGGAGGGTGCCGGACTGGGGGACAAAGGTACCCTCGATCCCGGG
GGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTGAGCTCGGAGGGGGGGATCATGTCCA
CCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGCCGAAAGCAAGTTC
CGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGATGACCGGCTGCAGGTG
GTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCCACCTCGTTCATCATCTCGC
GCACGTGCATGTTCTCGCGCACCAGTTCCGCCAGGAGGCGCTCTCCCCCCAGGGATAGGAGCTCC

TGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGCCATGGGCATTTTGGAGAGGGTTTG
TTGCAAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGAC
CTCCTCGTTTCGCGGGTTGGGACGGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGCAG
CCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTCACGGTGAAGGGG
TGCGCGCCGGGCTGGGCGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAAAACCGCTC
CCGATCGGCGCCCTGCGCGTCGGCCAGGTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCTCGG
CCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGGGAC
TTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGGGGCGTAGGCGTCCGCGCCGCAGTG
GGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCAGTT
TCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGGGTG
ACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCCGCG
GTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAAGG
AGGCCACGTGGGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAAA
CACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACCGGG
GGTCCCGGCCGGGGGGGTATAAAAGGGGGGGGTCCCTGCTCGTCCTCACTGTCTTCCGGATCGCT
GTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCA
GGTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAG
AGCCCCTCGTCCATCTGGTCAGAAAAGACGATCTTTTTGTTGTCGAGCTTGGTGGCGAAGGAGCC
GTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTGGTTTTTTTCCTTGTCGGCGC
GCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCCACGCACTTCCATTCGGGGAAGACG
GTGGTCAGCTCGTCGGGCACGATTCTGACCTGCCAGCCCCGATTATGCAGGGTGATGAGGTCCAC
ACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGCCCTTGCGCGAGC
AGAAGGGGGGCAGGGGGTCCAGCATGACCTCGTCGGGGGGGTCGGCATCGATGGTGAAGATGCCG
GGCAGGAGGTCGGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCATTC
GCGCACGGCCAGCGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCG
CGGAGGCGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTG
GGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAG
GAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGG
CATGCGAGTTGGAGGAGATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACC
GAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGCAGCTTGGCGACGAGCTCGGCGGTGACTAGGAC
GTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATGATGTCATACTTGAGCTGTCCCTTTTGTTTCC
ACAGCTCGCGGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCGAGGGGGAACCCGTCC
TGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTGTAGGCGCAGCAGCCCTT
CTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGT
CCCTGACCATGACCTTGAGGAACTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCCCAG
AGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGAG
GATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGT
TGTTGATGACCTGGGCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCCACGATGTAG
AGTTCCACGAATCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTC
GTCGGGGTCGCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGA
AGGAAGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGTACTGACGGAACTGCTGC
CCGACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCCCCGTGCCAGCGATCCCA
TTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCGGTCGTCCCCGGAGAGTTTCATGA
CCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCAGGTGTAGGTTTCCACATCGTAG
GTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACTGGATCTCCTGCCACCA
ATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAACACTCGTGCT
TGTGTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACGAGCTGT
ACCTGAGTTCCTTTGACGAGGAATTTCAGTGGGAAGTGGAGTCGTGGCGCCTGCATCTCGTGCTG
TACTACGTCGTGGTGGTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTCATGCTGACGAGCCCGC
GCGGGAGGCAGGTCCAGACCTCGGCGCGAGCGGGTCGGAGAGCGAGGACGAGGGCGCGCAGGCCG
GAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGTTGAC
TTGCAGGAGTTTTTCCAGGGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGG
TGGCGACGTCGATGGCTTGCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTCCCCCGTTTCTTC
TTGGGCGGCTGGGGCGACGGGGGGGGTGCCTCTTCCATGGTTAGAAGCGGCGGCGAGGACGCGCG
CCGGGCGGCAGGGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTCGGCGCCGCGCGC
GGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACGTCCT
GGATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTGAAAGAGAGTTCGACA

GAATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCTTGCACGTCGCCCGAGTTGTC
CTGGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCCGGCGC
GCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTCATGCCC
GCCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCgGGCGCGCATGACCACCTG
GGCGAGGTTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGT
TGAGCGTGGTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGGAGCGGCATCTCG
CTGACGTCGCCCAGCGCCTCCAAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAAA
CTGGGAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGATGAGCTCGGCGATGGTGG
CGCGCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCCTCTTCTTCCTCCTCCACTAAC
ATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAGGGGGCCTGCGTCGCCGGCGGCG
CACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCGGCGTCGCATGGTCTCGGTGA
CGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAGACGCCGCCGCGCATCTCCAGGTGGCCGGGG
GGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCCCGTAGGGACTCC
GCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGAAGGCTTCGAGCC
AGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGGGGGTCATGTTGGTTGGGAGCGGGG
GGGGCGATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAG
CACCAGGTCTTTGGGCCCGGCTTGCTGGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGGTCCT
GACACCTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCGCTCCACGGGCACCTCCTCCTCGCCC
GCGCGGCCGTGCATGCGCGTGAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTCGGCGAC
GACGCGCTCGGCGAGGATGGCTTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGTCGA
CGAAGCGGTGGTAGGCTCCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACGGACCAGTTGACG
GTCTGGTGGCCCGGACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGAAGAT
GTAGTCGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGGCGGT
AGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTGGTAG
CCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGAACTCGCG
GACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTGGCCCGTGA
GGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGGCTCGACTCCGTG
GCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAATCTCGAATCAGGCTG
GAGCCGCAGCTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAGGAT
ACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCGGATGAGACTAGTAAGCGCGGAAAGCG
GCCGACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCC
CCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAGACC
CCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCTCTTTTGTTTTGTTTGTTTTTGCCAG
ATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTCCACCGCAACAACAGCCCCCTCCACA
GCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGACCGCCGCGGCCGCCGTGAGCG
GGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTGGCGCGCCTGGGG
GCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGAGGCCTACGTGCC
CAAGCAGAACCTGTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGATGCGCGCGGCCCGGTTCC
ACGCGGGGCGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGTGCTGAGGGACGAGGATTTCGAG
GCGGACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCACGGC
GTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGCACCC
TGATCGCGCGCGAGGAGGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTGGAGGCCATCGTG
CAGAACCCCACCAGCAAGCCGCTGACGGCGCAGCTGTTCCTGGTGGTGCAGCATAGTCGGGACAA
CGAAGCGTTCAGGGAGGCGCTGCTGAATATCACCGAGCCCGAGGGCCGCTGGCTCCTGGACCTGG
TGAACATTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGCTGCCGCTGTCCGAGAAGCTGGCGGCC
ATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGGAAGATCTACAAGACCCCGTACGT
GCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTGACCC
TGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAGCGCCAGCAGGCGG
CGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGGGACCGA
GGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCCTTGGAGG
CGGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGAA
GACTGATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGCGATG
CGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCATGCA
ACGCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGGC
TCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCCCACGCACGAGAAGGTCCTGGCC
ATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCGACGAGGCCGGCCTGGTGTACAACGC
GCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAACGTGCAGACCAACCTGGACCGCATGGTGA

CCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCACCGCGAGTCCAACCTGGGATCCATG
GTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCGCCAACGTGCCCCGGGGCCAGGAGGACTACAC
CAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGGTGTACCAGTCCG
GGCCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCGTGAACCTGAGCCAGGCTTTC
AAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCCCCGGTCGGGGACCGCGCGACGGTGTCGAGCCT
GCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGGCAGCATCA
ACCGCAACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTG
GACGAGCAGACCTACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGGCAA
CCTGGAAGCCACCCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGATCCCCCCCCCAGTACGCGC
TCAGCACCGAGGAGGAGCGCATCCTGCGTTACGTGCAGCAGAGCGTGGGCCTGTTCCTGATGCAG
GAGGGGGCCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGGAGCCCAGCATGTACGC
CAGCAACCGCCCGTTCATCAATAAACTGATGGACTACTTGCATCGGGCGGCCGCCATGAACTCTG
ACTATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCGCCGGGGTTCTACACGGGCGAG
TACGACATGCCCGACCCCAATGACGGGTTCCTGTGGGACGATGTGGACAGCAGCGTGTTCTCCCC
CCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCCCGTCCTCGG
CGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCTTTCCCGAGCTTG
CCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGG
CGAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACG
GGATAGAAAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGAT
CCCCGGGCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCGGTGGCACGACAG
GCAGCGGGGACAGATGTGGGACGATGAGGACTCCGCCGACGACAGCAGCGTGTTGGACTTGGGTG
GGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGGCGCATGATGTAAGAGAAACCGAAA
ATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCTGTTGTTGTTGTA
TCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTCCCTCGTACGAGAGCGTGATGCAGCA
GGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTACGTGCCCCCGCGGTACCTGG
CGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTACGATACCACCCGG
TTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACGACCACAGCAA
CTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCACCCAGACCATCA
ACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCCAAC
GTGAACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAA
TGGGGTGACAGTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAAT
TTGAGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATC
GACAATTACTTGGCGGTGGGGCGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCGA
CACTAGGAACTTCAGGCTGGGCTGGGACCCCGTGACCGAGCTGGTCATGCCCGGGGTGTACACCA
ACGAGGCTTTCCATCCCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTCACCGAGAGCCGC
CTCAGCAACCTGCTGGGCATTCGCAAGAGGCAGCCCTTCCAGGAAGGCTTCCAGATCATGTACGA
GGATCTGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTCGACGCCTATGAGAAAAGCAAGGAGG
ATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTCAGGGGCGATAATTTT
GCAAGCGCCGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCATTCAGCC
GGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTACC
GCAGCTGGTACCTAGCCTACAACTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCTC
ACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATGCAAGA
CCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCTGC
CCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTCACC
TCGCTTACGCACGTCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCCGCGCCCAC
CATTACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGCCGCTGCGCAGCA
GTATCCGGGGAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTGCCCCTACGTCTAC
AAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTAAATGTCCATTCTCAT
CTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGATGTACGGAGGCGCTCGCC
AACGCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGGGGCGCCCTCAAG
GGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGCGCGCAA
CTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCGTGGTGGCCGACG
CGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACCGGAGCACCCCC
GCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTCAG
GGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGG
CGGCAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCC
GCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCCCCTCGCACTTGAAGATGTTCACTTCGCGA

TGTTGATGTGTCCCAGCGGCGAGGAGGATGTCCAAGCGCAAATTCAAGGAAGAGATGCTCCAGGT
CATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGAAGGAGGAAAGAAAGCCCCGCAAAATCAAGC
GGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTGATGTGGACGGATTGGTGGAGTTTGTGCGCGAG
TTCGCCCCCCGGCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCGGTGCTGAGACCCGGCAC
CACCGTGGTCTTCACGCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCCTACGACGAGGTGT
ACGGGGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGCTTACGGCAAGCGC
AGCCGTTCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCCACGCCGAG
CCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGGTTCAAGCGCG
AGGGCGAGGATCTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTG
CTGGAGACCATGAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGT
GGCCCCGGGCCTGGGCGTGCAGACCGTGGACATCAAGATTCCCACGGAGCCCATGGAAACGCAGA
CCGAGCCCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCG
GCTCCTAGTCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTGCTGATGCCCAACTACGCGCT
GCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGCTTCTACCGCGGTCATACCAGCA
GCCGCCGCCGCAAGACCACCACTCGCCGCCGCCGTCGCCGCACCGCCGCTGCAACCACCCCTGCC
GCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCTGCCGCGCGCGCGCTACCA
CCCGAGCATCGCCATTTAAACTTTCGCCTGCTTTGCAGATCAATGGCCCTCACATGCCGCCTTCG
CGTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACGGGATGC
GTCGCCACCACCACCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGAGGCTTCCTGCCCGCG
CTGATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGC
CTCTCAGCGCCACTGAGACACACTTGGAAACATCTTGTAATAAACCaATGGACTCTGACGCTCCT
GGTCCTGTGATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTTCGTCCCTGGCTCCGCGACA
CGGCACGCGGCCGTTCATGGGCACCTGGAGCGACATCGGCACCAGCCAACTGAACGGGGGCGCCT
TCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGAATTTCGGGTCCACGCTTAAAACCTATGGCAGC
AAGGCGTGGAACAGCACCACAGGGCAGGCGCTGAGGGATAAGCTGAAAGAGCAGAACTTCCAGCA
GAAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGGGGTGGTGGACCTGGCCAACCAGGCCGTGC
AGCGGCAGATCAACAGCCGCCTGGACCCGGTGCCGCCCGCCGGCTCCGTGGAGATGCCCCAGGTG
GAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCCCGATGCGGAGGAGAC
GCTGCTGACGCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAACTGGGTCTGCCCACCA
CGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCTGGACTTG
CCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCG
CGCGCGACCCGGGGGCACCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGG
GTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCT
TGTCTGTGTGTGTATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCG
CGTCGCCGAGTTGCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCC
GGACAGGACGCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTGCCCGCGCCACAGACACCTA
CTTCAGTCTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCACGATGTGACCACCGACC
GCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGACAACACCTACTCGTACAAA
GTGCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACATGGCCAGCACCTACTTTGACAT
CCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTCCGGCACCGCCTACAACAGTCTGG
CCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAAAGCCGATGGTGAAACTGCCACAGAA
AAAACCTATACATATGGAAATGCACCCGTGCAGGGCATTAACATCACAAAAGATGGTATTCAACT
TGGAACTGACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGTGG
GTGATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGCCT
GATACCAAAATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTCAGGC
AAATGTGAAAACAGGAACAGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTGACAACA
GAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGATTTGGAA
ACTCCAGATACCCATATTGTATACAAAGCAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGG
TCAGCAAGCCATGCCCAACAGACCTAACTACATTGGTTTCAGAGACAACTTTATCGGGCTCATGT
ACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCAGGCTTCTCAGCTGAATGCTGTGGTT
GACTTGCAAGACAGAAACACCGAGCTGTCCTACCAGCTCTTGCTTGACTCTCTGGGTGACAGAAC
CCGGTATTTCAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTGCGCATTATTGAAA
ATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGCAGAACAGAT
ACTTATCAGGGAATTAAGGCTAATGGAACTGATCAAACCACATGGACCAAAGATGACAGTGTCAA
TGATGCTAATGAGATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACATCCAAGCCAACCTGT
GGAGGAACTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAGTACACGCCGGCC
AATGTTACCCTGCCCACCAACACCAACACCTACGATTACATGAACGGCCGGGTGGTGGCGCCCTC

GCTGGTGGACTCCTACATCAACATCGGGGCGCGCTGGTCGCTGGATCCCATGGACAACGTGAACC
CCTTCAACCACCACCGCAATGCGGGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTAC
GTGCCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCGG
GTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACATGATCCTGCAGAGCTCCCTCGCCA
ACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCAGCATCAACCTCTACGCCACCTTCTTC
CCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAACGACACCAACGACCAGTC
CTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAACGCCACCAACGTGC
CCATCTCCATCCCCTCGCGCAACTGGGCCGCCTTCCGCGGCTGGTCCTTCACGCGTCTCAAGACC
AAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCCTA
CCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCG
TCAGCTGGCCCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAAGCGCACCGTCGAC
GGCGAGGGCTACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGC
CCACTACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCT
TCTTCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAG
GCCGTCACCCTGGCCTACCAGCACAACAACTGGGGCTTCGTCGGCTACCTCGCGCCCACCATGCG
CCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGTCACCAGCG
TCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCAACTTCATG
TCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGCCCACGCGCTAGA
CATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTTCGAAGTCTTCG
ACGTCGTCCGAGTGCACCACCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCGCACCCCCTTC
TCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGGCGAG
CAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCTGGGCACCTTCGATAA
GCGCTTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCG
AGACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGCTCGAACACCTGCTACCTCTTC
GACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCCTGCT
GCGCCGCAGCGCCCTGGCCACCGAGGACCGCTGCGTCACCCTGGAAAAGTCCACCCAGACCGTGC
AGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTGCATGTTCCTGCACGCCTTCGTGCACTGG
CCCGACCGCCCCATGGACAAGAACCCCACCATGAACTTGCTGACGGGGGTGCCCAACGGCATGCT
CCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTCTACCGCTTCCTCAACT
CCCACTCCGCCTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCACCGCCTTCGACCGCATG
AATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCACTTTCATGTTACA
CATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTTCTGCCGGGTCTCGGCATGGCCCGCGG
GCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACTCGGGGATCAGCAGTTTGGGC
AGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTC
GGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACGG
GGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTCGCGTCGGTG
ATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCATCCCGAAGGGGGTCATCTTGCAGGTCTGCCT
TCCCATGGTGGGCACGCACCCGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCT
GGGCCTGGTCGGCGTTCATCCCCGGGTACATGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCC
TGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCCGCAGGACTTGCTAGAGAACTGGTTGGTGGC
GCACCCGGCGTCGTGCACGCAGCAGCGCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCCCC
AGCGGTTCTGGGTGATCTTGGCCCGGTCGGGGTTCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTC
GCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGCTT
GCCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCCAGTTCTTGTGGG
CGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGGTCTTG
TTGCTAGTGAAGGTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCG
GTACACCTCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCTCCACGCGGTAGC
GGTCCATCAGCATAGTCATGATTTCCATACCCTTCTGCCAGGCCGAGACGATGGGCAGGCTCATA
GGGTTCTTCACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCTCGTCCAGGGTCTC
AAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGTAGCTGAAGCCCACGGCCGCCA
GCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCTGCAGGACCACATGCTTG
GTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGTTGGAGATGGCGAGGGGGAGCG
CGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCGAGGCCACGCGGCGGTAGGTAT
GTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCTCGCCGCCGCGACTTGGGGGATGGCTGGCA
GAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCCTCCGCGGCCGGC
CATTGTGTTCTCCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACCTCGCCATCTG
CCCCCACCGCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCCGCCCAGCCCC

GCCACCTCCGACGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGG
CTATGTGACGCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTTCACAAGAAGAGATACACC
AAGAACAGCCAGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGAC
TACCTCCACCTGAGCGGGGGGGAGGACGCGCTCATCAAGCATCTGGCCCGGCAGGCCACCATCGT
CAAGGATGCGCTGCTCGACCGCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTACG
AGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCAAGCGCCAGCCCAATGGCACCTGCGAGCCCAAC
CCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTACCACATCTTTTT
CAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCACCCGCGCCGACGCCCTTTTCAACC
TGGGTCCCGGCGCCCCCCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAAGATCTTCGAGGGT
CTGGGCAGCGACGAGACTCGGGCCCCGAACGCTCTGCAAGGAGAAGGAGGAGAGCATGAGCACCA
CAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGAGC
TGACCCATTTCGCCTACCCGGCTCTGAACCTGCCCCCCAAAGTCATGAGCGCGGTCATGGACCAG
GTGCTCATCAAGCGCGCGTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAA
GCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCCTAATGCTAGTCCCCAGAGTTTGG
AAGAGCGGCGCAAACTCATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGC
TTCTTCGCCGACGCGGAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCACGG
GTTCGTGCGCCAGGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCA
TCTTGCACGAGAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCGGGGAGGCCCGGCGC
GACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGACGGGCATGGGCGTGTG
GCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCCTGCAGAAGAACCTCAAGG
GTCTGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACCTCATTTTCCCCGAG
CGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGCATGTTGCAAAACTT
TCGCTCTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGGACT
TCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGCGCCTG
GCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTCAGCGGCGAGGGCCTGCTCGAGTG
CCACTGCCGCTGCAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCG
AGACCCAGATCATCGGCACCTTCGAGTTGCAAGGGCCCAGCGAAGGCGAGGGTTCAGCCGCCAAG
GGGGGTCTGAAACTCACCCCGGGGCTGTGGACCTCGGCCTACTTGCGCAAGTTCGTGCCCGAGGA
CTACCATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTCGG
CCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCCATCCAGAAATCCCGCCAAGAA
TTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAGACCGGTGAGGAGCTCAACCCCGG
CTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAGCTGCCCCCCGTGGAGGATTTG
GAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGACTGGGACAGCAC
TCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGGAGGAGG
TGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCTCGGGGGGGGAGAAAGCAAGCAGCACGGATACC
ATCTCCGCTCCGGGTCGGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCCC
GAACCCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCCTGGCGGGGGCACAAAA
ACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCTC
TTCCACCGCGGGGTGAACTTTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCCTA
CTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAAAGACCAGCAGAAAACCAGCAGCTAGAAAATCC
ACAGCGGCGGCAGCAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAG
GAACCGGATCTTTCCCACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGGAGCAGGAACTGA
AAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAGAGCGAAGACCAA
CTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCTCACTCTTAAAGA
GTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGTGCCCTTCGCCCTAG
CCGCCTCCACCCATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTGGAGCTACCAGCCCCA
GATGGGCCTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGC
CCGCGATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGATACTCCTAGAACAGTCA
GCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACCA
GGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGACTA
ACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTCGTCACCGCCCCGCTCAGGGTATAAAG
CGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTCT
GCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTCCTTCACGCCTCGTCAGGCCG
TCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGTGGCATCGGCACTCTCCAGTTCGTG
GAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGGCCACTACCCGGACGA
GTTCATCCCGAACTTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATTGAAACTAATCACCCC
CTTATCCAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGAT

TTGAAATAAAGATACAATCATATTGATGATTTGAGTTTAACAAAAAAATAAAGAATCACTTACTT
GAAATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGC
TCTGGTACTGCAGGCCCCGGGGGCTGCAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCT
CCTGTCCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATG
ACTTCGACCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCC
TTCGTCTCTTCAGATGGATTCCAAGAGAAGCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCC
CGTCACCACCAAGAACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCTCGG
GAAAACTCATCTCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCATT
TCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACC
ATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGAC
TCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATA
AAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTG
GGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTTG
GAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGGC
CTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTG
GACAACACCTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAACTAACACTTT
GCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAGGAAGTGGAAAC
CTAAACCCCATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACGGTGT
TCTTTTAACAGAACATTCTACACTAAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATG
GCACTCCATATACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCTTATCCAAAGTCACAAAGT
TCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTATGCT
TCTCACTATAACCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATACA
CCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACATC
GCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCCTTCCCACCCCACTCTGTGGAACAA
ACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGATTCG
AGCAGTTATTTTTCCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCCGCACAGCCT
TGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCAGAG
CGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCTCACA
GCTCAACAGCTGAGGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCG
GCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCCCCGCAGCAGTC
GCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATG
ATGCCCACGGCCCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATGCGGATCTCGCT
CAGGTCGCTGCAGTACGTGCAACACAGAACCACCAGGTTGTTCAACAGTCCATAGTTCAACACGC
TCCAGCCGAAACTCATCGCGGGAAGGATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTAA
ATCAAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACATGATCTCCTTGGGCATGTGGCGGTT
CACCACCTCCCGGTACCACATCACCCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCGGAACC
ACAGGGCCAGCACCGCCCCCCCCGCCATGCAGCGAAGAGACCCCGGGTCCCGGCAATGGCAATGG
AGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACAG
GCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCATATCCCAGGGCA
CGGGGAACTCTTGCAGGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACTTACATTG
TGCATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTC
GGTCTCCTCACAGCGTGGTAAGGGGGCCGGCCGATACGGGTGATGGCGGGACGCGGCTGATCGTG
TTCGCGACCGTGTCATGATGCAGTTGCTTTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGT
CCGGGCGCTGCACACCGATCGCCGGCGGCGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGT
AAAACAGCCACTCTCTCAGACCGTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCA
TGCCTGATGGCTCTGATCACATCGACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATT
TTGTTGGGTTTCGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCA
AACGGTCTCGGAGTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGG
TGGAAAATAACAGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAA
AGCCTCCACGCGCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAA
TCATCATGTTACACTCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGA
ACTAGTTCCTGAGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGG
CATTCTTAAGCACACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAGATTGAC
AAGCGGAATATCAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACT
CTTTCATATCCTCTCCGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACA
GTACAGATAAACCGAAGTGCTGCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTG
GCTAGACCCGGTGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAAT

CAACAAAAGAAAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAA
GCGGTGCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCAT
GCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGG
CGCGACCCTCGTAAAAATTGTCGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCG
GCGTGAATGATTCGACAAGATGAATACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAAGCG
CCCGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGAT
GAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGCC
CCCGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAA
CAGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAG
CCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCA
GCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGCGCACTTCCTCAAACCCCCAAAACT
GCCGTCATTTCCGGGTTCCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTA
AAAACGTCACCCGCCCCGCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCC
AAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATTGATGATGG

XV.B. ChAd Neoantigen Cassette Delivery Vector Testing

XV.B.1. ChAd Vector Evaluation Methods and Materials

Transfection of HEK293A Cells Using Lipofectamine

[0856] DNA for the ChAdV68 constructs (ChAdV68.4WTnt.GFP, ChAdV68.5WTnt.GFP, ChAdV68.4WTnt.MAG25mer and ChAdV68. 5WTnt.MAG25mer) was prepared and transfected into HEK293A cells using the following protocol.

[0857] 10 μg of plasmid DNA was digested with PadI to liberate the viral genome. DNA was then purified using GeneJet DNA cleanup Micro columns (Thermno Fisher) according to manufacturer's instructions for long DNA fragments, and eluted in 20 ul of pre-heated water; columns were left at 37 degrees for 0.5-1 hours before the elution step.

[0858] HEK293A cells were introduced into 6-well plates at a cell density of 10.sup.6 cells/well 14-18 hours prior to transfection. Cells were overlaid with 1 ml of fresh medium (DMEM-10% hiFBS with pen/strep and glutamate) per well. 1-2 ug of purified DNA was used per well in a transfection with twice the ul volume (2-4 ul) of Lipofectamine2000, according to the manufacturer's protocol. 0.5 ml of OPTI-MEM medium containing the transfection mix was added to the 1 ml of normal growth medium in each well, and left on cells overnight.

[0859] Transfected cell cultures were incubated at 37° C. for at least 5-7 days. If viral plaques were not visible by day 7 post-transfection, cells were split 1:4 or 1:6, and incubated at 37° C. to monitor for plaque development. Alternatively, transfected cells were harvested and subjected to 3 cycles of freezing and thawing and the cell lysates were used to infect HEK293A cells and the cells were incubated until virus plaques were observed.

Transfection of ChAdV68 Vectors into HEK293A Cells Using Calcium Phosphate and Generation of the Tertiary Viral Stock

[0860] DNA for the ChAdV68 constructs (ChAdV68.4WTnt.GFP, ChAdV68.5WTnt.GFP, ChAdV68.4WTnt.MAG25mer, ChAdV68.5WTnt.MAG25mer) was prepared and transfected into HEK293A cells using the following protocol.

[0861] HEK293A cells were seeded one day prior to the transfection at 10.sup.6 cells/well of a 6 well plate in 5% BS/DMEM/1×P/S, 1×Glutamax. Two wells are needed per transfection. Two to four hours prior to transfection the media was changed to fresh media. The ChAdV68.4WTnt.GFP plasmid was linearized with PacI. The linearized DNA was then phenol chloroform extracted and precipitated using one tenth volume of 3M Sodium acetate pH 5.3 and two volumes of 100% ethanol. The precipitated DNA was pelleted by centrifugation at 12,000×g for 5 min before washing 1× with 70% ethanol. The pellet was air dried and re-suspended in 50 μL of sterile water. The DNA concentration was determined using a NanoDrop™ (ThermoFisher) and the volume adjusted to 5 μg of DNA/50 μL.

[0862] 169 μL of sterile water was added to a microfuge tube. 5 μL of 2M CaCl.sub.2 was then added to the water and mixed gently by pipetting. 50 μL of DNA was added dropwise to the CaCl.sub.2 water solution. Twenty six μL of 2M CaCl.sub.2 was then added and mixed gently by pipetting twice with a micro-pipetor. This final solution should consist of 5 μg of DNA in 250 μL of 0.25M CaCl.sub.2. A second tube was then prepared containing 250 μL of 2×HBS (Hepes buffered solution). Using a 2 mL sterile pipette attached to a Pipet-Aid air was slowly bubbled through the 2×HBS solution. At the same time the DNA solution in the 0.25M CaCl.sub.2 solution was added in a dropwise fashion. Bubbling was continued for approximately 5 seconds after addition of the final DNA droplet. The solution was then incubated at room temperature for up to 20 minutes before adding to 293A cells. 250 μL of the DNA/Calcium phosphate solution was added dropwise to a monolayer of 293A cells that had been seeded one day prior at 10.sup.6 cells per well of a 6 well plate. The cells were returned to the incubator and incubated overnight. The media was changed 24h later. After 72 h the cells were split 1:6 into a 6

well plate. The monolayers were monitored daily by light microscopy for evidence of cytopathic effect (CPE). 7-10 days post transfection viral plaques were observed and the monolayer harvested by pipetting the media in the wells to lift the cells. The harvested cells and media were transferred to a 50 mL centrifuge tube followed by three rounds of freeze thawing (at −80° C. and 37° C.). The subsequent lysate, called the primary virus stock was clarified by centrifugation at full speed on a bench top centrifuge (4300×g) and a proportion of the lysate 10-50%) used to infect 293A cells in a T25 flask. The infected cells were incubated for 48h before harvesting cells and media at complete CPE. The cells were once again harvested, freeze thawed and clarified before using this secondary viral stock to infect a T150 flask seeded at 1.5×10.sup.7 cells per flask. Once complete CPE was achieved at 72h the media and cells were harvested and treated as with earlier viral stocks to generate a tertiary stock.

Production in 293F Cells

[0863] ChAdV68 virus production was performed in 293F cells grown in 293 FreeStyle™ (ThermoFisher) media in an incubator at 8% CO.sub.2. On the day of infection cells were diluted to 10.sup.6 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 1×PBS, re-centrifuged and then re-suspended in 20 mL of 10 mM Tris pH7.4. The cell pellet was lysed by freeze thawing 3× and clarified by centrifugation at 4,300×g for 5 minutes.

Purification by CsCl Centrifugation

[0864] 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.

[0865] 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 10° C. The tube was then removed to a laminar flow cabinet and the virus band pulled using an 18 guage 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 2×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 3×, 1h per buffer exchange. The virus was then aliquoted for storage at −80° C.

Viral Assays

[0866] VP concentration was performed by using an OD 260 assay based on the extinction coefficient of 1.1×10.sup.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, 1 mM 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.sup.12VP.

[0867] An infectious unit (IU) titer was calculated by a limiting dilution assay of the viral stock. The virus was initially diluted 100× in DMEM/5% NS/1×PS and then subsequently diluted using 10-fold dilutions down to 1×10.sup.−7. 100 μL of these dilutions were then added to 293A cells that were seeded at least an hour before at 3e5 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 1×PBS 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 1×PBS then blocked in 1×PBS/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 4×with 0.5 mL PBS per well. A HRP conjugated Goat anti-Rabbit antibody (Bethyl Labs, Montgomery Texas) diluted 1000×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% H.sub.2O2. Wells were developed for 5 min prior to counting. Cells were counted under a 10×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 mm.sup.2 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 florescent can be used rather than capsid staining to determine the number of GFP expressing virions per mL.

Immunizations

[0868] C57BL/6J female mice and Balb/c female mice were injected with 1×10.sup.8 viral particles (VP) of ChAdV68.5WTnt.MAG25mer in 100 uL volume, bilateral intramuscular injection (50 uL per leg).

Splenocyte Dissociation

[0869] Spleen and lymph nodes 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 (150 mM NH.sub.4Cl, 10 mM KHCO.sub.3, 0.1 mM Na.sub.2EDTA). Cells were filtered again through a 30 micron filter and then resuspended in complete RPMI. Cells were counted on the Attune NxT flow cytometer (Thermo Fisher) 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

[0870] ELISPOT analysis was performed according to ELISPOT harmonization guidelines {DOI: 10.1038/nprot.2015.068} with the mouse IFNg ELISpotPLUS kit (MABTECH). 5×10.sup.4 splenocytes were incubated with 10 uM of the indicated peptides for 16 hours in 96-well IFNg antibody coated plates. Spots were developed using alkaline phosphatase. The reaction was timed for 10 minutes and was terminated by running plate under tap water. Spots were counted using an AID vSpot Reader Spectrum. For ELISPOT analysis, wells with saturation >50% were recorded as “too numerous to count”. Samples with deviation of replicate wells >10% were excluded from analysis. Spot counts were then corrected for well confluency using the formula: spot count+2×(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.

XV.B.2. Production of ChAdV68 Viral Delivery Particles after DNA Transfection

[0871] In one example, ChAdV68.4WTnt.GFP (FIG. **21**) and ChAdV68.5WTnt.GFP (FIG. **22**) DNA was transfected into HEK293A cells and virus replication (viral plaques) was observed 7-10 days after transfection. ChAdV68 viral plaques were visualized using light (FIG. **21** A and **22**A) and fluorescent microscopy (FIG. **21** B-C and FIG. **22** B-C). GFP denotes productive ChAdV68 viral delivery particle production.

XV.B.3. ChAdV68 Viral Delivery Particles Expansion

[0872] In one example, ChAdV68.4WTnt.GFP, ChAdV68.5WTnt.GFP, and ChAdV68.5WTnt.MAG25mer viruses were expanded in HEK293F cells and a purified virus stock produced 18 days after transfection (FIG. **23**). Viral particles were quantified in the purified ChAdV68 virus stocks and compared to adenovirus type 5 (Ad5) and ChAdVY25 (a closely related ChAdV; Dicks, 2012, *PloS ONE* 7, e40385) viral stocks produced using the same protocol. ChAdV68 viral titers were comparable to Ad5 and ChAdVY25 (Table 7).

TABLE-US-00015 TABLE 7 Adenoviral vector production in 293F suspension cells Construct Average VP/cell +/− SD Ad5-Vectors (Multiple vectors) 2.96e4 +/− 2.26e4 Ad5-GFP 3.89e4 chAdY25-GFP 1.75e3 +/− 6.03e1 ChAdV68.4WTnt.GFP 1.2e4 +/− 6.5e32 ChAdV68.5WTnt.GFP 1.8e3 ChAdV68.5WTnt.MAG25mer 1.39e3 +/− 1.1e3 *SD is only reported where multiple Production runs have been performed

XV.B.4. Evaluation of Immunogenicity in Tumor Models

[0873] C68 vector expressing mouse tumor antigens were evaluated in mouse immunogenicity studies to demonstrate the C68 vector elicits T-cell responses. T-cell responses to the MHC class I epitope SIINFEKL (SEQ ID NO: 57) were measured in C57BL/6J female mice and the MHC class I epitope AH1-A5 (Slansky et al., 2000, Immunity13:529-538) measured in Balb/c mice. As shown in FIG. **29**, strong T-cell responses relative to control were measured after immunization of mice with ChAdV68.5WTnt.MAG25mer. Mean cellular immune responses of 8957 or 4019 spot forming cells (SFCs) per 10.sup.6 splenocytes were observed in ELISpot assays when C57BL/6J or Balb/c mice were immunized with ChAdV68.5WTnt.MAG25mer, respectively, 10 days after immunization.

XVI. Alphavirus Neoantigen Cassette Delivery Vector

XVI.A. Alphavirus Delivery Vector Evaluation Materials and Methods

In Vitro Transcription to Generate RNA

[0874] For in vitro testing: plasmid DNA was linearized by restriction digest with PmeI, column purified following manufacturer's protocol (GeneJet DNA cleanup kit, Thermo) and used as template. In vitro transcription was performed using the RiboMAX Large Scale RNA production System (Promega) with the m.sup.7G cap analog (Promega) according to manufacturer's protocol. mRNA was purified using the RNeasy kit (Qiagen) according to manufacturer's protocol.

[0875] For in vivo studies: RNA was generated and purified by TriLInk Biotechnologies and capped with Enzymatic Cap1.

Transfection of RNA

[0876] HEK293A cells were seeded at 6e4 cells/well for 96 wells and 2e5 cells/well for 24 wells, −16 hours prior to transfection. Cells were transfected with mRNA using MessengerMAX lipofectamine (Invitrogen) and following manufacturer's protocol. For 96-wells, 0.15 uL of lipofectamine and 10 ng of mRNA was used per well, and for 24-wells, 0.75 uL of lipofectamine and 150 ng of mRNA was used per well. A GFP expressing mRNA (TriLink Biotechnologies) was used as a transfection control.

Luciferase Assay

[0877] Luciferase reporter assay was performed in white-walled 96-well plates with each condition in triplicate using the ONE-Glo luciferase assay (Promega) following manufacturer's protocol. Luminescence was measured using the SpectraMax.

qRT-PCR

[0878] Transfected cells were rinsed and replaced with fresh media 2 hours post transfection to remove any untransfected mRNA. Cells were then harvested at various timepoints in RLT plus lysis buffer (Qiagen), homogenized using a QiaShredder (Qiagen) and RNA was extracted using the RNeasy kit (Qiagen), all according to manufacturer's protocol. Total RNA was quantified using a Nanodrop (Thermo Scientific). qRT-PCR was performed using the Quantitect Probe One-Step RT-PCR kit (Qiagen) on the qTower.sup.3 (Analytik Jena) according to manufacturer's protocol, using 20 ng of total RNA per reaction. Each sample was run in triplicate for each probe. Actin or GusB were used as reference genes. Custom primer/probes were generated by IDT (Table 8).

TABLE-US-00016 TABLE 8 qPCR primers/probes SEQ ID Target NO: Luci Primer1 GTGGTGTGCAGCGAGAATAG 142 Primer2 CGCTCGTTGTAGATGTCGTTAG 143 Probe /56- 144 FAM/TTGCAGTTC/ZEN/TTCATGCCCGTGTTG/3IABkFQ/ GusB Primer1 GTTTTTGATCCAGACCCAGATG 145 Primer2 GCCCATTATTCAGAGCGAGTA 146 Probe /56- 147 FAM/TGCAGGGTT/ZEN/TCACCAGGATCCAC/3IABkFQ/ ActB Primer1 CCTTGCACATGCCGGAG 148 Primer2 ACAGAGCCTCGCCTTTG 149 Probe /56-FAM/TCATCCATG/ZEN/GTGAGCTGGCGG/3IABkFQ/ 150 MAG-25mer Primer1 CTGAAAGCTCGGTTTGCTAATG 151 Set1 Primer2 CCATGCTGGAAGAGACAATCT 152 Probe /56- 153 FAM/CGTTTCTGA/ZEN/TGGCGCTGACCGATA/3IABkFQ/ MAG-25mer Primer1 TATGCCTATCCTGTCTCCTCTG 154 Set2 Primer2 GCTAATGCAGCTAAGTCCTCTC 155 Probe /56- 156 FAM/TGTTTACCC/ZEN/TGACCGTGCCTTCTG/3IABkFQ/

B16-OVA Tumor Model

[0879] C57BL/6J mice were injected in the lower left abdominal flank with 10.sup.5 B 16-OVA cells/animal. Tumors were allowed to grow for 3 days prior to immunization.

CT26 Tumor Model

[0880] Balb/c mice were injected in the lower left abdominal flank with 10.sup.6 CT26 cells/animal. Tumors were allowed to grow for 7 days prior to immunization.

Immunizations

[0881] For srRNA vaccine, mice were injected with 10 μg of RNA in 100 uL volume, bilateral intramuscular injection (50 uL per leg). For Ad5 vaccine, mice were injected with 5×10.sup.10 viral particles (VP) in 100 uL volume, bilateral intramuscular injection (50 uL per leg). Animals were injected with anti-CTLA-4 (clone 9D9, BioXcell), anti-PD-1 (clone RMP1-14, BioXcell) or anti-IgG (clone MPC-1 1, BioXcell), 250 ug dose, 2 times per week, via intraperitoneal injection.

In Vivo Bioluminescent Imaging

[0882] At each timepoint mice were injected with 150 mg/kg luciferin substrate via intraperitoneal injection and bioluminescence was measured using the IVIS In vivo imaging system (PerkinElmer) 10-15 minutes after injection.

Splenocyte Dissociation

[0883] Spleen and lymph nodes 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 (150 mM NH.sub.4Cl, 10 mM KHCO.sub.3, 0.1 mM Na.sub.2EDTA). Cells were filtered again through a 30 micron filter and then resuspended in complete RPMI. Cells were counted on the Attune NxT flow cytometer (Thermo Fisher) 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

[0884] ELISPOT analysis was performed according to ELISPOT harmonization guidelines {DOI: 10.1038/nprot.2015.068} with the mouse IFNg ELISpotPLUS kit (MABTECH). 5×10.sup.4 splenocytes were incubated with 10 uM of the indicated peptides for 16 hours in 96-well IFNg antibody coated plates. Spots were developed using alkaline phosphatase. The reaction was timed for 10 minutes and was terminated by running plate under tap water. Spots were counted using an AID vSpot Reader Spectrum. For ELISPOT analysis, wells with saturation >50% were recorded as “too numerous to count”. Samples with deviation of replicate wells >10% were excluded from analysis. Spot counts were then corrected for well confluency using the formula: spot count+2×(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.

XVI.B. Alphavirus Vector

XVI.B.1. Alphavirus Vector In Vitro Evaluation

[0885] In one implementation of the present invention, a RNA alphavirus backbone for the neoantigen expression system was generated from a Venezuelan Equine Encephalitis (VEE) (Kinney, 1986, Virology 152: 400-413) based self-replicating RNA (srRNA) vector. In one example, the sequences encoding the structural proteins of VEE located 3′ of the 26S sub-genomic promoter were deleted (VEE sequences 7544 to 11,175 deleted; numbering based on Kinney et al 1986; SEQ ID NO:6) and replaced by antigen sequences (SEQ ID NO:14 and SEQ ID NO:4) or a luciferase reporter (e.g., VEE-Luciferase, SEQ ID NO:15) (FIG. **24**). RNA was transcribed from the srRNA DNA vector in vitro, transfected into HEK293A cells and luciferase reporter expression was measured. In addition, an (non-replicating) mRNA encoding luciferase was transfected for comparison. An ~30,000-fold increase in srRNA reporter signal was observed for VEE-Luciferase srRNA when comparing the 23 hour measurement vs the 2 hour measurement (Table 9). In contrast, the mRNA reporter exhibited a less than 10-fold increase in signal over the same time period (Table 9).

TABLE-US-00017 TABLE 9 Expression of luciferase from VEE self-replicating vector increases over time. HEK293A cells transfected with 10 ng of VEE-Luciferase srRNA or 10 ng of non-replicating luciferase mRNA (TriLink L-6307) per well in 96 wells. Luminescence was measured at various times post transfection. Luciferase expression is reported as relative luminescence units (RLU). Each data point is the mean +/− SD of 3 transfected wells. Standard Dev Construct Timepoint (hr) Mean RLU (triplicate wells) mRNA 2 878.6666667 120.7904522 mRNA 5 1847.333333 978.515372 mRNA 9 4847 868.3271273 mRNA 23 8639.333333 751.6816702 SRRNA 2 27 15 SRRNA 5 4884.333333 2955.158935 SRRNA 9 182065.5 16030.81784 SRRNA 23 783658.3333 68985.05538

[0886] In another example, replication of the srRNA was confirmed directly by measuring RNA levels after transfection of either the luciferase encoding srRNA (VEE-Luciferase) or an srRNA encoding a multi-epitope cassette (VEE-MAG25mer) using quantitative reverse transcription polymerase chain reaction (qRT-PCR). An ~150-fold increase in RNA was observed for the VEE-luciferase srRNA (Table 10), while a 30-50-fold increase in RNA was observed for the VEE-MAG25mer srRNA (Table 11). These data confirm that the VEE srRNA vectors replicate when transfected into cells.

TABLE-US-00018 TABLE 10 Direct measurement of RNA replication in VEE-Luciferase srRNA transfected cells. HEK293A cells transfected with VEE-Luciferase srRNA (150 ng per well, 24-well) and RNA levels quantified by qRT-PCR at various times after transfection. Each measurement was normalized based on the Actin reference gene and fold-change relative to the 2 hour timepoint is presented. Timepoint Relative Fold (hr) Luciferase Ct Actin Ct dCt Ref dCt ddCt change 2 20.51 18.14 2.38 2.38 0.00 1.00 4 20.09 18.39 1.70 2.38 −0.67 1.59 6 15.50 18.19 −2.69 2.38 −5.07 33.51 8 13.51 18.36 −4.85 2.38 −7.22 149.43

TABLE-US-00019 TABLE 11 Direct measurement of RNA replication in VEE-MAG25mer srRNA transfected cells. HEK293 cells transfected with VEE-MAG25mer srRNA (150 ng per well, 24-well) and RNA levels quantified by qRT-PCR at various times after transfection. Each measurement was normalized based on the GusB reference gene and fold- change relative to the 2 hour timepoint is presented. Different lines on the graph represent 2 different qPCR primer/probe sets, both of which detect the epitope cassette region of the srRNA. Primer/ Timepoint GusB Relative probe (hr) Ct Ct dCt Ref dCt ddCt Fold-Change Set1 2 18.96 22.41 −3.45 −3.45 0.00 1.00 Set1 4 17.46 22.27 −4.81 −3.45 −1.37 2.58 Set1 6 14.87 22.04 −7.17 −3.45 −3.72 13.21 Set1 8 14.16 22.19 −8.02 −3.45 −4.58 23.86 Set1 24 13.16 22.01 −8.86 −3.45 −5.41 42.52 Set1 36 13.53 22.63 −9.10 −3.45 −5.66 50.45 Set2 2 17.75 22.41 −4.66 −4.66 0.00 1.00 Set2 4 16.66 22.27 −5.61 −4.66 −0.94 1.92 Set2 6 14.22 22.04 −7.82 −4.66 −3.15 8.90 Set2 8 13.18 22.19 −9.01 −4.66 −4.35 20.35 Set2 24 12.22 22.01 −9.80 −4.66 −5.13 35.10 Set2 36 13.08 22.63 −9.55 −4.66 −4.89 29.58

XVI.B.2. Alphavirus Vector In Vivo Evaluation

[0887] In another example, VEE-Luciferase reporter expression was evaluated in vivo. Mice were injected with 10 μg of VEE-Luciferase srRNA encapsulated in lipid nanoparticle (MC3) and imaged at 24 and 48 hours, and 7

and 14 days post injection to determine bioluminescent signal. Luciferase signal was detected at 24 hours post injection and increased over time and appeared to peak at 7 days after srRNA injection (FIG. **25**).

XVI.B.3. Alphavirus Vector Tumor Model Evaluation

[0888] In one implementation, to determine if the VEE srRNA vector directs antigen-specific immune responses in vivo, a VEE srRNA vector was generated (VEE-UbAAY, SEQ ID NO:14) that expresses 2 different MHC class I mouse tumor epitopes, SIINFEKL (SEQ ID NO: 57) and AH1-A5 (Slansky et al., 2000, *Immunity* 13:529-538). The SFL (SIINFEKL (SEQ ID NO: 57)) epitope is expressed by the B16-OVA melanoma cell line, and the AH1-A5 (SPSYAYHQF (SEQ ID NO: 58); Slansky et al., 2000, Immunity) epitope induces T cells targeting a related epitope (AH1/SPSYVYHQF (SEQ ID NO: 193); Huang et al., 1996, *Proc Natl Acad Sci USA* 93:9730-9735) that is expressed by the CT26 colon carcinoma cell line. In one example, for in vivo studies, VEE-UbAAY srRNA was generated by in vitro transcription using T7 polymerase (TriLink Biotechnologies) and encapsulated in a lipid nanoparticle (MC3).

[0889] A strong antigen-specific T-cell response targeting SFL, relative to control, was observed two weeks after immunization of B16-OVA tumor bearing mice with MC3 formulated VEE-UbAAY srRNA. In one example, a median of 3835 spot forming cells (SFC) per 10.sup.6 splenocytes was measured after stimulation with the SFL peptide in ELISpot assays (FIG. **26**A, Table 12) and 1.8% (median) of CD8 T-cells were SFL antigen-specific as measured by pentamer staining (FIG. **26**B, Table 12). In another example, co-administration of an anti-CTLA-4 monoclonal antibody (mAb) with the VEE srRNA vaccine resulted in a moderate increase in overall T-cell responses with a median of 4794.5 SFCs per 10.sup.6 splenocytes measured in the ELISpot assay (FIG. **26**A, Table 12).

TABLE-US-00020 TABLE 12 Results of ELISPOT and MHCI-pentamer staining assays 14 days post VEE srRNA immunization in B16-OVA tumor bearing C57BL/6J mice. Pentamer Pentamer SFC/1e6 positive SFC/1e6 positive Group Mouse splenocytes (% of CD8) Group Mouse splenocytes (% of CD8) Control 1 47 0.22 Vax 1 6774 4.92 2 80 0.32 2 2323 1.34 3 0 0.27 3 2997 1.52 4 0 0.29 4 4492 1.86 5 0 0.27 5 4970 3.7 6 0 0.25 6 4.13 7 0 0.23 7 3835 1.66 8 87 0.25 8 3119 1.64 aCTLA4 1 0 0.24 Vax + 1 6232 2.16 2 0 0.26 aCTLA4 2 4242 0.82 3 0 0.39 3 5347 1.57 4 0 0.28 4 6568 2.33 5 0 0.28 5 6269 1.55 6 0 0.28 6 4056 1.74 7 0 0.31 7 4163 1.14 8 6 0.26 8 3667 1.01 *Note that results from mouse #6 in the Vax group were excluded from analysis due to high variability between triplicate wells.

[0890] In another implementation, to mirror a clinical approach, a heterologous prime/boost in the B 16-OVA and CT26 mouse tumor models was performed, where tumor bearing mice were immunized first with adenoviral vector expressing the same antigen cassette (Ad5-UbAAY), followed by a boost immunization with the VEE-UbAAY srRNA vaccine 14 days after the Ad5-UbAAY prime. In one example, an antigen-specific immune response was induced by the Ad5-UbAAY vaccine resulting in 7330 (median) SFCs per 10.sup.6 splenocytes measured in the ELISpot assay (FIG. **27**A, Table 13) and 2.9% (median) of CD8 T-cells targeting the SFL antigen as measured by pentamer staining (FIG. **27**C, Table 13). In another example, the T-cell response was maintained 2 weeks after the VEE-UbAAY srRNA boost in the B16-OVA model with 3960 (median) SFL-specific SFCs per 10.sup.6 splenocytes measured in the ELISpot assay (FIG. **271**B, Table 13) and 3.1% (median) of CD8 T-cells targeting the SFL antigen as measured by pentamer staining (FIG. **27**D, Table 13).

TABLE-US-00021 TABLE 13 Immune monitoring of B16-OVA mice following heterologous prime/boost with Ad5 vaccine prime and srRNA boost. Pentamer Pentamer SFC/1e6 positive SFC/1e6 positive Group Mouse splenocytes (% of CD8) Group Mouse splenocytes (% of CD8) Day 14 Control 1 0 0.10 Vax 1 8514 1.87 2 0 0.09 2 7779 1.91 3 0 0.11 3 6177 3.17 4 46 0.18 4 7945 3.41 5 0 0.11 5 8821 4.51 6 16 0.11 6 6881 2.48 7 0 0.24 7 5365 2.57 8 37 0.10 8 6705 3.98 aCTLA4 1 0 0.08 Vax + 1 9416 2.35 2 29 0.10 aCTLA4 2 7918 3.33 3 0 0.09 3 10153 4.50 4 29 0.09 4 7212 2.98 5 0 0.10 5 11203 4.38 6 49 0.10 6 9784 2.27 7 0 0.10 8 7267 2.87 8 31 0.14 Day 28 Control 2 0 0.17 Vax 1 5033 2.61 4 0 0.15 2 3958 3.08 6 20 0.17 4 3960 3.58 aCTLA4 1 7 0.23 Vax + 4 3460 2.44 2 0 0.18 aCTLA4 5 5670 3.46 3 0 0.14

[0891] In another implementation, similar results were observed after an Ad5-UbAAY prime and VEE-UbAAY srRNA boost in the CT26 mouse model. In one example, an AH1 antigen-specific response was observed after the Ad5-UbAAY prime (day 14) with a mean of 5187 SFCs per 10.sup.6 splenocytes measured in the ELISpot assay (FIG. **28**A, Table 14) and 3799 SFCs per 10.sup.6 splenocytes measured in the ELISpot assay after the VEE-UbAAY srRNA boost (day 28) (FIG. **28**B, Table 14).

TABLE-US-00022 TABLE 14 Immune monitoring after heterologous prime/ boost in CT26 tumor mouse model. Day 12 Day 21 SFC/1e6 SFC/1e6 Group Mouse splenocytes Group Mouse splenocytes Control 1 1799 Control 9 167 2 1442 10 115 3 1235 11 347 aPD1 1 737 aPD1 8 511 2 5230 11 758 3 332 Vax 9 3133 Vax 1 6287 10 2036 2 4086 11 6227 Vax + 1 5363 Vax + 8 3844 aPD1 2 6500 aPD1 9 2071 11 4888

XVII. ChAdV/srRNA Combination Tumor Model Evaluation

[0892] Various dosing protocols using ChAdV68 and self-replicating RNA (srRNA) were evaluated in murine

CT26 tumor models.

XVII.A ChAdV/srRNA Combination Tumor Model Evaluation

Methods and Materials

Tumor Injection

[0893] Balb/c mice were injected with the CT26 tumor cell line. 7 days after tumor cell injection, mice were randomized to the different study arms (28-40 mice per group) and treatment initiated. Balb/c mice were injected in the lower left abdominal flank with 10.sup.6 CT26 cells/animal. Tumors were allowed to grow for 7 days prior to immunization. The study arms are described in detail in Table 15.

TABLE-US-00023 TABLE 15 ChAdV/srRNA Combination Tumor Model Evaluation Study Arms Group N Treatment Dose Volume Schedule Route 1 40 chAd68 control 1e11 vp 2 × 50 uL day 0 IM srRNA control 10 ug 50 uL day 14, 28, 42 IM Anti-PD1 250 ug 100 uL 2×/week (start day 0) IP 2 40 chAd68 control 1e11 vp 2 × 50 uL day 0 IM srRNA control 10 ug 50 uL day 14, 28, 42 IM Anti-IgG 250 ug 100 uL 2×/week (start day 0) IP 3 28 chAd68 vaccine 1e11 vp 2 × 50 uL day 0 IM srRNA vaccine 10 ug 50 uL day 14, 28, 42 IM Anti-PD1 250 ug 100 uL 2×/week (start day 0) IP 4 28 chAd68 vaccine 1e11 vp 2 × 50 uL day 0 IM srRNA vaccine 10 ug 50 uL day 14, 28, 42 IM Anti-IgG 250 ug 100 uL 2×/week (start day 0) IP 5 28 srRNA vaccine 10 ug 50 uL day 0, 28, 42 IM chAd68 vaccine 1e11 vp 2 × 50 uL day 14 IM Anti-PD1 250 ug 100 uL 2×/week (start day 0) IP 6 28 srRNA vaccine 10 ug 50 uL day 0, 28, 42 IM chAd68 vaccine 1e11 vp 2 × 50 uL day 14 IM Anti-IgG 250 ug 100 uL 2×/week (start day 0) IP 7 40 srRNA vaccine 10 ug 50 uL day 0, 14, 28, 42 IM Anti-PD1 250 ug 100 uL 2×/week (start day 0) IP 8 40 srRNA vaccine 10 ug 50 uL day 0, 14, 28, 42 IM Anti-IgG 250 ug 100 uL 2×/week (start day 0) IP

Immunizations

[0894] For srRNA vaccine, mice were injected with 10 μg of VEE-MAG25mer srRNA in 100 uL volume, bilateral intramuscular injection (50 uL per leg). For C68 vaccine, mice were injected with 1×10.sup.11 viral particles (VP) of ChAdV68.5WTnt.MAG25mer in 100 uL volume, bilateral intramuscular injection (50 uL per leg). Animals were injected with anti-PD-1 (clone RMP1-14, BioXcell) or anti-IgG (clone MPC-11, BioXcell), 250 ug dose, 2 times per week, via intraperitoneal injection.

Splenocyte Dissociation

[0895] Spleen and lymph nodes 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 (150 mM NH.sub.4Cl, 10 mM KHCO.sub.3, 0.1 mM Na.sub.2EDTA). Cells were filtered again through a 30 micron filter and then resuspended in complete RPML. Cells were counted on the Attune NxT flow cytometer (Thermo Fisher) 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

[0896] ELISPOT analysis was performed according to ELISPOT harmonization guidelines {DOI: 10.1038/nprot.2015.068} with the mouse IFNg ELISpotPLUS kit (MABTECH). 5×10.sup.4 splenocytes were incubated with 10 uM of the indicated peptides for 16 hours in 96-well IFNg antibody coated plates. Spots were developed using alkaline phosphatase. The reaction was timed for 10 minutes and was terminated by running plate under tap water. Spots were counted using an AID vSpot Reader Spectrum. For ELISPOT analysis, wells with saturation >50% were recorded as “too numerous to count”. Samples with deviation of replicate wells >10% were excluded from analysis. Spot counts were then corrected for well confluency using the formula: spot count+2×(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.

XVII.B ChAdV/srRNA Combination Evaluation in a CT26 Tumor Model

[0897] The immunogenicity and efficacy of the ChAdV68.5WTnt.MAG25mer/VEE-MAG25mer srRNA heterologous prime/boost or VEE-MAG25mer srRNA homologous prime/boost vaccines were evaluated in the CT26 mouse tumor model. Balb/c mice were injected with the CT26 tumor cell line. 7 days after tumor cell injection, mice were randomized to the different study arms and treatment initiated. The study arms are described in detail in Table 15 and more generally in Table 16.

TABLE-US-00024 TABLE 16 Prime/Boost Study Arms Group Prime Boost 1 Control Control 2 Control + anti-PD-1 Control + anti-PD-1 3 ChAdV68.5WTnt.MAG25mer VEE-MAG25mer srRNA 4 ChAdV68.5WTnt.MAG25mer + anti-PD-1 VEE-MAG25mer srRNA + anti-PD-1 5 VEE-MAG25mer srRNA ChAdV68.5WTnt.MAG25mer 6 VEE-MAG25mer srRNA + anti-PD-1 ChAdV68.5WTnt.MAG25mer + anti-PD-1 7 VEE-MAG25mer srRNA VEE-MAG25mer srRNA 8 VEE-MAG25mer srRNA + anti-PD-1 VEE-MAG25mer

srRNA + anti-PD-1
[0898] Spleens were harvested 14 days after the prime vaccination for immune monitoring. Tumor and body weight measurements were taken twice a week and survival was monitored. Strong immune responses relative to control were observed in all active vaccine groups.
[0899] Median cellular immune responses of 10,630, 12,976, 3319, or 3745 spot forming cells (SFCs) per 10.sup.6 splenocytes were observed in ELISpot assays in mice immunized with ChAdV68.5WTnt.MAG25mer (ChAdV/group 3), ChAdV68.5WTnt.MAG25mer+ant-PD-1 (ChAdV+PD-1/group 4), VEE-MAG25mer srRNA (srRNA/median for groups 5 & 7 combined), or VEE-MAG25mer srRNA+ant-PD-1 (srRNA+PD-1/median for groups 6 & 8 combined), respectively, 14 days after the first immunization (FIG. **30** and Table 17). In contrast, the vaccine control (group 1) or vaccine control with anti-PD-1 (group 2) exhibited median cellular immune responses of 296 or 285 SFC per 10.sup.6 splenocytes, respectively.
TABLE-US-00025 TABLE 17 Cellular immune responses in a CT26 tumor model Treatment Median SFC/10.sup.6 Splenocytes Control 296 PD1 285 ChAdV68.5WTnt.MAG25mer 10630 (ChAdV) ChAdV68.5WTnt.MAG25mer + 12976 PD1 (ChAdV + PD-1) VEE-MAG25mer srRNA 3319 (srRNA) VEE-MAG25mer srRNA + 3745 PD-1 (srRNA + PD1)
[0900] Consistent with the ELISpot data, 5.6, 7.8, 1.8 or 1.9% of CD8 T cells (median) exhibited antigen-specific responses in intracellular cytokine staining (ICS) analyses for mice immunized with ChAdV68.5WTnt.MAG25mer (ChAdV/group 3), ChAdV68.5WTnt.MAG25mer+anti-PD-1 (ChAdV+PD-1/group 4), VEE-MAG25mer srRNA (srRNA/median for groups 5 & 7 combined), or VEE-MAG25mer srRNA+anti-PD-1 (srRNA+PD-1/median for groups 6 & 8 combined), respectively, 14 days after the first immunization (FIG. **31** and Table 18). Mice immunized with the vaccine control or vaccine control combined with anti-PD-1 showed antigen-specific CD8 responses of 0.2 and 0.1%, respectively.
TABLE-US-00026 TABLE 18 CD8 T-Cell responses in a CT26 tumor model Median % CD8 IFN- Treatment gamma Positive Control 0.21 PD1 0.1 ChAdV68.5WTnt.MAG25mer 5.6 (ChAdV) ChAdV68.5WTnt.MAG25mer + 7.8 PD1 (ChAdV + PD-1) VEE-MAG25mer srRNA 1.8 (srRNA) VEE-MAG25mer srRNA + 1.9 PD-1 (srRNA + PD1)
[0901] Tumor growth was measured in the CT26 colon tumor model for all groups, and tumor growth up to 21 days after treatment initiation (28 days after injection of CT-26 tumor cells) is presented. Mice were sacrificed 21 days after treatment initiation based on large tumor sizes (>2500 mm.sup.3); therefore, only the first 21 days are presented to avoid analytical bias. Mean tumor volumes at 21 days were 1129, 848, 2142, 1418, 2198 and 1606 mm.sup.3 for ChAdV68.5WTnt.MAG25mer prime/VEE-MAG25mer srRNA boost (group 3), ChAdV68.5WTnt.MAG25mer prime/VEE-MAG25mer srRNA boost+anti-PD-1 (group 4), VEE-MAG25mer srRNA prime/ChAdV68.5WTnt.MAG25mer boost (group 5), VEE-MAG25mer srRNA prime/ChAdV68.5WTnt.MAG25mer boost+anti-PD-1 (group 6), VEE-MAG25mer srRNA prime/VEE-MAG25mer srRNA boost (group 7) and VEE-MAG25mer srRNA prime/VEE-MAG25mer srRNA boost+anti-PD-1 (group 8), respectively (FIG. **32** and Table 19). The mean tumor volumes in the vaccine control or vaccine control combined with anti-PD-1 were 2361 or 2067 mm.sup.3, respectively. Based on these data, vaccine treatment with ChAdV68.5WTnt.MAG25mer/VEE-MAG25mer srRNA (group 3), ChAdV68.5WTnt.MAG25mer/VEE-MAG25mer srRNA+anti-PD-1 (group 4), VEE-MAG25mer srRNA/ChAdV68.5WTnt.MAG25mer+anti-PD-1 (group 6) and VEE-MAG25mer srRNA/VEE-MAG25mer srRNA+anti-PD-1 (group 8) resulted in a reduction of tumor growth at 21 days that was significantly different from the control (group 1).
TABLE-US-00027 TABLE 19 Tumor size at day 21 measured in the CT26 model Treatment Tumor Size (mm.sup.3) SEM Control 2361 235 PD1 2067 137 chAdV/srRNA 1129 181 chAdV/srRNA + 848 182 PD1 srRNA/chAdV 2142 233 srRNA/chAdV + 1418 220 PD1 srRNA 2198 134 srRNA + PD1 1606 210
[0902] Survival was monitored for 35 days after treatment initiation in the CT-26 tumor model (42 days after injection of CT-26 tumor cells). Improved survival was observed after vaccination of mice with 4 of the combinations tested. After vaccination, 64%, 46%, 41% and 36% of mice survived with ChAdV68.5WTnt.MAG25mer prime/VEE-MAG25mer srRNA boost in combination with anti-PD-1 (group 4; P<0.0001 relative to control group 1), VEE-MAG25mer srRNA prime/VEE-MAG25mer srRNA boost in combination with anti-PD-1 (group 8; P=0.0006 relative to control group 1), ChAdV68.5WTnt.MAG25mer prime/VEE-MAG25mer srRNA boost (group 3; P=0.0003 relative to control group 1) and VEE-MAG25mer srRNA prime/ChAdV68.5WTnt.MAG25mer boost in combination with anti-PD-1 (group 6; P=0.0016 relative to control group 1), respectively (FIG. **33** and Table 20). Survival was not significantly different from the control group 1 (<14%) for the remaining treatment groups [VEE-MAG25mer srRNAprime/ChAdV68.5WTnt.MAG25mer boost (group 5), VEE-MAG25mer srRNA prime/VEE-MAG25mer srRNA boost (group 7) and anti-PD-1 alone (group 2)].

TABLE-US-00028 TABLE 20 Survival in the CT26 model chAdV/ srRNA/ chAdV/ srRNA + srRNA/ chAdV + srRNA + Timepoint Control PD1 srRNA PD1 chAdV PD1 srRNA PD1 0 100 100 100 100.00 100.00 100 100 100 21 96 100 100 100 100 95 100 100 24 54 64 91 100 68 82 68 71 28 21 32 68 86 45 68 21 64 31 7 14 41 64 14 36 11 46 35 7 14 41 64 14 36 11 46

[0903] In conclusion, ChAdV68.5WTnt.MAG25mer and VEE-MAG25mer srRNA elicited strong T-cell responses to mouse tumor antigens encoded by the vaccines, relative to control. Administration of a ChAdV68.5WTnt.MAG25mer prime and VEE-MAG25mer srRNA boost with or without co-administration of anti-PD-1, VEE-MAG25mer srRNA prime and ChAdV68.5WTnt.MAG25mer boost in combination with anti-PD-1 or administration of VEE-MAG25mer srRNA as a homologous prime boost immunization in combination with anti-PD-1 to tumor bearing mice resulted in improved survival.

XVIII. Non-Human Primate Studies

[0904] Various dosing protocols using ChAdV68 and self-replicating RNA (srRNA) were evaluated in non-human primates (NHP).

Materials and Methods

[0905] A priming vaccine was injected intramuscularly (IM) in each NHP to initiate the study (vaccine prime). One or more boosting vaccines (vaccine boost) were also injected intramuscularly in each NHP. Bilateral injections per dose were administered according to groups outlined in tables and summarized below.

Immunizations

[0906] Mamu-A*01 Indian rhesus macaques were immunized bilaterally with 1×10.sup.12 viral particles (5×10.sup.11 viral particles per injection) of ChAdV68.5WTnt.MAG25mer, 30 μg of VEE-MAG25 MER srRNA, 100 μg of VEE-MAG25mer srRNA or 300 μg of VEE-MAG25mer srRNA formulated in LNP-1 or LNP-2. Vaccine boosts of 30 ug, 100 μg or 300 ug VEE-MAG25mer srRNA were administered intramuscularly at the indicated time after prime vaccination.

Immune Monitoring

[0907] PBMCs were isolated at indicated times after prime vaccination using Lymphocyte Separation Medium (LSM, MP Biomedicals) and LeucoSep separation tubes (Greiner Bio-One) and resuspended in RPMI containing 10% FBS and penicillin/streptomycin. Cells were counted on the Attune NxT flow cytometer (Thermo Fisher) using propidium iodide staining to exclude dead and apoptotic cells. Cell were then adjusted to the appropriate concentration of live cells for subsequent analysis. For each monkey in the studies, T cell responses were measured using ELISpot or flow cytometry methods. T cell responses to 6 different rhesus macaque Mamu-A*01 class I epitopes encoded in the vaccines were monitored from PBMCs by measuring induction of cytokines, such as IFN-gamma, using ex vivo enzyme-linked immunospot (ELISpot) analysis. ELISpot analysis was performed according to ELISPOT harmonization guidelines {DOI: 10.1038/nprot.2015.068} with the monkey IFNg ELISpotPLUS kit (MABTECH). 200,000 PBMCs were incubated with 10 uM of the indicated peptides for 16 hours in 96-well IFNg antibody coated plates. Spots were developed using alkaline phosphatase. The reaction was timed for 10 minutes and was terminated by running plate under tap water. Spots were counted using an AID vSpot Reader Spectrum. For ELISPOT analysis, wells with saturation >50% were recorded as “too numerous to count”. Samples with deviation of replicate wells >10% were excluded from analysis. Spot counts were then corrected for well confluency using the formula: spot count+2×(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.

[0908] Specific CD4 and CD8 T cell responses to 6 different rhesus macaque Mamu-A*01 class I epitopes encoded in the vaccines were monitored from PBMCs by measuring induction of intracellular cytokines, such as IFN-gamma, using flow cytometry. The results from both methods indicate that cytokines were induced in an antigen-specific manner to epitopes.

Immunogenicity in Rhesus Macaques

[0909] This study was designed to (a) evaluate the immunogenicity and preliminary safety of VEE-MAG25mer srRNA 30 μg and 100 μg doses as a homologous prime/boost or heterologous prime/boost in combination with ChAdV68.5WTnt.MAG25mer; (b) compare the immune responses of VEE-MAG25mer srRNA in lipid nanoparticles using LNP1 versus LNP2; (c) evaluate the kinetics of T-cell responses to VEE-MAG25mer srRNA and ChAdV68.5WTnt.MAG25mer immunizations.

[0910] The study arm was conducted in Mamu-A*01 Indian rhesus macaques to demonstrate immunogenicity. Select antigens used in this study are only recognized in Rhesus macaques, specifically those with a Mamu-A*01 MHC class I haplotype. Mamu-A*01 Indian rhesus macaques were randomized to the different study arms (6 macaques per group) and administered an IM injection bilaterally with either ChAdV68.5WTnt.MAG25mer or VEE-MAG25mer srRNA vector encoding model antigens that includes multiple Mamu-A*01 restricted epitopes.

The study arms were as described below.

TABLE-US-00029 TABLE 21 Non-GLP immunogenicity study in Indian Rhesus Macaques Group Prime Boost 1 Boost 2 1 VEE- VEE- VEE- MAG25mer MAG25mer MAG25mer srRNA- srRNA- srRNA- LNP1(30 μg) LNP1 (30 μg) LNP1 (30 μg) 2 VEE- VEE- VEE- MAG25mer MAG25mer MAG25mer srRNA- srRNA- srRNA- LNP1 (100 μg) LNP1 (100 μg) LNP1 (100 μg) 3 VEE- VEE- VEE- MAG25mer MAG25mer MAG25mer srRNA- srRNA- srRNA- LNP2 (100 μg) LNP2 (100 μg) LNP2 (100 μg) 4 ChAdV68.5WTnt. VEE- VEE- MAG25mer MAG25mer MAG25mer srRNA- srRNA- LNP1 (100 μg) LNP1 (100 μg)

[0911] PBMCs were collected prior to immunization and on weeks 1, 2, 3, 4, 5, 6, 8, 9, and 10 after the initial immunization for immune monitoring.

Results

[0912] Antigen-specific cellular immune responses in peripheral blood mononuclear cells (PBMCs) were measured to six different Mamu-A*01 restricted epitopes prior to immunization and 1, 2, 3, 4, 5, 6, 8, 9, and 10 weeks after the initial immunization. Animals received a boost immunization with VEE-MAG25mer srRNA on weeks 4 and 8 with either 30 μg or 100 μg doses, and either formulated with LNP1 or LNP2, as described in Table 21. Combined immune responses to all six epitopes were plotted for each immune monitoring timepoint (FIG. **34**A-D and Tables 22-25).

[0913] Combined antigen-specific immune responses were observed at all measurements with 170, 14, 15, 11, 7, 8, 14, 17, 12 SFCs per 10.sup.6 PBMCs (six epitopes combined) 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after an initial VEE-MAG25mer srRNA-LNP1(30 μg) prime immunization, respectively (FIG. **34**A). Combined antigen-specific immune responses were observed at all measurements with 108, -3, 14, 1, 37, 4, 105, 17, 25 SFCs per 10.sup.6 PBMCs (six epitopes combined) 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after an initial VEE-MAG25mer srRNA-LNP1(100 μg) prime immunization, respectively (FIG. **34**B). Combined antigen-specific immune responses were observed at all measurements with -17, 38, 14, -2, 87, 21, 104, 129, 89 SFCs per 10.sup.6 PBMCs (six epitopes combined) 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after an initial VEE-MAG25mer srRNA-LNP2(100 μg) prime immunization, respectively (FIG. **34**C). Negative values are a result of normalization to pre-bleed values for each epitope/animal.

[0914] Combined antigen-specific immune responses were observed at all measurements with 1218, 1784, 1866, 973, 1813, 747, 797, 1249, and 547 SFCs per 10.sup.6 PBMCs (six epitopes combined) 1, 2, 3, 4, 5, 6, 8, 9, or 10 weeks after an initial ChAdV68.5WTnt.MAG25mer prime immunization, respectively (FIG. **34**D). The immune response showed the expected profile with peak immune responses measured -2-3 weeks after the prime immunization followed by a contraction in the immune response after 4 weeks. Combined antigen-specific cellular immune responses of 1813 SFCs per 10.sup.6 PBMCs (six epitopes combined) were measured 5 weeks after the initial immunization with ChAdV68.5WTnt.MAG25mer (i.e., 1 week after the first boost with VEE-MAG25mer srRNA). The immune response measured 1 week after the first boost with VEE-MAG25mer srRNA (week 5) was comparable to the peak immune response measured for the ChAdV68.5WTnt.MAG25mer prime immunization (week 3) (FIG. **34**D). Combined antigen-specific cellular immune responses of 1249 SFCs per 10.sup.6 PBMCs (six epitopes combined) was measured 9 weeks after the initial immunization with ChAdV68.5WTnt.MAG25mer, respectively (i.e., 1 week after the second boost with VEE-MAG25mer srRNA). The immune responses measured 1 week after the second boost with VEE-MAG25mer srRNA (week 9) was -2-fold higher than that measured just before the boost immunization (FIG. **34**D).

TABLE-US-00030 TABLE 22 Mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope ± SEM for VEE-MAG25mer srRNA-LNP1(30 μg) (Group 1) Antigen Wk Env CL9 Env TL9 Gag CM9 Gag LW9 Pol SV9 Tat TL8 1 0 ± 0 0 ± 0 0 ± 0 0 ± 0 0 ± 0 0 ± 0 2 39.7 ± 22.7 35.4 ± 25.1 3.2 ± 3.6 .sup. 33 ± 28.1 30.9 ± 20.3 28.3 ± 17.5 3 .sup. 2 ± 2.4 0.2 ± 1.8 1.8 ± 2.4 3.7 ± 1.9 1.7 ± 2.8 4.9 ± 2.3 4 .sup. 1 ± 1.8 0.3 ± 1.2 5.5 ± 3.6 2.3 ± 2.2 5.7 ± 2.7 0.8 ± 0.8 5 0.5 ± 0.9 1.4 ± 3.8 3.1 ± 1.6 2.3 ± 2.7 1.9 ± 2.sup. 1.4 ± 1.2 6 1.9 ± 1.8 −0.3 ± 3 .sup. 1.7 ± 1.2 1.4 ± 1.4 0.8 ± 1.1 1.1 ± 1.sup. 8 −0.4 ± 0.8 −0.9 ± 2.9 0.5 ± 1.3 .sup. 3 ± 1.1 2.2 ± 2.1 3.7 ± 2.sup. 9 .sup. 1 ± 1.7 1.2 ± 4.2 7.2 ± 3.9 0.5 ± 0.7 1.6 ± 3.sup. 3 ± 1 10 3.8 ± 1.8 11 ± 5 −1.1 ± 1.1 1.9 ± 0.9 1.3 ± 1.6 0.2 ± 0.5

TABLE-US-00031 TABLE 23 Mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope ± SEM for VEE-MAG25mer srRNA-LNP1(100 μg) (Group 2) Antigen Wk Env CL9 Env TL9 Gag CM9 Gag LW9 Pol SV9 Tat TL8 1 0 ± 0 0 ± 0 0 ± 0 0 ± 0 0 ± 0 0 ± 0 2 7.9 ± 17.2 23.2 ± 17.4 11.4 ± 4.9 41.7 ± 16.5 .sup. 15 ± 13.5 8.9 ± 6.2 3 −3.1 ± 4.6 −7.2 ± 6.5 2.3 ± 2.3 −0.3 ± 2.7 2.7 ± 5.1 2.2 ± 1.4 4 1.9 ± 3.8 −6.2 ± 7.6 10.5 ± 4.1 1.2 ± 2.9 5.6 ± 4.9 1.1 ± 0.8 5 −2.6 ± 7 .sup. −8 ± 5.9 1.5 ± 1.7 6.4 ± 2.3 0.7 ± 4.3 3.3 ± 1.3 6 6.3 ± 6.3 4.4 ± 8.3 6.6 ± 4.4 5.2 ± 5.2 3.9 ± 5.sup. 10.8 ± 6.9 8 −3.6 ± 7.2 −6.8 ± 7.3 −0.8 ± 1.2 3.4 ± 4.2 6.4 ± 7.5 5.7 ± 2.7 9 8.1 ± 2.4 20.6 ± 23.4 18.9 ± 5.7 8.1 ± 8.9 9 ± 11.2 .sup. 40 ± 17.6 10 3.1 ± 8.sup. −3.9 ± 8.5 3.3 ± 1.8 0.6 ± 2.9 7.4 ± 6.4 6.1 ± 2.5

TABLE-US-00032 TABLE 24 Mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope ± SEM for

VEE-MAG25mer srRNA-LNP2(100 μg) (Group 3) Antigen Wk Env CL9 Env TL9 Gag CM9 Gag LW9 Pol SV9 Tat TL8 1 0 ± 0 0 ± 0 0 ± 0 0 ± 0 0 ± 0 0 ± 0 2 −5.9 ± 3.8 −0.3 ± 0.5 −0.5 ± 1.5 −5.7 ± 6.1 −1 ± 1.3 −3.2 ± 5.5 3 0.7 ± 5.2 3.4 ± 2.4 4.2 ± 4.6 18.3 ± 15.5 11.9 ± 5.1 −0.4 ± 8.2 4 −3.8 ± 5.5 2.3 ± 1.8 11.3 ± 6.1 −3.1 ± 5.6 8.5 ± 4.sup. −1.5 ± 6.1 5 −3.7 ± 5.7 −0.1 ± 0.7 −0.2 ± 1.6 3.4 ± 8.5 .sup. 3 ± 3.1 −4.6 ± 5 .sup. 6 12.3 ± 15.sup. 7.8 ± 4.9 24.7 ± 19.8 23.2 ± 22.5 18.7 ± 15.8 0.5 ± 6.2 8 5.9 ± 12.3 −0.1 ± 0.7 −0.5 ± 1.3 8.8 ± 14.4 8.7 ± 8.sup. −1.3 ± 4 .sup. 9 16.1 ± 13.4 16.5 ± 4 22.9 ± 4.2 .sup. 13 ± 13.2 16.4 ± 7.8 19.6 ± 9.2 10 29.9 ± 21.8 .sup. 22 ± 19.5 0.5 ± 2.6 22.2 ± 22.6 35.3 ± 15.8 19.4 ± 17.3

TABLE-US-00033 TABLE 25 Mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope ± SEM for ChAdV68.5WTnt.MAG25mer prime Antigen Wk Env CL9 Env TL9 Gag CM9 Gag LW9 Pol SV9 Tat TL8 1 178 ± 68.7 206.5 ± 94.8 221.2 ± 120.sup. 15.4 ± 16.7 33.3 ± 25.9 563.5 ± 174.4 2 311.2 ± 165.5 278.8 ± 100.9 344.6 ± 110.8 46.3 ± 13.5 181.6 ± 76.8 621.4 ± 220.9 3 277.3 ± 101.1 359.6 ± 90.5 468.2 ± 106.6 41.7 ± 11.1 169.8 ± 57.8 549.4 ± 115.7 4 140 ± 46.5 169.6 ± 46.8 239.4 ± 37 26.5 ± 11.4 75 ± 31.6 322.2 ± 50.7 5 155.6 ± 62.1 406.7 ± 96.4 542.7 ± 143.3 35.1 ± 16.6 134.2 ± 53.7 538.5 ± 91.9 6 78.9 ± 42.5 95.5 ± 29.4 220.9 ± 75.3 −1.4 ± 5.3 43.4 ± 19.6 308.1 ± 42.6 8 88.4 ± 30.4 162.1 ± 30.3 253.4 ± 78.6 21.4 ± 11.2 53.7 ± 22.3 217.8 ± 45.2 9 158.5 ± 69 322.3 ± 87.2 338.2 ± 137.1 5.6 ± 12.4 109.2 ± 17.9 314.8 ± 43.4 10 97.3 ± 32.5 133.2 ± 27.sup. 154.9 ± 59.2 10 ± 6 26 ± 16.7 125.5 ± 27.7

Non-GLP RNA Dose Ranging Study (Higher Doses) in Indian Rhesus Macaques

[0915] This study was designed to (a) evaluate the immunogenicity of VEE-MAG25mer srRNAat a dose of 300 μg as a homologous prime/boost or heterologous prime/boost in combination with ChAdV68.5WTnt.MAG25mer; (b) compare the immune responses of VEE-MAG25mer srRNA in lipid nanoparticles using LNP1 versus LNP2 at the 300 μg dose; and (c) evaluate the kinetics of T-cell responses to VEE-MAG25mer srRNA and ChAdV68.5WTnt.MAG25mer immunizations.

[0916] The study arm was conducted in Mamu-A*01 Indian rhesus macaques to demonstrate immunogenicity. Vaccine immunogenicity in nonhuman primate species, such as Rhesus, is the best predictor of vaccine potency in humans. Furthermore, select antigens used in this study are only recognized in Rhesus macaques, specifically those with a Mamu-A*01 MHC class I haplotype. Mamu-A*01 Indian rhesus macaques were randomized to the different study arms (6 macaques per group) and administered an IM injection bilaterally with either ChAdV68.5-WTnt.MAG25mer or VEE-MAG25mer srRNA encoding model antigens that includes multiple Mamu-A*01 restricted antigens. The study arms were as described below.

[0917] PBMCs were collected prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after the initial immunization for immune monitoring for group 1 (heterologous prime/boost). PBMCs were collected prior to immunization and 4, 5, 7, 8, 10, 11, 12, 13, 14, or 15 weeks after the initial immunization for immune monitoring for groups 2 and 3 (homologous prime/boost).

TABLE-US-00034 TABLE 26 Non-GLP immunogenicity study in Indian Rhesus Macaques Group Prime Boost 1 Boost 2 Boost 3 1 ChAdV68.5WTnt. VEE-MAG25mer VEE-MAG25mer VEE-MAG25mer MAG25mer srRNA- srRNA- srRNA- LNP2 (300 μg) LNP2 (300 μg) LNP2 (300 μg) 2 VEE-MAG25mer VEE-MAG25mer VEE-MAG25mer srRNA- srRNA- srRNA- LNP2 (300 μg) LNP2 (300 μg) LNP2 (300 μg) 3 VEE-MAG25mer VEE-MAG25mer VEE-MAG25mer srRNA- srRNA- srRNA- LNP1 (300 μg) LNP1 (300 μg) LNP1 (300 μg)

Results

[0918] Mamu-A*01 Indian rhesus macaques were immunized with ChAdV68.5-WTnt.MAG25mer. Antigen-specific cellular immune responses in peripheral blood mononuclear cells (PBMCs) were measured to six different Mamu-A*01 restricted epitopes prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after the initial immunization (FIG. **35** and Table 27). Animals received boost immunizations with VEE-MAG25mer srRNA using the LNP2 formulation on weeks 4, 12, and 20. Combined antigen-specific immune responses of 1750, 4225, 1100, 2529, 3218, 1915, 1708, 1561, 5077, 4543, 4920, 5820, 3395, 2728, 1996, 1465, 4730, 2984, 2828, or 3043 SFCs per 10.sup.6 PBMCs (six epitopes combined) were measured 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24 weeks after the initial immunization with ChAdV68.5WTnt.MAG25mer (FIG. **35**). Immune responses measured 1 week after the second boost immunization (week 13) with VEE-MAG25mer srRNA were -3-fold higher than that measured just before the boost immunization (week 12). Immune responses measured 1 week after the third boost immunization (week 21) with VEE-MAG25mer srRNA, were -3-fold higher than that measured just before the boost immunization (week 20), similar to the response observed for the second boost.

[0919] Mamu-A*01 Indian rhesus macaques were also immunized with VEE-MAG25mer srRNA using two different LNP formulations (LNP1 and LNP2). Antigen-specific cellular immune responses in peripheral blood mononuclear cells (PBMCs) were measured to six different Mamu-A*01 restricted epitopes prior to immunization and 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, or 15 weeks after the initial immunization (FIGS. **36** and **37**, Tables 28 and 29). Animals received boost immunizations with VEE-MAG25mer srRNA using the respective

LNP1 or LNP2 formulation on weeks 4 and 12. Combined antigen-specific immune responses of 168, 204, 103, 126, 140, 145, 330, 203, and 162 SFCs per 106 PBMCs (six epitopes combined) were measured 4, 5, 7, 8, 10, 11, 13, 14, 15 weeks after the immunization with VEE-MAG25mer srRNA-LNP2 (FIG. **36**). Combined antigen-specific immune responses of 189, 185, 349, 437, 492, 570, 233, 886, 369, and 381 SFCs per 10.sup.6 PBMCs (six epitopes combined) were measured 4, 5, 7, 8, 10, 11, 12, 13, 14, 15 weeks after the immunization with VEE-MAG25mer srRNA-LNP1 (FIG. **37**).

TABLE-US-00035 TABLE 27 Mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope ± SEM for priming vaccination with ChAdV68.5WTnt.MAG25mer (Group 1) Antigen Wk Env CL9 Env TL9 Gag CM9 Gag LW9 Pol SV9 Tat TL8 4 173 ± 41.6 373.5 ± 87.3 461.4 ± 74.2 38.4 ± 26.1 94.5 ± 26 609.2 ± 121.9 5 412.7 ± 138.4 987.8 ± 283.3 1064.4 ± 266.9 85.6 ± 31.2 367.2 ± 135.2 1306.8 ± 332.8 6 116.2 ± 41.2 231.1 ± 46.3 268.3 ± 90.7 86.1 ± 42 174.3 ± 61.sup. 223.9 ± 38.1 7 287.4 ± 148.7 588.9 ± 173.9 693.2 ± 224.8 92.1 ± 33.5 172.9 ± 55.6 694.6 ± 194.8 8 325.4 ± 126.6 735.8 ± 212 948.9 ± 274.5 211.3 ± 62.7 179.1 ± 50.sup. 817.3 ± 185.2 10 .sup. 312 ± 129.7 543.2 ± 188.4 618.6 ± 221.7 −5.7 ± 4.1 136.5 ± 51.3 309.9 ± 85.6 11 248.5 ± 81.1 348.7 ± 129.8 581.1 ± 205.5 −3.1 ± 4.4 .sup. 119 ± 51.2 413.7 ± 144.8 12 261.9 ± 68.2 329.9 ± 83.sup. 486.5 ± 118.6 −1.2 ± 5.1 132.8 ± 31.8 350.9 ± 69.3 13 389.3 ± 167.7 1615.8 ± 418.3 1244.3 ± 403.6 1.3 ± 8.1 522.5 ± 155 1303.3 ± 385.6 14 406.3 ± 121.6 .sup. 1616 ± 491.7 1142.3 ± 247.2 6.6 ± 11.1 322.7 ± 94.1 1048.6 ± 215.6 15 446.8 ± 138.7 1700.8 ± 469.1 1306.3 ± 294.4 .sup. 43 ± 24.5 421.2 ± 87.9 1001.5 ± 236.4 16 686.8 ± 268.8 1979.5 ± 541.7 1616.8 ± 411.8 2.4 ± 7.8 381.9 ± 116.4 1152.8 ± 352.7 17 375.8 ± 109.3 1378.6 ± 561.2 773.1 ± 210.3 −1.4 ± 4.3 177.6 ± 93.7 691.7 ± 245 18 255.9 ± 99.7 1538.4 ± 498.1 498.7 ± 152.3 −5.3 ± 3.3 26.2 ± 13.4 413.9 ± 164.8 19 133 ± 62.6 955.9 ± 456.8 491.1 ± 121.8 −5.7 ± 4.1 50.3 ± 25.4 371.2 ± 123.7 20 163.7 ± 55.8 641.7 ± 313.5 357.9 ± 91.1 2.6 ± 7.5 41.4 ± 24.2 257.8 ± 68.9 21 319.9 ± 160.5 2017.1 ± 419.9 1204.8 ± 335.2 −3.7 ± 5.1 268.1 ± 109.6 924.1 ± 301 22 244.7 ± 105.6 1370.9 ± 563.5 780.3 ± 390 −3.6 ± 5.1 118.2 ± 68.1 473.3 ± 249.3 23 176.7 ± 81.8 1263.7 ± 527.3 838.6 ± 367.9 −5.7 ± 4.1 73.6 ± 49 480.9 ± 163.9 24 236.5 ± 92 1324.7 ± 589.3 879.7 ± 321 −0.4 ± 5.7 .sup. 104 ± 53.1 498 ± 135.8

TABLE-US-00036 TABLE 28 Mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope ± SEM for priming vaccination with VEE-MAG25mer srRNA-LNP2 (300 μg) (Group 2) Antigen Wk Env CL9 Env TL9 Gag CM9 Gag LW9 Pol SV9 Tat TL8 4 .sup. 46 ± 27.1 18.4 ± 6.8 58.3 ± 45.8 29.9 ± 20.8 4.9 ± 2.3 10.7 ± 4 5 85.4 ± 54.sup. 5.2 ± 5.8 52.4 ± 51.2 34.5 ± 35.sup. 11.8 ± 12.2 14.4 ± 7.9 7 18.6 ± 32.5 1.9 ± 1.7 59.4 ± 55.7 9.3 ± 10.7 3.3 ± 3.sup. 10.7 ± 6.1 8 36.6 ± 39.4 6.3 ± 3.9 48.7 ± 39.9 13.5 ± 8.8 3.8 ± 3.6 17.2 ± 9.7 10 69.1 ± 59.1 4.4 ± 1.9 39.3 ± 38.sup. 14.7 ± 10.8 4.4 ± 5.3 8.5 ± 5.3 11 .sup. 43 ± 38.8 22.6 ± 21.1 30.2 ± 26.2 3.3 ± 2.2 5.8 ± 3.5 40.3 ± 25.5 13 120.4 ± 78.3 68.2 ± 43.9 54.2 ± 36.8 21.8 ± 7.4 17.7 ± 6.1 47.4 ± 27.3 14 .sup. 76 ± 44.8 .sup. 28 ± 19.5 65.9 ± 64.3 −0.3 ± 1.3 2.5 ± 2.sup. 31.1 ± 26.5 15 58.9 ± 41.4 19.5 ± 15.1 55.4 ± 51.sup. 2.5 ± 2.sup. 5.5 ± 3.6 20.1 ± 15.7

TABLE-US-00037 TABLE 29 Mean spot forming cells (SFC) per 10.sup.6 PBMCs for each epitope ± SEM for priming vaccination with VEE-MAG25mer srRNA-LNP1 (300 μg) (Group 3) Antigen Wk Env CL9 Env TL9 Gag CM9 Gag LW9 Pol SV9 Tat TL8 4 19.5 ± 8.7 13.3 ± 13.1 16.5 ± 15.3 10.5 ± 7.3 35.9 ± 24.8 92.9 ± 91.6 5 87.9 ± 43.9 12.7 ± 11.7 37.2 ± 31.9 21.1 ± 23.8 13.2 ± 13.7 12.6 ± 13.7 7 21.1 ± 13.3 48.8 ± 48.4 51.7 ± 39.5 9.1 ± 10.5 58.6 ± 55.8 159.4 ± 159.sup. 8 47.7 ± 21.7 66.4 ± 52.2 59.8 ± 57.4 49.4 ± 28.sup. 79.4 ± 63.sup. 133.8 ± 132.1 10 .sup. 49 ± 30.2 42.2 ± 41.1 139.3 ± 139.3 51.6 ± 51.2 78.2 ± 75.8 131.7 ± 131.6 11 .sup. 42 ± 26.8 20.9 ± 21.4 177.1 ± 162.sup. −6.3 ± 4.3 104.3 ± 104.1 231.5 ± 230.1 12 40.2 ± 19.sup. 20.3 ± 11.9 42.2 ± 46.7 3.7 ± 6.7 .sup. 57 ± 44.7 .sup. 70 ± 69.2 13 81.2 ± 48.9 38.2 ± 37.6 259.4 ± 222.2 −4 ± 4.1 164.1 ± 159.3 347.3 ± 343.5 14 34.5 ± 31.8 5.3 ± 11.6 138.6 ± 137.3 −4.7 ± 5.2 52.3 ± 52.9 142.6 ± 142.6 15 49 ± 24 6.7 ± 9.8 167.1 ± 163.8 −6.4 ± 4.2 47.8 ± 42.3 116.6 ± 114.5

srRNA Dose Ranging Study

[0920] In one implementation of the present invention, an srRNA dose ranging study can be conducted in mamu A01 Indian rhesus macaques to identify which srRNA dose to progress to NHP immunogenicity studies. In one example, Mamu A01 Indian rhesus macaques can be administered with an srRNA vector encoding model antigens that includes multiple mamu A01 restricted epitopes by IM injection. In another example, an anti-CTLA-4 monoclonal antibody can be administered SC proximal to the site of IM vaccine injection to target the vaccine draining lymph node in one group of animals. PBMCs can be collected every 2 weeks after the initial vaccination for immune monitoring. The study arms are described in below (Table 30).

TABLE-US-00038 TABLE 30 Non-GLP RNA dose ranging study in Indian Rhesus Macaques Group Prime Boost 1 Boost 2 1 srRNA-LNP srRNA-LNP srRNA-LNP (Low Dose) (Low Dose) (Low Dose) 2 srRNA-LNP srRNA-LNP srRNA-LNP (Mid Dose) (Mid Dose) (Mid Dose) 3 srRNA-LNP srRNA-LNP srRNA-LNP (High Dose) (High Dose) (High Dose) 4 srRNA-LNP srRNA-LNP srRNA-LNP (High Dose) + (High Dose) + (High Dose) + anti-CTLA-4 anti-CTLA-4 anti-CTLA-4 * Dose range of srRNA to be determined with the high dose

≤300 □ g.

Immunogenicity Study in Indian Rhesus Macaques

[0921] In one implementation of the present invention, vaccine studies can be conducted in mamu A01 Indian rhesus macaques to demonstrate immunogenicity. In one example, Mamu A01 Indian rhesus macaques can be administered an IM injection with a ChAdV and/or srRNA vector encoding model antigens that includes multiple mamu A01 restricted antigens. In another example, an anti-CTLA-4 monoclonal antibody will be administered SC proximal to the site of IM vaccine injection to some of the groups. PBMCs can be collected every 2 weeks after the initial vaccination for immune monitoring. The study arms are described in below (Table 31).

TABLE-US-00039 TABLE 31 Non-GLP immunogenicity study in Indian Rhesus Macaques Group Prime Boost 1 Boost 2 1 ChAdV srRNA-LNP* srRNA-LNP 2 srRNA-LNP ChAdV srRNA-LNP 3 srRNA-LNP srRNA-LNP ChAdV 4 srRNA-LNP + srRNA-LNP + srRNA-LNP + anti-CTLA-4 anti-CTLA-4 anti-CTLA-4 5 ChAdV + srRNA-LNP + srRNA-LNP + anti-CTLA-4 anti-CTLA-4 anti-CTLA-4 6 srRNA-LNP + ChAdV + srRNA-LNP + anti-CTLA-4 anti-CTLA-4 anti-CTLA-4 *srRNA dose to be determined based on srRNA dose range study.

XIX. Identification of MHC/Peptide Target-Reactive T Cells and TCRs

[0922] T cells can be isolated from blood, lymph nodes, or tumors of patients. T cells can be enriched for antigen-specific T cells, e.g., by sorting antigen -MHC tetramer binding cells or by sorting activated cells stimulated in an in vitro co-culture of T cells and antigen -pulsed antigen presenting cells. Various reagents are known in the art for antigen-specific T cell identification including antigen-loaded tetramers and other MHC-based reagents.

[0923] Antigen-relevant alpha-beta (or gamma-delta) TCR dimers can be identified by single cell sequencing of TCRs of antigen-specific T cells. Alternatively, bulk TCR sequencing of antigen-specific T cells can be performed and alpha-beta pairs with a high probability of matching can be determined using a TCR pairing method known in the art.

[0924] Alternatively or in addition, antigen-specific T cells can be obtained through in vitro priming of naïve T cells from healthy donors. T cells obtained from PBMCs, lymph nodes, or cord blood can be repeatedly stimulated by antigen-pulsed antigen presenting cells to prime differentiation of antigen-experienced T cells. TCRs can then be identified similarly as described above for antigen-specific T cells from patients.

Certain Sequences

[0925] Sequences for vectors, cassettes, and antibodies are shown below.

TABLE-US-00040 Tremelimumab VL (SEQ ID NO: 16)
PSSLSASVGDRVTITCRASQSINSYLDWYQQKPCKAPKLLIYAASSLQSGVPSRFSGSGSGTDF
TLTISSLQPEDFATYYCQQYYSTPFTFGPGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPR
EAKV Tremelimumab VH (SEQ ID NO: 17)
GVVQPGRSLRLSCAASGFTFSSYGMHWVRQAPGKGLEWVAVIWYDGSNKYYADSVKGRFTI
SRDNSKNTLYLQMNSLRAEDTAVYYCARDPRGATLYYYYYGMDVWGQGTTVTVSSASTKGPSVFPL
APCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVH Tremelimumab VH CDR1 (SEQ ID
NO: 18) GFTFSSYGMH Tremelimumab VH CDR2 (SEQ ID NO: 19) VIWYDGSNKYYADSV
Tremelimumab VH CDR3 (SEQ ID NO: 20) DPRGATLYYYYYGMDV Tremelimumab VL
CDR1 (SEQ ID NO: 21) RASQSINSYLD Tremelimumab VL CDR2 (SEQ ID NO: 22)
AASSLQS Tremelimumab VL CDR3 (SEQ ID NO: 23) QQYYSTPFT Durvalumab (MEDI4736)
VL (SEQ ID NO: 24)
EIVLTQSPGTLSLSPGERATLSCRASQRVSSSYLAWYQQKPGQAPRLLIYDASSRATGIPDRFSGSGSGT
DFTLTISRLEPEDFAVYYCQQYGSLPWTFGQGTKVEIK MEDI4736 VH (SEQ ID NO: 25)
EVQLVESGGGLVQPGGSLRLSCAASGFTFSRYWMSWVRQAPGKGLEWVANIKQDGSEKYYVDSVKG
RFTISRDNAKNSLYLQMNSLRAEDTAVYYCAREGGWFGELAFDYWGQGTLVTVSS MEDI4736 VH
CDR1 (SEQ ID NO: 26) RYWMS MEDI4736 VH CDR2 (SEQ ID NO: 27)
NIKQDGSEKYYVDSVKG MEDI4736 VH CDR3 (SEQ ID NO: 28) EGGWFGELAFDY
MEDI4736 VL CDR1 (SEQ ID NO: 29) RASQRVSSSYLA MEDI4736 VL CDR2 (SEQ ID
NO: 30) DASSRAT MEDI4736 VL CDR3 (SEQ ID NO: 31) QQYGSLPWT UbA76-25
merPDTT nucleotide (SEQ ID NO: 32)
GCCCGGGCATTTAAATGCGATCGCATCGATtacgactctagaatagtctagtccgcaggccaccatgCAGATCT
TCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATCGAGAA
CGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCCG
GCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCAC
CTGGTCCTTCGCCTGAGAGGTGcCatgtttcaggcgctgagcgaaggctgcaccccgtatgatattaaccagatgctgaacgtgctgggcga
tcatcaggtctcaggccttgagcagcttgagagtataatcaactttgaaaaactgactgaatggaccagttctaatgttatgCCTATCCTGTCTCCTCT
GACAAAGGGCATCCTGGGCTTCGTGTTTACCCTGACCGTGCCTTCTGAGAGAGGACTTagctgcattagc
gaagcggatgcgaccaccccggaaagcgcgaacctgggcgaagaaattctgagccagctgtatctttggccaagggtgacctaccattcccctagttatgcttac

caccaatttgaaagacgagccaaatataaaagaCACTTCCCCGGCTTTGGCCAGAGCCTGCTGTTTGGCTACCCTGTGT
ACGTGTTCGGCGATTGCGTGCAGGGCGATtgggatgcgattcgctttcgctattgcgcgccgccgggctatgcgctgctgcgctgcaa
cgataccaactatagcgctctgctggctgtgggggccctagaaggacccaggaatcaggactggcttggtgtcccaagacaacttgtaactCGGATGCA
GGCTATTCAGAATGCCGGCCTGTGTACCCTGGTGGCCATGCTGGAAGAGACAATCTTCTGGCTGC
AAgcgtttctgatggcgctgaccgatagcggcccgaaaaccaacattattgtggatagccagtatgtgatgggcattagcaaaccgagctttcaggaatttgtgg
attgggaaaacgtgagcccggaactgaacagcaccgatcagccgtttTGGCAAGCCGGAATCCTGGCCAGAAATCTGGTGCC
TATGGTGGCCACAGTGCAGGGCCAGAACCTGAAGTACCAGggtcagtcactagtcatctctgcttctatcattgtcttcaacct
gCtggaactggaaggtgattatcgagatgatggcaacgtgtgggtgcataccccgctgagcccgcgcaccctgaacgcgtgggtgaaagcggtggaagaaaa
aaaaggtattccagttcacctagagctggccagtatgaccaacaTggagctcatgagcagtattgtgcatcagcaggtcAGAACATACGGCCCCG
TGTTCATGTGTCTCGGCGGACTGCTTACAATGGTGGCTGGTGCTGTGTGGCTGACAGTGcgagtgctcga
gctgttccgggccgcgcagctggccaacgacgtggtcctccagatcatggagctttgtggtgcagcgtttcgccaggtgtgccataccaccgtgccgtggccga
acgcgagcctgaccccgaaatggaacaacgaaaccacccagccccagatcgccaactgcagcgtgtatgacttttttgtgtggctccattattattctgttcgagac
acactttggccaagggtgacctaccatatgaacaaatatgcgtatcatatgctggaaagacgagccaaatataaaagaGGACCAGGACCTGGCGC
TAAATTTGTGGCCGCCTGGACACTGAAAGCCGCTGCTGGTCCTGGACCTGGCCAGTACATCAAGG
CCAACAGCAAGTTCATCGGCATCACCGAACTCGGACCCGGACCAGGCTGATGATTTCGAAATTTA
AATAAGCTTGCGGCCGCTAGGGATAACAGGGTAATtatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtgg
UbA76-25 merPDTT polypeptide (SEQ ID NO: 33)
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKES
TLHLVLRLRGAMFQALSEGCTPYDINQMLNVLGDHQVSGLEQLESIINFEKLTEWTSSNVMPILSPLTK
GILGFVFTLTVPSERGLSCISEADATTPESANLGEEILSQLYLWPRVTYHSPSYAYHQFERRAKYKRHFP
GFGQSLLFGYPVYVFGDCVQGDWDAIRFRYCAPPGYALLRCNDTNYSALLAVGALEGPRNQDWLGV
PRQLVTRMQAIQNAGLCTLVAMLEETIFWLQAFLMALTDSGPKTNIIVDSQYVMGISKPSFQEFVDWE
NVSPELNSTDQPFWQAGILARNLVPMVATVQGQNLKYQGQSLVISASIIVFNLLELEGDYRDDGNVW
VHTPLSPRTLNAWVKAVEEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVFMCLGGLLTMVAGA
VWLTVRVLELFRAAQLANDVVLQIMELCGAAFRQVCHTTVPWPNASLTPKWNNETTQPQIANCSVY
DFFVWLHYYSVRDTLWPRVTYHMNKYAYHMLERRAKYKRGPGPGAKFVAAWTLKAAAGPGPGQY
IKANSKFIGITELGPGPG MAG-25 merPDTT nucleotide (SEQ ID NO: 34)
ATGGCCGGGATGTTCCAGGCACTGTCCGAAGGCTGCACACCCTATGATATTAACCAGATGCTGAA
TGTCCTGGGAGACCACCAGGTCTCTGGCCTGGAGCAGCTGGAGAGCATCATCAACTTCGAGAAGC
TGACCGAGTGGACAAGCTCCAATGTGATGCCTATCCTGTCCCCACTGACCAAGGGCATCCTGGGC
TTCGTGTTTACCCTGACAGTGCCTTCTGAGCGGGGCCTGTCTTGCATCAGCGAGGCAGACGCAACC
ACACCAGAGTCCGCCAATCTGGGCGAGGAGATCCTGTCTCAGCTGTACCTGTGGCCCCGGGTGAC
ATATCACTCCCCTTCTTACGCCTATCACCAGTTCGAGCGGAGAGCCAAGTACAAGAGACACTTCCC
AGGCTTTGGCCAGTCTCTGCTGTTCGGCTACCCCGTGTACGTGTTCGGCGATTGCGTGCAGGGCGA
CTGGGATGCCATCCGGTTTAGATACTGCGCACCACCTGGATATGCACTGCTGAGGTGTAACGACA
CCAATTATTCCCCCCTGCTGGCAGTGGGCGCCCTGGAGGGCCCTCGCAATCAGGATTGGCTGGGC
GTGCCAAGGCAGCTGGTGACACGCATGCAGGCCATCCAGAACGCAGGCCTGTGCACCCTGGTGGC
AATGCTGGAGGAGACAATCTTCTGGCTGCAGGCCTTTCTGATGGCCCTGACCGACAGCGGCCCCA
AGACAAACATCATCGTGGATTCCCAGTACGTGATGGGCATCTCCAAGCCTTCTTTCCAGGAGTTTG
TGGACTGGGAGAACCTGAGCCCAGAGCTGAATTCCACCGATCAGCCATTCTGGCAGGCAGGAATC
CTGGCAAGGAACCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAATCTGAAGTACCAGGGCC
AGAGCCTGGTCATCAGCGCCTCCATCATCGTGTTTAACCTGCTGGAGCTGGAGGGCGACTATCGG
GACGATGGCAACGTGTGGGTGCACACCCCACTGAGCCCCAGAACACTGAACGCCTGGGTGAAGG
CCGTGGAGGAGAAGAAGGGCATCCCAGTGCACCTGGAGCTGGCCTCCATGACCAATATGGAGCT
GATGTCTAGCATCGTGCACCAGCAGGTGAGGACATACGGACCCGTGTTCATGTGCCTGGGAGGCC
TGCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAGTGCGGGTGCTGGAGCTGTTCAGAGCCGCC
CAGCTGGCCAACGATGTGGTGCTGCAGATCATGGAGCTGTGCGGAGCAGCCTTTCGCCAGGTGTG
CCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAGTQGAACAATGAGACAACACAGC
CTCAGATCGCCAACTGTAGCGTGTACGACTTCTTCGTGTGGCTGCACTACTATAGCGTGAGGGATA
CCCTGTGGCCCCCCGTGACATACCACATGAATAAGTACGCCTATCACATGCTGGAGAGGCGCGCC
AAGTATAAGAGAGGCCCTGGCCCAGGCGCAAAGTTTGTGGCAGCATGGACCCTGAAGGCCGCCG
CCGGCCCCGGCCCCGGCCAGTATATCAAGGCTAACAGTAAGTTCATTGGAATCACAGAGCTGGGA
CCCGGACCTGGA MAG-25 merPDTT polypeptide (SEQ ID NO: 35)
MAGMFQALSEGCTPYDINQMLNVLGDHQVSGLEQLESIINFEKLTEWTSSNVMPILSPLTKGIL
GFVFTLTVPSERGLSCISEADATTPESANLGEEILSQLYLWPRVTYHSPSYAYHQFERRAKYKRHFPGF
GQSLLFGYPVYVFGDCVQGDWDAIRFRYCAPPGYALLRCNDTNYSALLAVGALEGPRNQDWLGVPR
QLVTRMQAIQNAGLCTLVAMLEETIFWLQAFLMALTDSGPKTNIIVDSQYVMGISKPSFQEFVDWENV

SPELNSTDQPFWQAGILARNLVPMVATVQGQNLKYQGQSLVISASIIVFNLLELEGDYRDDGNVWVHT
PLSPRTLNAWVKAVEEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVFMCLGGLLTMVAGAVWL
TVRVLELFRAAQLANDVVLQIMELCGAAFRQVCHTTVPWPNASLTPKWNNETTQPQIANCSVYDFFV
WLHYYSVRDTLWPRVTYHMNKYAYHMLERRAKYKRGPGPGAKFVAAWTLKAAAGPGPGQYIKAN
SKFIGITELGPGPG Ub7625 merPDTT_NoSFL nucleotide (SEQ ID NO: 36)
GCCCGGGCATTTAAATGCGATCGCATCGATtacgactctagaatagtctagtccgcaggccaccatgCAGATCT
TCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATCGAGAA
CGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCCG
GCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCAC
CTGGTCCTTCGCCTGAGAGGTGcCatgtttcaggcgctgagcgaaggctgcaccccgtatgatattaaccagatgctgaacgtgctgggcga
tcatcagtttaagcacatcaaagcctttgaccggacatttgctaacaacccaggtcccatggttgtgtttgccacacctgggCCTATCCTGTCTCCTCT
GACAAAGGGCATCCTGGGCTTCGTGTTTACCCTGACCGTGCCTTCTGAGAGAGGACTTagctgcattagc
gaagcggatgcgaccaccccggaaagcgcgaacctgggcgaagaaattctgagccagctgtatctttggccaagggtgacctaccattcccctagttatgcttac
caccaatttgaaagacgagccaaatataaaagaCACTTCCCCGGCTTTGGCCAGAGCCTGCTGTTTGGCTACCCTGTGT
ACGTGTTCGGCGATTGCGTGCAGGGCGATtgggatgcgattcgctttcgctattgcgcgccgccgggctatgcgctgctgcgctgcaa
cgataccaactatagcgctctgctggctgtgggggccctagaaggacccaggaatcaggactggcttggtgtcccaagacaacttgtaactCGGATGCA
GGCTATTCAGAATGCCGGCCTGTGTACCCTGGTGGCCATGCTGGAAGAGACAATCTTCTGGCTGC
AAgcgtttctgatggcgctgaccgatagcggcccgaaaaccaacattattgtggatagccagtatgtgatgggcattagcaaaccgagctttcaggaatttgtgg
attgggaaaacgtgagcccggaactgaacagcaccgatcagccgtttTGGCAAGCCGGAATCCTGGCCAGAAATCTGGTGCC
TATGGTGGCCACAGTGCAGGGCCAGAACCTGAAGTACCAGggtcagtcactagtcatctctgcttctatcattgtcttcaacct
gCtggaactggaaggtgattatcgagatgatggcaacgtgtgggtgcataccccgctgagcccgcgcaccctgaacgcgtgggtgaaagcggtggaagaaaa
aaaaggtattccagttcacctagagctggccagtatgaccaacaTggagctcatgagcagtattgtgcatcagcaggtcAGAACATACGGCCCCG
TGTTCATGTGTCTCGGCGGACTGCTTACAATGGTGGCTGGTGCTGTGTGGCTGACAGTGcgagtgctcga
gctgttccgggccgcgcagctggccaacgacgtggtcctccagatcatggagctttgtggtgcagcgtttcgccaggtgtgccataccaccgtgccgtggccga
acgcgagcctgaccccgaaatggaacaacgaaaccacccagccccagatcgccaactgcagcgtgtatgacttttttgtgtggctccattattattctgttcgagac
acactttggccaagggtgacctaccatatgaacaaatatgcgtatcatatgctggaaagacgagccaaatataaaagaGGACCAGGACCTGGCGC
TAAATTTGTGGCCGCCTGGACACTGAAAGCCGCTGCTGGTCCTGGACCTGGCCAGTACATCAAGG
CCAACAGCAAGTTCATCGGCATCACCGAACTCGGACCCGGACCAGGCTGATGATTTCGAAATTTA
AATAAGCTTGCGGCCGCTAGGGATAACAGGGTAATtatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtgg
Ub7625 merPDTT_NoSFL polypeptide (SEQ ID NO: 37)
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKES
TLHLVLRLRGAMFQALSEGCTPYDINQMLNVLGDHQFKHIKAFDRTFANNPGPMVVFATPGPILSPLT
KGILGFVFTLTVPSERGLSCISEADATTPESANLGEEILSQLYLWPRVTYHSPSYAYHQFERRAKYKRHF
PGFGQSLLFGYPVYVFGDCVQGDWDAIRFRYCAPPGYALLRCNDTNYSALLAVGALEGPRNQDWLG
VPRQLVTRMQAIQNAGLCTLVAMLEETIFWLQAFLMALTDSGPKTNIIVDSQYVMGISKPSFQEFVDW
ENVSPELNSTDQPFWQAGILARNLVPMVATVQGQNLKYQGQSLVISASIIVFNLLELEGDYRDDGNVW
VHTPLSPRTLNAWVKAVEEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVFMCLGGLLTMVAGA
VWLTVRVLELFRAAQLANDVVLQIMELCGAAFRQVCHTTVPWPNASLTPKWNNETTQPQIANCSVY
DFFVWLHYYSVRDTLWPRVTYHMNKYAYHMLERRAKYKRGPGPGAKFVAAWTLKAAAGPGPGQY
IKANSKFIGITELGPGPG ChAdV68.5WTnt.MAG25 mer (SEQ ID NO: 2); AC_000011.1 with
El (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 CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGA
ATTTGGGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGAC
GTTTTGATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAA
CGAGGTGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTTT
CTGGGCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAATCT
GAGTAATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGATTA
CGTGGGGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGTGTTTT
TACGTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCCACTCTTG
AGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAAGTAGGGAT
AACAGGGTAATgacattgattattgactagttGttaaTAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATAT
ATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCCCTGGCTGACCGCCCAACGACCCCCG
CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCA
ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTC
CGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAC

GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGgTGATGCGGTTTTG
GCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG
ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCC
GCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAgcTCGTTTA
GTGAACCGTCAGATCGCCTGGAACGCCATCCACGCTGTTTTGACCTCCATAGAAGACAGCGATCG
CGccaccATGGCCGGGATGTTCCAGGCACTGTCCGAAGGCTGCACACCCTATGATATTAACCAGATG
CTGAATGTCCTGGGAGACCACCAGGTCTCTGGCCTGGAGCAGCTGGAGAGCATCATCAACTTCGA
GAAGCTGACCGAGTGGACAAGCTCCAATGTGATGCCTATCCTGTCCCCACTGACCAAGGGCATCC
TGGGCTTCGTGTTTACCCTGACAGTGCCTTCTGAGCGGGGCCTGTCTTGCATCAGCGAGGCAGACG
CAACCACACCAGAGTCCGCCAATCTGGGCGAGGAGATCCTGTCTCAGCTGTACCTGTGGCCCCGG
GTGACATATCACTCCCCTTCTTACGCCTATCACCAGTTCGAGCGGAGAGCCAAGTACAAGAGACA
CTTCCCAGGCTTTGGCCAGTCTCTGCTGTTCGGCTACCCCGTGTACGTGTTCGGCGATTGCGTGCA
GGGCGACTGGGATGCCATCCGGTTTAGATACTGCGCACCACCTGGATATGCACTGCTGAGGTGTA
ACGACACCAATTATTCCGCCCTGCTGGCAGTGGGCGCCCTGGAGGGCCCTCGCAATCAGGATTGG
CTGGGCGTGCCAAGGCAGCTGGTGACACGCATGCAGGCCATCCAGAACGCAGGCCTGTGCACCCT
GGTGGCAATGCTGGAGGAGACAATCTTCTGGCTGCAGGCCTTTCTGATGGCCCTGACCGACAGCG
GCCCCAAGACAAACATCATCGTGGATTCCCAGTACGTGATGGGCATCTCCAAGCCTTCTTTCCAGG
AGTTTGTGGACTGGGAGAACGTGAGCCCAGAGCTGAATTCCACCGATCAGCCATTCTGGCAGGCA
GGAATCCTGGCAAGGAACCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAATCTGAAGTACCA
GGGCCAGAGCCTGGTCATCAGCGCCTCCATCATCGTGTTTAACCTGCTGGAGCTGGAGGGCGACT
ATCGGGACGATGGCAACGTGTGGGTGCACACCCCACTGAGCCCCAGAACACTGAACGCCTGGGTG
AAGGCCGTGGAGGAGAAGAAGGGCATCCCAGTGCACCTGGAGCTGGCCTCCATGACCAATATGG
AGCTGATGTCTAGCATCGTGCACCAGCAGGTGAGGACATACGGACCCGTGTTCATGTGCCTGGGA
GGCCTGCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAGTGCGGGTGCTGGAGCTGTTCAGAGC
CGCCCAGCTGGCCAACGATGTGGTGCTGCAGATCATGGAGCTGTGCGGAGCAGCCTTTCGCCAGG
TGTGCCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAGTGGAACAATGAGACAACA
CAGCCTCAGATCGCCAACTGTAGCGTGTACGACTTCTTCGTGTGGCTGCACTACTATAGCGTGAGG
GATACCCTGTGGCCCCGCGTGACATACCACATGAATAAGTACGCCTATCACATGCTGGAGAGGCG
CGCCAAGTATAAGAGAGGCCCTGGCCCAGGCGCAAAGTTTGTGGCAGCATGGACCCTGAAGGCC
GCCGCCGGCCCCGGCCCCGGCCAGTATATCAAGGCTAACAGTAAGTTCATTGGAATCACAGAGCT
GGGACCCGGACCTGGATAATGAGTTTAAACTCCCATTTAAATGTGAGGGTTAATGCTTCGAGCAG
ACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTT
ATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAAC
AACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCAAG
TAAAACCTCTACAAATGTGGTAAAATAACTATAACGGTCCTAAGGTAGCGAGTGAGTAGTGTTCT
GGGGCGGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTGTTGCAG
CAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCC
CCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCC
GCGAACTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGCA
GCTGCTGCATCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCT
GGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGA
TGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAG
GAGCAGACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAATAAATAAACG
GAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTCGCGCGCGGTAGGCCCT
GGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAGGACCCGGTAGAGGTGGGCTT
GGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCTCCATTGCAGGGCCTCG
TGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATGGTGTTGCACAAT
ATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTGTTGA
GCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATG
TTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACCACCAGCACGGTGTATCCGGTGCA
CTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCC
CGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCGATGGGCCCGTGGGGGGGGCCTGGGCA
AAGACGTTTCGGGGGTCGGACACATCATAGTTGTGGTCCTGGGTGAGGTCATCATAGGCCATTTT
AATGAATTTGGGGCGGAGGGTGCCGGACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAG
TTCCCCTCACAGATCTGCATCTCCCAGGCTTTGAGCTCGGAGGGGGGGATCATGTCCACCTGCGGG
GCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGCCGAAAGCAAGTTCCGGAGCA
GCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGATGACCGGCTGCAGGTGGTAGTTG

AGGGAGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCCACCTCGTTCATCATCTCGCGCACGTG
CATGTTCTCGCCCACCAGTTCCGCCAGGAGGCGCTCTCCCCCCAGGGATAGGAGCTCCTGGAGCG
AGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGCCATGGGCATTTTGGAGAGGGTTTGTTGCAAG
AGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGT
TTCGCGGGTTGGGACGGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGT
CCGGTCCTTCCAGGGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTCACGGTGAAGGGGTGCGCGC
CGGGCTGGGCGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCG
GCGCCCTGCGCGTCGGCCAGGTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTG
GCCTTTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGGGACTTGAGGG
CGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGGGGCGTAGGCGTCCGCGCCGCAGTGGGCGCA
GACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCAGTTTCCCGC
CGTTCTTTTTGATGCCTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGGGTGACAAAGAG
GCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCCGCGGTCCTCCTC
GTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAAGGAGGCCACG
TGGGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAAACACATGTC
CCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCCGG
CCGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCTCACTGTCTTCCGGATCGCTGTCCAGG
AGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTTGTC
AGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCT
CGTCCATCTGGTCAGAAAAGACGATCTTTTTGTTGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGG
GCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTGGTTTTTTTCCTTGTCGGCGCGCTCCTTG
GCGGCGATGTTGAGCTGCACGTACTCGCGCGCCACGCACTTCCATTCGGGGAAGACGGTGGTCAG
CTCGTCGGGCACGATTCTGACCTGCCAGCCCCGATTATGCAGGGTGATGAGGTCCACACTGGTGG
CCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGCCCTTGCGCGAGCAGAAGGGG
GGCAGGGGGTCCAGCATGACCTCGTCGGGGGGGTCGGCATCGATGGTGAAGATGCCGGGCAGGA
GGTCGGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCATTCGCGCACG
GCCAGCGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGAGG
CGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAG
CAGCGCCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCC
CCGGGCCCAGGTTGGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGC
GAGTTGGAGGAGATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGT
CGCGGATGAAGTGGGCGTAGGAGTCTTGCAGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCC
AGAGCGCAGTAGTCGAGGGTCTCCTGGATGATGTCATACTTGAGCTGTCCCTTTTGTTTCCACAGC
TCGCGGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCGAGGGGGAACCCGTCCTGATCT
GCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTGTAGGCGCAGCAGCCCTTCTCCAC
GGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGTCCCTG
ACCATGACCTTGAGGAACTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCCCAGAGCTG
GAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGAGGATCT
TGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTG
ATGACCTGGGCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCCACGATGTAGAGTTC
CACGAATCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGG
GGTCGCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGA
AGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGTACTGACGGAACTGCTGCCCGA
CGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCCCCGTGCCAGCGATCCCATTTG
AGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCGGTCGTCCCCGGAGAGTTTCATGACCA
GCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCAGGTGTAGGTTTCCACATCGTAGGTG
AGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACTGGATCTCCTGCCACCAATT
GGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAACACTCGTGCTTGT
GTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACGAGCTGTACC
TGAGTTCCTTTGACGAGGAATTTCAGTGGGAAGTGGAGTCGTGGCGCCTGCATCTCGTGCTGTACT
ACGTCGTGGTGGTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGG
AGGCAGGTCCAGACCTCGGCGCGAGCGGGTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAG
CTGTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGTTGACTTG
CAGGAGTTTTTCCAGGGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGGTGG
CGACGTCGATGGCTTGCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGG
GCGGCTGGGGCGACGGGGGGGGTGCCTCTTCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCCG
GGCGGCAGGGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTCGGCGCCGCGCGCG

GGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACGTCCTG
GATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTGAAAGAGAGTTCGACAG
AATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCTTGCACGTCGCCCGAGTTGTCCT
GGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCCGGCGCGCT
CCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTCATGCCCGCC
TCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGC&GGCGCGCATGACCACCTGGGC
GAGGTTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGA
GCGTGGTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGGAGCGGCATCTCGCTG
ACGTCGCCCAGCGCCTCCAAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAAACTG
GGAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGATGAGCTCGGCGATGGTGGCGC
GCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCCTCTTCTTCCTCCTCCACTAACATCTC
TTCTACTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAGGGGGCCTGCGTCGCCGGCGGCGCACGG
GCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCGGCGTCGCATGGTCTCGGTGACGGCG
CGCCCGTCCTCGCGGGGCCGCAGCGTGAAGACGCCGCCGCGCATCTCCAGGTGGCCGGGGGGGTC
CCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCCCGTAGGGACTCCGCGCA
AGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGAAGGCTTCGAGCCAGTCG
CAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGGGGGC
GATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACC
AGGTCTTTGGGCCCGGCTTGCTGGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGGTCCTGACA
CCTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCG
GCCGTGCATGCGCGTGAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTCGGCGACGACG
CGCTCGGCGAGGATGGCTTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAA
GCGGTGGTAGGCTCCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACGGACCAGTTGACGGTCT
GGTGGCCCGGACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGAAGATGTA
GTCGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGGCGGTAG
AGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTGGTAGC
CGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGAACTCGCG
GACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTGGCCCGTGA
GGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGGCTCGACTCCGTG
GCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAATCTCGAATCAGGCTG
GAGCCGCAGCTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAGGAT
ACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCGGATGAGACTAGTAAGCGCGGAAAGCG
GCCGACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCC
CCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAGACC
CCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCTCTTTTGTTTTGTTTGTTTTTGCCAGAT
GCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTCCACCGCAACAACAGCCCCCTCCACAGC
CGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGACCGCCGCGGCCGCCGTGAGCGGG
GCTGGACAGAGTTATGATCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTGGCGCGCCTGGGGG
CGTCGTCGCCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGAGGCCTACGTGCC
CAAGCAGAACCTGTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGATGCGCGCGGCCCGGTTC
CACGCGGGGGGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGTGCTGAGGGACGAGGATTTCG
AGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCAC
GGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGC
ACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTGGAGGCCAT
CGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGCAGCTGTTCCTGGTGGTGCAGCATAGTCGGG
ACAACGAAGCGTTCAGGGAGGCGCTGCTGAATATCACCGAGCCCGAGGGCCGCTGGCTCCTGGAC
CTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGCTGCCGCTGTCCGAGAAGCTGGC
GGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGGAAGATCTACAAGACCCCGT
ACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTG
ACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAGCGCCAGCA
GGCGGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGG
GACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCC
TTGGAGGCGGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGT
ACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATC
CCGCGATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACCATTGGACCCAG
GCCATGCAACGCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGC
CAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCCCACGCACGAGAAGG

TCCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCGACGAGGCCGGCCTGGTG
TACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAACGTGCAGACCAACCTGGACCG
CATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCACCGCGAGTCCAACCTGG
GATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCGCCAACGTGCCCCGGGGCCAGGAG
GACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGGTGTA
CCAGTCCGGGCCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCGTGAACCTGAGCC
AGGCTTTCAAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCCCCGGTCGGGGACCGCGCGACGGTG
TCGAGCCTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGG
CAGCATCAACCGCAACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGG
CGCACGTGGACGAGCAGACCTACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGA
CCCGGGCAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGATCCCGCCCC
AGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTGCGTTACGTGCAGCAGAGCGTGGGCCTGTTC
CTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGGAGCCCA
GCATGTACGCCAGCAACCGCCCGTTCATCAATAAACTGATGGACTACTTGCATCGGGCGGCCGCC
ATGAACTCTGACTATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCGCCGGGGTTCTAC
ACGGGCGAGTACGACATGCCCGACCCCAATACGGGTTCCTGTGGGACGATGTGGACAGCAGCGT
GTTCTCCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCC
CGTCCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCTTTC
CCGAGCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCG
CTTGCTGGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCC
CCAATAACGGGATAGAAAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCA
CAGGGACGATCCCCGGGCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCGG
TGGCACGACAGGCAGCGGGGACAGATGTGGGACGATGAGGACTCCGCCGACGACAGCAGCGTGT
TGGACTTGGGTGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGGCGCATGATGTAA
GAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCT
GTTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTCCCTCGTACGAGAGC
GTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTACGTGCCCCC
GCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTACG
ATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAAC
GACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCAC
CCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGGGCCAGCTGAAAACCATCATGCACACCA
ACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGC
AAGACCCCCAATGGGGTGACAGTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATG
AATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATCGACCTGATGAAC
AACGCCATCATCGACAATTACTTGGCGGTGGGGCGGCAGAACGGGGTGCTGGAGAGCGACATCG
GCGTGAAGTTCGACACTAGGAACTTCAGGCTGGGCTGGGACCCCGTGACCGAGCTGGTCATGCCC
GGGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTC
ACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCAGCCCTTCCAGGAAGGCTTCCA
GATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTCGACGCCTATGAGA
AAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTCAG
GGGCGATAATTTTGCAAGCGCCGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAG
ATAGTCATTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGA
TAAACACCGCCTACCGCAGCTGGTACCTAGCCTACAACTATGGCGACCCCGAGAAGGGCGTGCGC
TCCTGGACGCTGCTCACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCC
GACATGATGCAAGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGG
CGCCGAGCTCCTGCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCT
GCGCGCCTTCACCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCC
GCCCGCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGC
CGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTGC
CCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTAAATG
TCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGATGTACGG
AGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGGG
GCGCCCTCAAGGGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCC
GACGCGCGCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCGT
GGTGGCcGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACC
GGAGCACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAG
GGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGC

GCGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGGAACGTGTACT
GGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCCCCTCGCACTTGAAGA
TGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCCAAGCGCAAATTCAAGGAA
GAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGAAGGAGGAAAGAAAGC
CCCCCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTGATGTGGACGGATTGGT
GGAGTTTGTGCGCGAGTTCGCCCCCCCGGCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCG
GTGCTGAGACCCGGCACCACCGTGGTCTTCACGCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCG
CTCCTACGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGT
TTGCTTACGGCAAGCGCAGCCGTTCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCAC
GGCAACCCCACGCCGAGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCG
CCGGGGGTTCAAGCGCGAGGGCGAGGATCTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCC
AGAAGCTGGAAGACGTGCTGGAGACCATGAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGT
GCGGCCCATCAAGCAGGTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACATCAAGATTCCCACGG
AGCCCATGGAAACGCAGACCGAGCCCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGAC
GGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTGC
TGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGCTTCT
ACCGCGGTCATACCAGCAGCCGCCGCCGCAAGACCACCACTCGCCGCCGCCGTCGCCGCACCGCC
GCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCT
GCCGCGCGCGCGCTACCACCCGAGCATCGCCATTTAAACTTTCGCCTGCTTTGCAGATCAATGGCC
CTCACATGCCGCCTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCT
GGCGGGGAACGGGATGCGTCGCCACCACCACCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGG
GGAGGCTTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGC
TTCCGTGGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACTTGGAAACATCTTGTAATAAACCaA
TGGACTCTGACGCTCCTGGTCCTGTGATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTTCGT
CCCTGGCTCCGCGACACGGCACGCGGCCGTTCATGGGCACCTGGAGCGACATCGGCACCAGCCAA
CTGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGAATTTCGGGTCCACGCT
TAAAACCTATGGCAGCAAGGCGTGGAACAGCACCACAGGGCAGGCGCTGAGGGATAAGCTGAAA
GAGCAGAACTTCCAGCAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGGGGGGGGGACCT
GGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCCGCCCGCCGGCTCCG
TGGAGATGCCGCAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCG
CCCCGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTG
AAACTGGGTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAA
GCCCGCGACCCTGGACTTGCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCC
GCCGGTGGCCGTGGCCCGCGCGCGACCCGGGGGCACCGCCCGCCCTCATGCGAACTGGCAGAGC
ACTCTGAACAGCATCGTGGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTA
CCGTAGCGCTTAACTTGCTTGTCTGTGTGTGTATGTATTATGTCGCCGCCGCCGCTGTCCACCAGA
AGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGATGGCCACCCCATCGATGCTGCCCCAGTG
GGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTG
CCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACG
CACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGA
CAACACCTACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACATGG
CCAGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTCCGGCA
CCGCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAAAGCCGAT
GGTGAAACTGCCACAGAAAAAACCTATACATATGGAAATGCACCCGTGCAGGGCATTAACATCAC
AAAAGATGGTATTCAACTTGGAACTGACACCGATGATCAGCCAATCTACGCAGATAAAACCTATC
AGCCTGAACCTCAAGTGGGTGATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGA
GGCAGAGCTCTTAAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAAT
AAAGAAGGAGGTCAGGCAAATGTGAAAACAGGAACAGGCACTACTAAAGAATATGACATAGACA
TGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAATTGTTTTGTATACTG
AAAATGTGGATTTGGAAACTCCAGATACCCATATTGTATACAAAGCAGGCACAGATGACAGCAGC
TCTTCTATTAATTTGGGTCAGCAAGCCATGCCCAACAGACCTAACTACATTGGTTTCAGAGACAAC
TTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCAGGCTTCTCA
GCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCCTACCAGCTCTTGCTTGACTC
TCTGGGTGACAGAACCCGGTATTTCAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATG
TGCGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTG
TTGGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACTGATCAAACCACATGGACCAAA
GATGACAGTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACAT

CCAAGCCAACCTGTGGAGGAACTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAA
GTACACGCCGGCCAATGTTACCCTGCCCACCAACACCAACACCTACGATTACATGAACGGCCGGG
TGGTGGCGCCCTCGCTGGTGGACTCCTACATCAACATCGGGGCGCGCTGGTCGCTGGATCCCATG
GACAACGTGAACCCCTTCAACCACCACCGCAATGCGGGGCTGCGCTACCGCTCCATGCTCCTGGG
CAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCAAGAGCCTCCT
GCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACATGATCCTGCAGA
GCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCAGCATCAACCTCTACG
CCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAACGACACC
AACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAACGC
CACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCTTCCGCGGCTGGTCCTTCACGCG
TCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCTC
CATCCCCTACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGA
CTCCTCCGTCAGCTGGCCCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAAGCGCA
CCGTCGACGGCGAGGGCTACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAG
ATGCTGGCCCACTACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCAT
GTACTCCTTCTTCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGG
ACTACCAGGCCGTCACCCTGGCCTACCAGCACAACAACTCGGGCTTCGTCGGCTACCTCGCGCCC
ACCATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGT
CACCAGCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCA
ACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGCCCACG
CGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTTCGAAG
TCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCGCACCC
CCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGG
CGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCTGGGCACCTTCG
ATAAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGC
CGCGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGCTCGAACACCTGCTACCT
CTTCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCC
TGCTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTGCGTCACCCTGGAAAAGTCCACCCAGACC
GTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTGCATGTTCCTGCACGCCTTCGTGCAC
TGGCCCGACCGCCCCATGGACAAGAACCCCACCATGAACTTGCTGACGGGGGTGCCCAACGGCAT
GCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTCTACCGCTTCCTCA
ACTCCCACTCCGCCTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCACCGCCTTCGACCGCA
TGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCACTTTCATGTTAC
ACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTTCTGCCGGGTCTCGGCATGGCCCGCG
GGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACTCGGGGATCAGCAGTTTGGG
CAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGGT
CGGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACG
GGGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTCGCGTCGGT
GATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCATCCCGAAGGGGGTCATCTTGCAGGTCTGCCT
TCCCATGGTGGGCACGCACCCGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCT
GGGCCTGGTCGGCGTTCATCCCCGGGTACATGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCCT
GCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCCGCAGGACTTGCTAGAGAACTGGTTGGTGGCG
CACCCGGCGTCGTGCACGCAGCAGCGCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCCCCA
GCGGTTCTGGGTGATCTTGGCCCGGTCGGGGTTCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTCGC
CACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGCTTGCC
CTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCCAGTTCTTGTGGGCGA
TCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGGTCTTGTTG
CTAGTGAAGGTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTA
CACCTCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCTCCACGCGGTAGCGGTC
CATCAGCATAGTCATGATTTCCATACCCTTCTCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTT
CTTCACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCTCGTCCAGGGTCTCAAAGCT
CCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGGTAGCTGAAGCCCACGGCCGCCAGCTCCTC
CTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCTGCAGGACCACATGCTTGGTCTTGCG
GGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGTTGGAGATGGCGAGGGGGAGCGCGAGTTCT
CGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCGAGGCCACGCGGCGGTAGGTATGTCTCTTCG
GGGGCAGAGGCGGAGGCGACGGGCTCTCGCCGCCGCGACTTGGCGGATGGCTGGCAGAGCCCCT
TCCGCCTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCCTCCGCGGCCGGCCATTGTGTT

CTCCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACCTCGCCATCTGCCCCCACCG
CCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCCGCCCAGCCCCGCCACCTC
CGACGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGA
CCCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTTCACAAGAAGAGATACACCAAGAACA
GCCAGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTC
CACCTGAGCGGGGGGGAGGACGCGCTCATCAAGCATCTGGCCCGGCAGGCCACCATCGTCAAGG
ATGCGCTGCTCGACCGCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGTTG
AACCTCTTCTCGCCGCGCGTGCCCCCCAAGCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCG
CCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAA
CCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCACCCGCGCCGACGCCCTTTTCAACCTGGGTCC
CGGCGCCCGCCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAAGATCTTCGAGGGTCTGGGCA
GCGACGAGACTCGGGCCGCGAACGCTCTGCAAGGAGAAGGAGGAGAGCATGAGCACCACAGCGC
CCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGAGCTGACCC
ATTTCGCCTACCCGGCTCTGAACCTGCCCCCCAAAGTCATGAGCGCGGTCATGGACCAGGTGCTC
ATCAAGCGCGCGTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCG
TGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCCTAATGCTAGTCCCCAGAGTTTGGAAGAG
CGGCGCAAACTCATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTT
CGCCGACGCGGAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCG
TGCGCCAGGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTG
CACGAGAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCGGGGAGGCCCGGCGCGACT
ACATCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGACGGGCATGGGCGTGTGGCAG
CAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCCTGCAGAAGAACCTCAAGGGTCT
GTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACCTCATTTTCCCCGAGCGCC
TCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGCATGTTGCAAAACTTTCGC
TCTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGGACTTCGTG
CCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCAAC
TACCTGGCCTACCACTCGGACGTGATCGAGGACGTCAGCGGCGAGGGCCTGCTCGAGTGCCACTG
CCGCTGCAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCC
AGATCATCGGCACCTTCGAGTTGCAAGGGCCCAGCGAAGGCGAGGGTTCAGCCGCCAAGGGGGG
TCTGAAACTCACCCCGGGGCTGTGGACCTCGGCCTACTTGCGCAAGTTCGTGCCCGAGGACTACC
ATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTCGGCCTGC
GTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCTT
GCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAGACCGGTGAGGAGCTCAACCCCGGCTTCC
CCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAGCTGCCGCCCGTGGAGGATTTGGAGG
AAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGACTGGGACAGCACTC
AGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGGAGGAGG
TGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCTCGGCGGGGGAGAAAGCAAGCAGCACGGATAC
CATCTCCGCTCCGGGTCGGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCC
CGAACCCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCCTGGCGGGGGCACAA
AAACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCT
CTTCCACCGCGGGGTGAACTTTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCCTA
CTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAAAGACCAGCAGAAAACCAGCAGCTAGAAAATC
CACAGCGGCGGCAGCAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAAACCCGGGAGCTG
AGGAACCGGATCTTTCCCACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGGAGCAGGAACT
GAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAGAGCGAAGACC
AACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCTCACTCTTAAA
GAGTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGTGCCCTTCGCCCT
AGCCGCCTCCACCCATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTGGAGCTACCAGCCC
CAGATGGGCCTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGG
GCCCGCGATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGATACTCCTAGAACAGT
CAGCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACC
AGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGACT
AACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTCGTCACCGCCCCGCTCAGGGTATAAA
GCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTC
TGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTCCTTCACGCCTCGTCAGGCCG
TCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGTGGCATCGGCACTCTCCAGTTCGTGG
AGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGGCCACTACCCGGACGAGT

TCATCCCGAACTTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATTGAAACTAATCACCCCCT
TATCCAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTG
AAATAAAGATACAATCATATTGATGATTTGAGTTTAACAAAAAAATAAAGAATCACTTACTTGAA
ATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGG
TACTGCAGGCCCCGGCGGGCTGCAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTG
TCCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTC
GACCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTCGTC
TCTTCAGATGGATTCCAAGAGAAGCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCAC
CACCAAGAACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCTCGGGAAAA
CTCATCTCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCATTTCCCTT
AACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACCATTAAAT
ATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACTCCGTGGC
TCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATAAAGCTTACC
TTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTGGGCTAAAG
GTTTAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTTGGAAGCAGT
AGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGGCCTTAGCTTT
GACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTGGACAACAC
CTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAACTAACACTTTGCTTGACT
AAATGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAGGAAGTGGAAACCTAAACCC
CATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACGGTGTTCTTTTAAC
AGAACATTCTACACTAAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATGGCACTCCA
TATACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCTTATCCAAAGTCACAAAGTTCTACTACT
AAAAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTATGCTTCTCACTAT
AACCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATACACCTGGACTAA
TGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACATCGCCCAAGAATG
AACACTGTATCCCACCCTGCATGCCAACCCTTCCCACCCCACTCTGTGGAACAAACTCTGAAACAC
AAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGATTCGAGCAGTTATTTTT
CCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCCGCACAGCCTTGAACATCTGAATG
CCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCAGAGCGAGCCAGTCTCGGG
TCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCTGAGG
ATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAATCATA
GTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCCCCGCAGCAGTCGCTGCCGCCGCCGCT
CCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCCACGGCCCTC
AGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATGCGGATCTCGCTCAGGTCGCTGCAGTA
CGTGCAACACAGAACCACCAGGTTGTTCAACAGTCCATAGTTCAACACGCTCCAGCCGAAACTCA
TCGCGGGAAGGATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTAAATCAAGTGGTGCCCC
CTCCAGAACACGCTGCCCACGTACATGATCTCCTTGGGCATGTGGCGGTTCACCACCTCCCGGTAC
CACATCACCCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCGGAACCACAGGGCCAGCACCGC
CCCGCCCGCCATGCAGCGAAGAGACCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCGCTCGT
ACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCATGCAT
CTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCATATGCCAGGGCACGGGGAACTCTTGCAG
GACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAGGGTAT
CGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGT
GGTAAGGGGGCCGGCCGATACGGGTGATGGGGGACGCGGCTGATCGTGTTCGCGACCGTGTCAT
GATGCAGTTGCTTTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCG
ATCGCCGGCGGCGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTC
AGACCGTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGAT
CACATCGACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTTCGGTGA
CGGCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCAAACGGTCTCGGAGTAC
TTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCA
GGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGCGCACA
TCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATCATGTTACACTC
CTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGAACTAGTTCCTGAGGTAAA
TCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCACACCCT
CATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAGATTGACAAGCGGAATATCAAAATCT
CTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCCTCTCCGAAA
TTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAACCGAAGTCC

TCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGGTGATATCTT
CCAGATAACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGAAAAATCCTC
CAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCGTTCCAGCATG
GTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCTGGCGAACAGG
TGGGTAAATCGTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGGCGCGACCCTCGTAAAAAT
TGTCGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGTGAATGATTCGACAA
GATGAATACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAAGCGCCCGAGGAAGCAATAAG
GCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAATTCTCA
GGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGCCCCCGATCCCTCCAGGTA
CACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAG
AGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCCAGATCTACACTGAC
GTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCAGCACACGCCCAGAAAC
CGGTGACACACTCAAAAAAATACGCGCACTTCCTCAAACGCCCAAAACTGCCGTCATTTCCGGGT
TCCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTAAAAACGTCACCCGCCCC
GCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTCAAACACCTCATT
TGCATATTAACGCGCACAAAAAGTTTGAGG Venezuelan equine encephalitis virus [VEE] (SEQ ID NO: 3) GenBank: L01442.2 atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct

cgcaggtacc cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc gcaattatga
tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag tcctccacca taatgaacac ccacagagtg acttttcttc
attcgtcagc aaattgaagg gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt
tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg tgcccaaata tgacataata tttgttaatg
tgaggacccc atataaatac catcactatc agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct
tgtctgcatc tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa gcatcattgg
tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct cacttgaaga gacggaagtt ctgtttgtat tcattgggta
cgatcgcaag gcccgtacgc acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg
aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag gagtgattat aaatgctgct
aacagcaaag gacaacctgg cggagggggtg tgcggagcgc tgtataagaa attcccggaa agcttcgatt tacagccgat
cgaagtagga aaagcgcgac tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca acgataacaa ttacaagtca
gtagcgattc cactgttgtc caccggcatc ttttccggga acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac
accactgatg cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg ctaggagaga
agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg atgcagagct ggtgagggtg catccgaaga
gttctttggc tggaaggaag ggctacagca caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag
gcggccaagg atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca tgtatatcct
cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg aagcctccac accacctagc acgctgcctt
gcttgtgcat ccatgccatg actccagaaa gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct caccgaaagt gcctgcgtat attcatccaa
ggaagtatct cgtggaaaca ccaccggtag acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct
gaacaaccac cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg aagaagagga
tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg aggcagacat tcacgggccg ccctctgtat
ctagctcatc ctggtccatt cctcatgcat ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca
gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc gaccggtgcc tgcgcctcga
acagtattca ggaaccctcc acatcccgct ccgcgcacaa gaacaccgtc acttgcaccc agcagggcct gctcgagaac
cagcctagtt tccacccccgc caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc
ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga ttacaagaga ggagtttgag
gcgttcgtag cacaacaaca atgacggttt gatgcgggtg catacatctt ttcctccgac accggtcaag ggcatttaca
acaaaaatca gtaaggcaaa cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta acagaagcag ataccagtcc
aggaaggtgg agaacatgaa agccataaca gctagacgta ttctgcaagg cctagggcat tatttgaagg cagaaggaaa
agtggagtgc taccgaaccc tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg cagtggaagc
ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta ttattccaga gtacgatgcc tatttggaca tggttgacgg
agcttcatgc tgcttagaca ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac ccacaatacg
atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag ctgccacaaa aagaaattgc aatgtcacgc
aaatgagaga attgcccgta ttggattcgg cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa aaggaccaaa agctgctgct ctttttgcga
agacacataa tttgaatatg ttgcaggaca taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact
ccaggaacaa aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag cgtatctgtg
cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga acattcatac actgtttgat atgtcggctg aagactttga
cgctattata gccgagcact tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg
acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt tgacgctgat tgaggcggct
ttcggcgaaa tttcatcaat acatttgccc actaaaacta aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt
gtgaacacag tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg cagcattcat
tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag acaggtgcgc cacctggttg aatatggaag
tcaagattat agatgctgtg gtgggcgaga aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc
gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg aacatgatga tgacaggaga
agggcattgc atgaagagtc aacacgctgg aaccgagtgg gtattctttc agagctgtgc aaggcagtag aatcaaggta
tgaaaccgta ggaacttcca tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag gggcccctat
aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa gatgttcccg ttccagccaa tgtatccgat gcagccaatg
ccctatcgca acccgttcgc ggccccgcgc aggccctggt tccccagaac cgacccttt ctggcgatgc aggtgcagga
attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg gccatccgct aagaaaccga
agaaggaggc ctcgcaaaaa cagaaagggg gaggccaagg gaagaagaag aagaaccaag ggaagaagaa ggctaagaca
gggccgccta atccgaaggc acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat
gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc ttgtgtggtc ggagggaagt
tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga cgttctggcc gcgcttaaga cgaagaaagc atccaaatac

gatcttgagt atgcagatgt gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta
cagctggcat catggagcag tccaatatga aaatgggcgt ttcacggtgc cgaaaggagt tggggccaag ggagacagcg
gacgacccat tctggataac cagggacggg tggtcgctat tgtgctggga ggtgtgaatg aaggatctag gacagccctt
tcagtcgtca tgtggaacga gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac
catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag aaaaccagca gagactttgg
ccatgctcag cgttaacgtt gacaacccgg gctacgatga gctgctggaa gcagctgtta agtgccccgg aaggaaaagg
agatccaccg aggagctgtt taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag
ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag acttcagact tcctcgcagt
atggcctgga ttcctccggc aacttaaagg gcaggaccat gcggtatgac atgcacggga ccattaaaga gataccacta
catcaagtgt cactccatac atctcgcccg tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgcccggc aggggactcc
atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc gtatgaagtg aaatttaatc ctgtaggcag agaactctat
actcatcccc cagaacacgg agtagagcaa gcgtgccaag tctacgcaca tgatgcacag aacagaggag cttatgtcga
gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt caccgtgaca cctcctgttg
ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa gatctccgag accatcaaca agacaaaaca gttcagccag
tgcacaaaga aggagcagtg cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc
agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaatg caccgtgcct ctagcaccag
aacctatgat aacctttggt ttcagatcag tgtcactgaa actgcaccct aagaatccca catatctaac cacccgccaa cttgctgatg
agcctcacta cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtg ggagtttgta
tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcacccgg aaatccacat gggctaccgc acgaggtgat
aactcattat taccacagat accctatgtc caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac
ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc taggatacca ttttgtctgg ctgtgctttg
ctgcgcccgc actgcccggg ccgagaccac ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga
ttcaattgct gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgcgtgt gctgtgtcgt gccttttta gtcatggccg
gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat gccgagccaa gcgggaatct cgtataacac tatagtcaac
agagcaggct acgcaccact ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt
cacctgccac tacaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga atgcactcca acttacaggc
ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt catgtggggt ggtgcatatt gcttttgcga cactgagaac acccaagtca
gcaaggccta cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc ctcagtgcag
gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta tgtgaatgga gaaactcctg tgaatttcaa tggggtcaaa
ttaactgcag gtccgctttc cacagcttgg acacccttg atcgcaaaat cgtgcagtat gccggggaga tctataatta tgattttcct
gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac agtctcaagc tcagatctgt atgccaatac
caacctagtg ctgcagagac ccaaagcagg agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga
agaaagataa agctccatca ttgaaattta ccgccccttt cggatgcgaa atatatacaa accccattcg cgccgaaaac tgtgctgtag
ggtcaattcc attagccttt gacattcccg acgccttgtt caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta
acgagtgcgt gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa gtgcgcagtc
catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac cgagcaaggg tcggcgacta tccatttctc
gaccgcaaat atccacccgg agttcaggct ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt cacccccccga
aagaccatat tgtgacacac cctcagtatc acgcccaaac atttacagcc gcggtgtcaa aaaccgcgtg gacgtggtta
acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct ggctactatt gtggccatgt acgtgctgac caaccagaaa
cataattgaa tacagcagca attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt ttatttttc
ttttcttttc cgaatcggat tttgttttta atatttc VEE-MAG25 mer (SEQ ID NO: 4); contains MAG-25 merPDTT nucleotide (bases 30-1755)
atgggcggcgcatgagagaagcccagaccaattacctacccaaaatggagaaagttcacgttgacatcgaggaagacagcccattcctcagagct
ttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttcgcatctggcttcaaaactgatcgaaac
ggaggtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtattctaagcacaagtatcattgtatctgtccgatgagatgtgcggaa
gatccggacagattgtataagtatgcaactaagctgaagaaaaactgtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtcatg
agcgaccctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgacgg
accgacaagtctctatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttatgtttaagaacttggctggagcatatccatc
atactctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgtagagggatgtccattcttaga
aagaagtatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggacttactgaggagctggcacctgccgtctgtatttcact
tacgtggcaagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggtacgtcgttaaaagaatagctatcagtccaggcctgtatgggaagccttca
ggctatgctgctacgatgcaccgcgagggattcttgtgctgcaaagtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagcta
cattgtgtgaccaaatgactggcatactggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagtcgtcaacggtcgca
cccagagaaacaccaataccatgaaaaattacctttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatgaaagg
ccactaggactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgcccggatacccaaaccatcatcaa
agtgaacagcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaagg
agccgtcacctctcattaccgccgaggacgtacaagaagctaagtgcgcagccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctac

caccttggcagctgatgttgaggagcccactctggaagccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttgataa
aggttaccagctacgctggcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttgcatccaccctctcgctgaac
aagtcatagtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaatacccgtccaggac
tttcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacacatggaggagcgctgaacactgatg
aagaatattacaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcgacaggaaacagtgcgtcaagaaagaactagtcactgggctagg
gctcacaggcgagctggtggatcctcccttccatgaattcgcctacgagagtctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatgg
cgtgccaggatcaggcaagtctggcatcattaaaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataagggac
gtcaagaaaatgaaagggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagcttttg
cttgtcatgcaggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggttttttaacatgatgtgcctga
aagtgcattttaaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtgacttcggtcgtctcaaccttgttttacgacaaaaaa
atgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtaccaaacctaagcaggacgatctcattctcacttgtttcagaggggggtga
agcagttgcaaatagattacaaaggcaacgaaataatgacggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaatgaaa
atcctctgtacgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaaca
ctgactgccaagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagaccggaccctaccga
cgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatgaccactgaacaatggaacactgtggattat
tttgaaacggacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttctttggactcgatctggactccggtctattttctgcacccactgttccgtt
atccattaggaataatcactgggataactccccgtcgcctaacatgtacgggctgaataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcgg
gcagttgccactggaagagtctatgacatgaacactggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttagt
cctccaccataatgaacacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaaaagttgtccgtcccaggcaaa
atggttgactggttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatatgacataatatttgttaatgtgagg
accccatataaataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagcttgtctgcatctgaatcccggcggaacctgtgtc
agcataggttatggttacgctgacagggccagcgaaagcatcattggtgctatagcgcggcagttcaagttttcccgggtatgcaaaccgaaatcctcacttgaag
agacggaagttctgtttgtattcattgggtacgatcgcaaggcccgtacgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactcca
cgaagccggatgtgcaccctcatatcatgtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaaggacaacctggcg
gaggggtgtgcggagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggtcaaaggtgcagctaaacatatc
attcatgccgtaggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgctaagattgtcaacgataacaatta
caagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcattgaaccatttgctgacagctttagacaccactgatgc
agatgtagccatatactgcagggacaagaaatgggaaatgactctcaaggaagcagtggctaggagagaagcagtggaggagatatgcatatccgacgactctt
cagtgacagaacctgatgcagagctggtgagggtgcatccgaagagttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatatttgg
aagggaccaagtttcaccaggcggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcatgtatatcctcgga
gaaagcatgagcagtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtgcatccatgccatgactccagaa
agagtacagcgcctaaaagcctcacgtccagaacaaattactgtgtgctcatcctttccattgccgaagtatagaatcactggtgtgcagaagatccaatgctccca
gcctatattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacaccaccggtagacgagactccggagccatcggcagagaacca
atccacagaggggacacctgaacaaccaccacttataaccgaggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatag
cataagtttgctgtcagatggcccgacccaccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcctcatgcat
ccgactttgatgtggacagtttatccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtat
ggagtttctggcgcgaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccgtcacttgcacccagcagggcct
gctcgagaaccagcctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgcttaccccgtcacgcactcctagcaggtcggtc
tcgagaaccagcctggtctccaacccgccaggcgtaaatagggtgattacaagagaggagtttgaggcgttcgtagcacaacaacaatgacggtttgatgcggg
tgcatacatcttttcctccgacaccggtcaagggcatttacaacaaaaatcagtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagattt
cgtatgccccgcgcctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagataccagtccaggaaggtg
gagaacatgaaagccataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgctaccgaaccctgcatcctgttcctt
tgtattcatctagtgtgaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttccgactgtggcttcttactgtattattc
cagagtacgatgcctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccctgcaaagctgcgcagctttccaaagaaacactcctattt
ggaacccacaatacgatcggcagtgccttcagcgatccagaacacgctccagaacgtcctggcagctgccacaaaaagaaattgcaatgtcacgcaaatgaga
gaattgcccgtattggattcggcggcctttaatgtggaatgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggcttactg
aagaaaacgtggtaaattacattaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcaggacataccaatggacaggt
ttgtaatggacttaaagagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatccaggctgccgatccgctagcaaca
gcgtatctgtgcggaatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgtttgatatgtcggctgaagactttgacgctatta
tagccgagcacttccagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaaagtgaggacgacgccatggctctgaccgcgttaatgattctggaa
gacttaggtgtggacgcagagctgttgacgctgattgaggcggctttcggcgaaatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatg
aaatctggaatgttcctcacactgtttgtgaacacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgcagcattcatt
ggagatgacaatatcgtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaagattatagatgctgtggtgggc
gagaaagcgccttatttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctgtttaagcttggcaaacct
ctggcagcagacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaaccgagtgggtattctttcagagctgtgcaaggcagt
agaatcaaggtatgaaaccgtaggaacttccatcatagttatggccatgactactctagctagcagtgttaaatcattcagctacctgagaggggcccctataactct
ctacggctaacctgaatggactacgactctagaatagtctttaatTAAGCCACCATGGCAGGCATGTTTCAGGCGCTGAGCGA

AGGCTGCACCCCGTATGATATTAACCAGATGCTGAACGTGCTGGGCGATCATCAGGTCTCAGGCC
TTGAGCAGCTTGAGAGTATAATCAACTTTGAAAAACTGACTGAATGGACCAGTTCTAATGTTATG
CCTATCCTGTCTCCTCTGACAAAGGGCATCCTGGGCTTCGTGTTTACCCTGACCGTGCCTTCTGAG
AGAGGACTTAGCTGCATTAGCGAAGCGGATGCGACCACCCCGGAAAGCGCGAACCTGGGCGAAG
AAATTCTGAGCCAGCTGTATCTTTGGCCAAGGGTGACCTACCATTCCCCTAGTTATGCTTACCACC
AATTTGAAAGACGAGCCAAATATAAAAGACACTTCCCCGGCTTTGGCCAGAGCCTGCTGTTTGGC
TACCCTGTGTACGTGTTCGGCGATTGCGTGCAGGGCGATTGGGATGCGATTCGCTTTCGCTATTGC
GCGCCGCCGGGCTATGCGCTGCTGCGCTGCAACGATACCAACTATAGCGCTCTGCTGGCTGTGGG
GGCCCTAGAAGGACCCAGGAATCAGGACTGGCTTGGTGTCCCAAGACAACTTGTAACTCGGATGC
AGGCTATTCAGAATGCCGGCCTGTGTACCCTGGTGGCCATGCTGGAAGAGACAATCTTCTGGCTG
CAAGCGTTTCTGATGGCGCTGACCGATAGCGGCCCGAAAACCAACATTATTGTGGATAGCCAGTA
TGTGATGGGCATTAGCAAACCGAGCTTTCAGGAATTTGTGGATTGGGAAAACGTGAGCCCGGAAC
TGAACAGCACCGATCAGCCGTTTTGGCAAGCCGGAATCCTGGCCAGAAATCTGGTGCCTATGGTG
GCCACAGTGCAGGGCCAGAACCTGAAGTACCAGGGTCAGTCACTAGTCATCTCTGCTTCTATCATT
GTCTTCAACCTGCTGGAACTGGAAGGTGATTATCGAGATGATGGCAACGTGTGGGTGCATACCCC
GCTGAGCCCGCGCACCCTGAACGCGTGGGTGAAAGCGGTGGAAGAAAAAAAAGGTATTCCAGTT
CACCTAGAGCTGGCCAGTATGACCAACATGGAGCTCATGAGCAGTATTGTGCATCAGCAGGTCAG
AACATACGGCCCCGTGTTCATGTGTCTCGGCGGACTGCTTACAATGGTGGCTGGTGCTGTGTGGCT
GACAGTGCGAGTGCTCGAGCTGTTCCGGGCCGCGCAGCTGGCCAACGACGTGGTCCTCCAGATCA
TGGAGCTTTGTGGTGCAGCGTTTCGCCAGGTGTGCCATACCACCGTGCCGTGGCCGAACGCGAGC
CTGACCCCGAAATGGAACAACGAAACCACCCAGCCCCAGATCGCCAACTGCAGCGTGTATGACTT
TTTTGTGTGGCTCCATTATTATTCTGTTCGAGACACACTTTGGCCAAGGGTGACCTACCATATGAA
CAAATATGCGTATCATATGCTGGAAAGACGAGCCAAATATAAAAGAGGACCAGGACCTGGCGCT
AAATTTGTGGCCGCCTGGACACTGAAAGCCGCTGCTGGTCCTGGACCTGGCCAGTACATCAAGGC
CAACAGCAAGTTCATCGGCATCACCGAACTCGGACCCGGACCAGGCTGATGATTcgaacggccgtatcacg
cccaaacatttacagccgcggtgtcaaaaaccgcgtggacgtggttaacatccctgctgggaggatcagccgtaattattataattggcttggtgctggctactattg
tggccatgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggcaagctgcttacatagaactcgcggcgattggcatgccgccttaaaatttttat
tttatttttcttttcttttccgaatcggatttgtttttaatatttcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
aaaaaaaaaaaaaaa Venezuelan equine encephalitis virus strain TC-83 [TC-83](SEQ ID NO: 5)
GenBank: L01443.1 atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg
ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg aggtagaagc caagcaggtc
actgataatg accatgctaa tgccagagcg ttttcgcatc tggcttcaaa actgatcgaa acggaggtgg acccatccga
cacgatcctt gacattggaa gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat
gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg aaataactga taaggaattg
gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc ctgacctgga aactgagact atgtgcctcc acgacgacga
gtcgtgtcgc tacgaagggc aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag
ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta agaacttggc tggagcatat
ccatcatact ctaccaactg ggccgacgaa accgtgttaa cggctcgtaa cataggccta tgcagctctg acgttatgga
gcggtcacgt agagggatgt ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga
ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact tacgtggcaa gcaaaattac
acatgtcggt gtgagactat agttagttgc gacgggtacg tcgttaaaag aatagctatc agtccaggcc tgtatgggaa
gccttcaggc tatgctgcta cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg
tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac tggcaacaga tgtcagtgcg
gacgacgcgc aaaaactgct ggttgggctc aaccagcgta tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat
gaaaaattac cttttgcccg tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa
ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc acaagataac atctatttat
aagcgcccgg atacccaaac catcatcaaa gtgaacagcg atttccactc attcgtgctg cccaggatag gcagtaacac
attggagatc gggctgagaa caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg
acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt tgcgcgcagc tctaccacct
ttggcagctg atgttgagga gcccactctg gaagccgatg tcgacttgat gttacaagag gctggggccg gctcagtgga
gacacctcgt ggcttgataa aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg
ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga taacacactc tggccgaaaa
gggcgttatg ccgtggaacc ataccatggt aaagtagtgg tgccagaggg acatgcaata cccgtccagg actttcaagc
tctgagtgaa agtgccacca ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag
gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg aatacctgta cgacatcgac
aggaaacagt gcgtcaagaa agaactagtc actgggctag ggctcacagg cgagctggtg gatcctccct tccatgaatt
cgcctacgag agtctgagaa cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag
gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg

agcgccaaga aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg ccagaactgt
ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata ttgacgaagc ttttgcttgt catgcaggta ctctcagagc
gctcatagcc attataagac ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc tgaaagtgca
ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc gttgcactaa atctgtgact tcggtcgtct caaccttgtt
ttacgacaaa aaaatgagaa cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc
aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca aaggcaacga aataatgacg
gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc
cacctcagaa catgtgaacg tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga
taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag cagagcatga tgccatcatg
aggcacatct tggagagacc ggaccctacc gacgtcttcc agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc
ggtgctgaag accgctggca tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact
cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg gtctattttc tgcacccact gttccgttat
ccattaggaa taatcactgg gataactccc cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct
cgcaggtacc cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc gcaattatga
tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag tcctccacca taatgaacac ccacagagtg acttttcttc
attcgtcagc aaattgaagg gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt
tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg tgcccaaata tgacataata tttgttaatg
tgaggacccc atataaatac catcactatc agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct
tgtctgcatc tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa gcatcattgg
tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct cacttgaaga gacggaagtt ctgtttgtat tcattgggta
cgatcgcaag gcccgtacgc acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg
aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag gagtgattat aaatgctgct
aacagcaaag gacaacctgg cggaggggtg tgcggagcgc tgtataagaa attcccggaa agcttcgatt tacagccgat
cgaagtagga aaagcgcgac tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt
cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca acgataacaa ttacaagtca
gtagcgattc cactgttgtc caccggcatc ttttccggga acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac
accactgatg cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg ctaggagaga
agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg atgcagagct ggtgagggtg catccgaaga
gttctttggc tggaaggaag ggctacagca caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag
gcggccaagg atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca tgtatatcct
cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg aagcctccac accacctagc acgctgcctt
gcttgtgcat ccatgccatg actccagaaa gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat
tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct caccgaaagt gcctgcgtat attcatccaa
ggaagtatct cgtggaaaca ccaccggtag acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct
gaacaaccac cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg aagaagagga
tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg aggcagacat tcacgggccg ccctctgtat
ctagctcatc ctggtccatt cctcatgcat ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca
gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc gaccggtgcc tgcgcctcga
acagtattca ggaaccctcc acatcccgct ccgcgcacaa gaacaccgtc acttgcaccc agcagggcct gctcgagaac
cagcctagtt tccacccccgc caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc
ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga ttacaagaga ggagtttgag
gcgttcgtag cacaacaaca atgacggttt gatgcgggtg catacatctt ttcctccgac accggtcaag ggcatttaca
acaaaaatca gtaaggcaaa cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc
tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta acagaagcag ataccagtcc
aggaaggtgg agaacatgaa agccataaca gctagacgta ttctgcaagg cctagggcat tatttgaagg cagaaggaaa
agtggagtgc taccgaaccc tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg cagtggaagc
ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta ttattccaga gtacgatgcc tatttggaca tggttgacgg
agcttcatgc tgcttagaca ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac ccacaatacg
atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag ctgccacaaa aagaaattgc aatgtcacgc
aaatgagaga attgcccgta ttggattcgg cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt
ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa aaggaccaaa agctgctgct ctttttgcga
agacacataa tttgaatatg ttgcaggaca taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact
ccaggaacaa aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag cgtatctgtg
cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga acattcatac actgtttgat atgtcggctg aagactttga
cgctattata gccgagcact tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg
acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt tgacgctgat tgaggcggct
ttcggcgaaa tttcatcaat acatttgccc actaaaacta aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt

gtgaacacag tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg cagcattcat
tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag acaggtgcgc cacctggttg aatatggaag
tcaagattat agatgctgtg gtgggcgaga aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc
gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg aacatgatga tgacaggaga
agggcattgc atgaagagtc aacacgctgg aaccgagtgg gtattctttc agagctgtgc aaggcagtag aatcaaggta
tgaaaccgta ggaacttcca tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag gggcccctat
aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa gatgttcccg ttccagccaa tgtatccgat gcagccaatg
ccctatcgca acccgttcgc ggccccgcgc aggccctggt tccccagaac cgaccctttt ctggcgatgc aggtgcagga
attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg gccatccgct aagaaaccga
agaaggaggc ctcgcaaaaa cagaaagggg gaggccaagg gaagaagaag aagaaccaag ggaagaagaa ggctaagaca
gggccgccta atccgaaggc acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat
gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc ttgtgtggtc ggagggaagt
tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga cgttctggcc gcgcttaaga cgaagaaagc atccaaatac
gatcttgagt atgcagatgt gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta
cagctggcat catggagcag tccaatatga aaatgggcgt ttcacggtgc cgaaaggagt tggggccaag ggagacagcg
gacgacccat tctggataac cagggacggg tggtcgctat tgtgctggga ggtgtgaatg aaggatctag gacagccctt
tcagtcgtca tgtggaacga gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac
catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag aaaaccagca gagactttgg
ccatgctcag cgttaacgtt gacaacccgg gctacgatga gctgctggaa gcagctgtta agtgccccgg aaggaaaagg
agatccaccg aggagctgtt taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag
ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag acttcagact tcctcgcagt
atggcctgga ttcctccggc aacttaaagg gcaggaccat gcggtatgac atgcacggga ccattaaaga gataccacta
catcaagtgt cactccatac atctcgcccg tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgcccggc aggggactcc
atcaccatgg aatttaagaa agattccgtc acacactcct gctaggtgcc gtatgaagtg aaatttaatc ctgtaggcag agaactctat
actcatcccc cagaacacgg agtagagcaa gcgtgccaag tctacgcaca tgatgcacag aacagaggag cttatgtcga
gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt caccgtgaca cctcctgttg
ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa gatctccgag accatcaaca agacaaaaca gttcagccag
tgcacaaaga aggagcagtg cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc
agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaatg caccgtgcct ctagcaccag
aacctatgat aacctttggt ttcagatcag tgtcactgaa actgcaccct aagaatccca catatctaac cacccgccaa cttgctgatg
agcctcacta cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtg ggagtttgta
tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcacccgg aaatccacat gggctaccgc acgaggtgat
aactcattat taccacagat accctatgtc caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac
ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc taggatacca ttttgtctgg ctgtgctttg
ctgcgcccgc actgcccggg ccgagaccac ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga
ttcaattgct gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgcgtgt gctgtgtcgt gccttttta gtcatggccg
gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat gccgagccaa gcgggaatct cgtataacac tatagtcaac
agagcaggct acgcaccact ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt
cacctgccac tacaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga atgcactcca acttacaggc
ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt catgtggggt ggtgcatatt gcttttgcga cactgagaac acccaagtca
gcaaggccta cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc ctcagtgcag
gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta tgtgaatgga gaaactcctg tgaatttcaa tggggtcaaa
ttaactgcag gtccgctttc cacagcttgg acacccttg atcgcaaaat cgtgcagtat gccggggaga tctataatta tgattttcct
gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac agtctcaagc tcagatctgt atgccaatac
caacctagtg ctgcagagac ccaaagcagg agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga
agaaagataa agctccatca ttgaaattta ccgccccttt cggatgcgaa atatatacaa accccattcg cgccgaaaac tgtgctgtag
ggtcaattcc attagccttt gacattcccg acgccttgtt caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta
acgagtgcgt gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa gtgcgcagtc
catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac cgagcaaggg tcggcgacta tccatttctc
gaccgcaaat atccacccgg agttcaggct ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt caccccccga
aagaccatat tgtgacacac cctcagtatc acgcccaaac atttacagcc gcggtgtcaa aaaccgcgtg gacgtggtta
acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct ggctactatt gtggccatgt acgtgctgac caaccagaaa
cataattgaa tacagcagca attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt ttatttttc
ttttcttttc cgaatcggat tttgttttta atatttc VEE Delivery Vector (SEQ ID NO: 6); VEE genome with nucleotides 7544-11175 deleted [alphavirus structural proteins removed]
ATGggcggcgcatgagagaagcccagaccaattacctacccaaaATGGagaaagttcacgttgacatcgaggaagacagcccattcctcagagctttgca
gcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttcgcatctggcttcaaaactgatcgaaacggag

gtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtattctaagcacaagtatcattgtatctgtccgatgagatgtgcggaagatcc
ggacagattgtataagtatgcaactaagctgaagaaaaactgtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtcatgagcg
accctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgacggaccga
caagtctctatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttatgtttaagaacttggctggagcatatccatcatactct
accaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgtagagggatgtccattcttagaaagaag
tatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggacttactgaggagctggcacctgccgtctgtatttcacttacgtgg
caagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggtacgtcgttaaaagaatagctatcagtccaggcctgtatgggaagccttcaggctatg
ctgctacgatgcaccgcgagggattcttgtgctgcaaagtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagctacattgtgt
gaccaaatgactggcatactggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagtcgtcaacggtcgcacccaga
gaaacaccaataccatgaaaaattacctttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatgaaaggccactag
gactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgcccggatacccaaaccatcatcaaagtgaac
agcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaaggagccgtc
acctctcattaccgccgaggacgtacaagaagctaagtgcgcagccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctaccaccttg
gcagctgatgttgaggagcccactctggaagccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttgataaaggttacc
agctacgctggcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttgcatccaccctctcgctgaacaagtcata
gtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaatacccgtccaggactttcaagct
ctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacacatggaggagcgctgaacactgatgaagaatatt
acaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcgacaggaaacagtgcgtcaagaaagaactagtcactgggctagggctcacag
gcgagctggtggatcctcccttccatgaattcgcctacgagagtctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatggcgtgccag
gatcaggcaagtctggcatcattaaaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaa
atgaaaggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagcttttgcttgtcatgc
aggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggtttttttaacatgatgtgcctgaaagtgcatttt
aaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtgacttcggtcgtctcaaccttgttttacgacaaaaaaatgagaacg
acgaatccgaaagagactaagattgtgattgacactaccggcagtaccaaacctaagcaggacgatctcattctcacttgtttcagaggggggtgaagcagttgc
aaatagattacaaaggcaacgaaataatgacggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaatgaaaatcctctgta
cgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaacactgactgcc
aagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagaccggaccctaccgacgtcttcca
gaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatgaccactgaacaatggaacactgtggattattttgaaacg
gacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttctttggactcgatctggactccggtctattttctgcacccactgttccgttatccattag
gaataatcactgggataactccccgtcgcctaacatgtacgggctgaataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcgggcagttgcc
actggaagagtctatgacatgaacactggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttagtcctccaccat
aatgaacacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaaaagttgtccgtcccaggcaaaatggttgactg
gttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatatgacataatatttgttaatgtgaggaccccatataa
ataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagcttgtctgcatctgaatcccggcggaacctgtgtcagcataggtta
tggttacgctgacagggccagcgaaagcatcattggtgctatagcgcggcagttcaagttttcccgggtatgcaaaccgaaatcctcacttgaagagacggaagtt
ctgtttgtattcattgggtacgatcgcaaggcccgtacgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactccacgaagccggat
gtgcaccctcatatcatgtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaaggacaacctggcggaggggtgtgc
ggagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggtcaaaggtgcagctaaacatatcattcatgccgta
ggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgctaagattgtcaacgataacaattacaagtcagtag
cgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcattgaaccatttgctgacagctttagacaccactgatgcagatgtagccat
atactgcagggacaagaaatgggaaatgactctcaaggaagcagtggctaggagagaagcagtggaggagatatgcatatccgacgactcttcagtgacagaa
cctgatgcagagctggtgagggtgcatccgaagagttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatatttggaagggaccaag
tttcaccaggcggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcatgtatatcctcggagaaagcatgagc
agtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtgcatccatgccatgactccagaaagagtacagcgc
ctaaaagcctcacgtccagaacaaattactgtgtgctcatccctttccattgccgaagtatagaatcactggtgtgcagaagatccaatgctcccagcctatattgttctc
accgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacaccaccggtagacgagactccggagccatcggcagagaaccaatccacagagggg
acacctgaacaaccaccacttataaccgaggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagtttgctgtc
agatggcccgacccaccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcctcatgcatccgactttgatgtgg
acagtttatccatacttgacaccctggaggggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtatggagtttctggcgc
gaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccgtcacttgcacccagcagggcctgctcgagaaccag
cctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgcttaccccgtcacgcactcctagcaggtcggtctcgagaaccagcc
tggtctccaacccgccaggcgtaaatagggtgattacaagagaggagtttgaggcgttcgtagcacaacaacaatgacggtttgatgcgggtgcatacatcttttcc
tccgacaccggtcaagggcatttacaacaaaaatcagtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagatttcgtatgccccgcgc
ctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagataccagtccaggaaggtggagaacatgaaagc
cataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgctaccgaaccctgcatcctgttccttgtattcatctagtgt
gaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttccgactgtggcttcttactgtattattccagagtacgatgc
ctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccctgcaaagctgcgcagctttccaaagaaacactcctatttggaacccacaata

cgatcggcagtgccttcagcgatccagaacacgctccagaacgtcctggcagctgccacaaaaagaaattgcaatgtcacgcaaatgagagaattgcccgtattg
gattcggcggcctttaatgtggaatgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggcttactgaagaaaacgtggta
aattacattaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcaggacataccaatggacaggtttgtaatggacttaaa
gagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatccaggctgccgatccgctagcaacagcgtatctgtgcgg
aatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgtttgatatgtcggctgaagactttgacgctattatagccgagcacttc
cagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaaagtgaggacgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtgga
cgcagagctgttgacgctgattgaggcggctttcggcgaaatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatgaaatctggaatgttc
ctcacactgtttgtgaacacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgcagcattcattggagatgacaatatc
gtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaagattatagatgctgtggtgggcgagaaagcgccttat
ttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctgtttaagcttggcaaacctctggcagcagacga
tgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaaccgagtgggtattctttcagagctgtgcaaggcagtagaatcaaggtatga
aaccgtaggaacttccatcatagttatggccatgactactctagctagcagtgttaaatcattcagctacctgagaggggcccctataactctctacggcTAAcct
gaatggactacgactatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtggacgtggttaacatccctgctgggaggatcagccgtaattattataatt
ggcttggtgctggctactattgtggccatgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggcaagctgcttacatagaactcgcggcgattg
gcatgccgccttaaaatttttatttattttttctttctttccgaatcggattttgtttttaatatttcAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA TC-83 Delivery Vector(SEQ ID NO: 7); TC-83 genome with nucleotides 7544-11175 deleted [alphavirus structural proteins removed]
ATAGGCGGCGCATGAGAGAAGCCCAGACCAATTACCTACCCAAAATGGAGAAAGTTCACGTTGA
CATCGAGGAAGACAGCCCATTCCTCAGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAG
CCAAGCAGGTCACTGATAATGACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCTTCAAAACTG
ATCGAAACGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCCGCAGAAT
GTATTCTAAGCACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGT
ATAAGTATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAA
AATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACTGAGACTATGTGCCTCCACG
ACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGATGTATACGCGGTTGACGGA
CCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCTACTGGATAGGCTTTGACAC
CACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATACTCTACCAACTGGGCCGACGA
AACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCTGACGTTATGGAGCGGTCACGTAGAG
GGATGTCCATTCTTAGAAAGAAGTATTTGAAACCATCCAACAATGTTCTATTCTCTGTTGGCTCGA
CCATCTACCACGAGAAGAGGGACTTACTGAGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGT
GGCAAGCAAAATTACACATGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAG
AATAGCTATCAGTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGG
GATTCTTGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACG
TATGTGCCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGA
CGCGCAAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAACA
CCAATACCATGAAAAATTACCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCAAAGGAA
TATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTAGTCATGGGGT
GTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCGGATACCCAAACCATC
ATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATAGGCAGTAACACATTGGAGAT
CGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCACAAGGAGCCGTCACCTCTCATTACC
GCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGATGAGGCTAAGGAGGTGCGTGAAGCCGAGG
AGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTGAGGAGCCCACTCTGGAAGCCGATGTC
GACTTGATGTTACAAGAGGCTGGGGCCGGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTAC
CAGCTACGATGGCGAGGACAAGATCGGCTCTTACGCTGTGCTTTCTCCGCAGGCTGTACTCAAGA
GTGAAAAATTATCTTGCATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGA
AAAGGGCGTTATGCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATAC
CCGTCCAGGACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTA
AACAGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTACAA
AACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTACGACATCGACAGGAAACAGTGCGTC
AAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCCTTCCATGAATT
CGCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCAACCATAGGGGTGTATG
GCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTCACCAAAAAAGATCTAGTGGTG
AGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGACGTCAAGAAAATGAAAGGGCTGGACG
TCAATGCCAGAACTGTGGACTCAGTGCTCTTGAATGGATGCAAACACCCCGTAGAGACCCTGTAT
ATTGACGAAGCTTTTGCTTGTCATGCAGGTACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAA
AAGGCAGTGCTCTGCGGGGATCCCAAACAGTGCGGTTTTTTTAACATGATGTGCCTGAAAGTGCA

TTTTAACCACGAGATTTGCACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGT
GACTTCGGTCGTCTCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCCGAAAGAGACTA
AGATTGTGATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCA
GAGGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCCTC
TCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCTCTGTACG
CACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATCGTGTGGAAAACA
CTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGGAATTTCACTGCCACGAT
AGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGGAGAGACCGGACCCTACC
GACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAGGCTTTAGTGCCGGTGCTGAAGACCGC
TGGCATAGACATGACCACTGAACAATGGAACACTGTGGATTATTTTGAAACGGACAAAGCTCACT
CAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTAT
TTTCTGCACCCACTGTTCCGTTATCCATTAGGAATAATCACTGGGATAACTCCCCGTCGCCTAACA
TGTACGGGCTGAATAAAGAAGTGGTCCGTCAGCTCTCTCGCAGGTACCCACAACTGCCTCGGGCA
GTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAA
CCTAGTACCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGA
GTGACTTTTCTTCATTCGTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGGAAAAGTTGT
CCGTCCCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTTCAGAGCTCGGCTG
GATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAATGTGAGGACCCCATAT
AAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATGTTGACCAAGAAAGC
TTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATAGGTTATGGTTACGCTGACAGGGCCA
GCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTCAAGTTTTCCCGGGTATGCAAACCGAAATCC
TCACTTGAAGAGACGGAAGTTCTGTTTGTATTCATTGGGTACGATCGCAAGGCCCGTACGCACAA
TCCTTACAAGCTTTCATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATG
TGCACCCTCATATCATGTGGTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATG
CTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGA
AAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACAT
ATCATTCATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGGTGACAAACAGTTGGC
AGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGATTCCAC
TGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAACCATTTGCTGA
CAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAGAAATGGGAAATGACT
CTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCTTCAG
TGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAGAGTTCTTTGGCTGGAAGGAAGGG
CTACAGCACAAGCGATGGCAAAACTTTCTCATATTTGGAAGGGACCAAGTTTCACCAGGCGGCCA
AGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATG
TATATCCTCGGAGAAAGCATGAGCAGTATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTC
CACACCACCTAGCACGCTGCCTTGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCC
TAAAAGCCTCACGTCCAGAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCA
CTGGTGTGCAGAAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTC
ATCCAAGGAAGTATCTCGTGGAAACACCACCGGTAGACGAGACTCCGGAGCCATCGGCAGAGAA
CCAATCCACAGAGGGGACACCTGAACAACCACCACTTATAACCGAGGATGAGACCAGGACTAGA
ACGCCTGAGCCGATCATCATCGAAGAGGAAGAAGAGGATAGCATAAGTTTGCTGTCAGATGGCCC
GACCCACCAGGTGCTGCAAGTCGAGGCAGACATTCACGGGCCGCCCTCTGTATCTAGCTCATCCT
GGTCCATTCCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCCTGGAGGGAG
CTAGCGTGACCAGCGGGGCAACGTCAGCCGAGACTAACTCTTACTTCGCAAAGAGTATGGAGTTT
CTGGCGCGACCGGTGCCTGCGCCTCGAACAGTATTCAGGAACCCTCCACATCCCGCTCCGCGCAC
AAGAACACCGTCACTTGCACCCAGCAGGGCCTGCTCGAGAACCAGCCTAGTTTCCACCCCGCCAG
GCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCACGCACTCCTAGCAGG
TCGGTCTCGAGAACCAGCCTGGTCTCCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGA
GTTTGAGGCGTTCGTAGCACAACAACAATGACGGTTTGATGCGGGTGCATACATCTTTTCCTCCGA
CACCGGTCAAGGGCATTTACAACAAAAATCAGTAAGGCAAACGGTGCTATCCGAAGTGGTGTTGG
AGAGGACCGAATTGGAGATTTCGTATGCCCCGCGCCTCGACCAAGAAAAAGAAGAATTACTACGC
AAGAAATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGA
ACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAGGA
AAAGTGGAGTGCTACCGAACCCTGCATCCTGTTCCTTTGTATTCATCTAGTGTGAACCGTGCCTTT
TCAAGCCCCAAGGTCGCAGTGGAAGCCTGTAACGCCATGTTGAAAGAGAACTTTCCGACTGTGGC
TTCTTACTGTATTATTCCAGAGTACGATGCCTATTTGGACATGGTTGACGGAGCTTCATGCTGCTT
AGACACTGCCAGTTTTTGCCCTGCAAAGCTGCGCAGCTTTCCAAAGAAACACTCCTATTTGGAACC

CACAATACGATCGGCAGTGCCTTCAGCGATCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCA
CAAAAAGAAATTGCAATGTCACGCAAATGAGAGAATTGCCCGTATTGGATTCGGCGGCCTTTAAT
GTGGAATGCTTCAAGAAATATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCAT
CAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAATTAAAAGGACCAAAAGCTGCTGCTC
TTTTTGCGAAGACACATAATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAATGGACT
TAAAGAGAGACGTGAAAGTGACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGGT
GATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGA
GATTAAATGCGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCGGCTGAAGACTTTGACG
CTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAAACTGACATCGCGTCGTTTGATA
AAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTCTGGAAGACTTAGGTGTGGACGCA
GAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAATTTCATCAATACATTTGCCCACTAAAACT
AAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCCTCACACTGTTTGTGAACACAGTCATT
AACATTGTAATCGCAAGCAGAGTGTTGAGAGAACGGCTAACCGGATCACCATGTGCAGCATTCAT
TGGAGATGACAATATCGTGAAAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACC
TGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGCCTTATTTCTGTGG
AGGGTTTATTTTGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCAGACCCCCTAAAAAGGC
TGTTTAAGCTTGGCAAACCTCTGGCAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTG
CATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGGCAGTAGAATC
AAGGTATGAAACCGTAGGAACTTCCATCATAGTTATGGCCATGACTACTCTAGCTAGCAGTGTTA
AATCATTCAGCTACCTGAGAGGGGCCCCTATAACTCTCTACGGCTAACCTGAATGGACTACGACT
ATCACGCCCAAACATTTACAGCCGCGGTGTCAAAAACCGCGTGGACGTGGTTAACATCCCTGCTG
GGAGGATCAGCCGTAATTATTATAATTGGCTTGGTGCTGGCTACTATTGTGGCCATGTACGTGCTG
ACCAACCAGAAACATAATTGAATACAGCAGCAATTGGCAAGCTGCTTACATAGAACTCGCGGCGA
TTGGCATGCCGCCTTAAAATTTTTATTTTATTTTTCTTTTCTTTTCCGAATCGGATTTTGTTTTTAAT
ATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAA VEE Production Vector (SEQ ID NO: 8); VEE genome with nucleotides 7544-11175 deleted, plus 5′ T7- promoter, plus 3′ restriction sites
TAATACGACTCACTATAGGATGggcggcgcatgagagaagcccagaccaattacctacccaaaATGGagaaagttcacgttgacatcg
aggaagacagcccattcctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttc
gcatctggcttcaaaactgatcgaaacggaggtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtattctaagcacaagtatcat
tgtatctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaaaaactgtaaggaaataactgataaggaattggacaaga
aaatgaaggagctcgccgccgtcatgagcgaccctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtt
taccaggatgtatacgcggttgacggaccgacaagtctctatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttatgttt
aagaacttggctggagcatatccatcatactctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggt
cacgtagagggatgtccattcttagaaagaagtatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggacttactgagga
gctggcacctgccgtctgtatttcacttacgtggcaagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggtacgtcgttaaaagaatagctatca
gtccaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattcttgtgctgcaaagtgacagacacattgaacggggagagggtctctttt
tcccgtgtgcacgtatgtgccagctacattgtgtgaccaaatgactggcatactggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaa
ccagcgtatagtcgtcaacggtcgcacccagagaaacaccaataccatgaaaaattacctttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatat
aaggaagatcaagaagatgaaaggccactaggactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagc
gcccggataccccaaaccatcatcaaagtgaacagcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctgagaacaagaatca
ggaaaatgttagaggagcacaaggagccgtcacctctcattaccgccgaggacgtacaagaagctaagtgcgcagccgatgaggctaaggaggtgcgtgaag
ccgaggagttgcgcgcagctctaccacctttggcagctgatgttgaggagcccactctggaagccgatgtcgacttgatgttacaagaggctggggccggctca
gtggagacacctcgtggcttgataaaggttaccagctacgctggcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaatt
atcttgcatccaccctctcgctgaacaagtcatagtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagag
ggacatgcaatacccgtccaggactttcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacac
atggaggagcgctgaacactgatgaagaatattacaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcgacaggaaacagtgcgtcaa
gaaagaactagtcactgggctagggctcacaggcgagctggtggatcctcccttccatgaattcgcctacgagagtctgagaacacgaccagccgctccttacca
agtaccaaccataggggtgtatggcgtgccaggatcaggcaagtctggcatcattaaaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaa
aactgtgcagaaattataagggacgtcaagaaaatgaaagggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtaga
gaccctgtatattgacgaagcttttgcttgtcatgcaggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagt
gcggttttttaacatgatgtgcctgaaagtgcattttaaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtgacttcggtcg
tctcaaccttgttttacgacaaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtaccaaacctaagcaggacgatctcatt
ctcacttgtttcagagggtgggtgaagcagttgcaaatagattacaaaggcaacgaaataatgacggcagctgcctctcaagggctgacccgtaaaggtgtgtatg
ccgttcggtacaaggtgaatgaaaatcctctgtacgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactag
ccggcgacccatggataaaaacactgactgccaagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatc
ttggagagaccggaccctaccgacgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatgaccact

gaacaatggaacactgtggattattttgaaacggacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttctttggactcgatctggactccggt
ctattttctgcacccactgttccgttatccattaggaataatcactgggataactccccgtcgcctaacatgtacgggctgaataaagaagtggtccgtcagctctctc
gcaggtacccacaactgcctcgggcagttgccactggaagagtctatgacatgaacactggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaa
cagaagactgcctcatgctttagtcctccaccataatgaacacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcgggga
aaagttgtccgtcccaggcaaaatggttgactggttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatat
gacataatatttgttaatgtgaggaccccatataaataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagcttgtctgcatct
gaatcccggcggaacctgtgtcagcataggttatggttacgctgacagggccagcgaaagcatcattggtgctatagcgcggcagttcaagttttcccgggtatgc
aaaccgaaatcctcacttgaagagacggaagttctgtttgtattcattgggtacgatcgcaaggcccgtacgcacaatccttacaagctttcatcaaccttgaccaac
atttatacaggttccagactccacgaagccggatgtgcaccctcatatcatgtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgcta
acagcaaaggacaacctggcggaggggtgtgcggagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggt
caaaggtgcagctaaacatatcattcatgccgtaggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgct
aagattgtcaacgataacaattacaagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcattgaaccatttgctga
cagctttagacaccactgatgcagatgtagccatatactgcagggacaagaaatgggaaatgactctcaaggaagcagtggctaggagagaagcagtggagga
gatatgcatatccgacgactcttcagtgacagaacctgatgcagagctggtgagggtgcatccgaagagttctttggctggaaggaagggctacagcacaagcg
atggcaaaactttctcatatttggaagggaccaagtttcaccaggcggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagc
aggtatgcatgtatatcctcggagaaagcatgagcagtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtg
catccatgccatgactccagaaagagtacagcgcctaaaagcctcacgtccagaacaaattactgtgtgctcatcctttccattgccgaagtatagaatcactggtgt
gcagaagatccaatgctcccagcctatattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacaccaccggtagacgagactccg
gagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccacttataaccgaggatgagaccaggactagaacgcctgagccgatcatcatc
gaagaggaagaagaggatagcataagtttgctgtcagatggcccgacccaccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctc
atcctggtccattcctcatgcatccgactttgatgtggacagtttatccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagact
aactcttacttcgcaaagagtatggagtttctggcgcgaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccg
tcacttgcacccagcagggcctgctcgagaaccagcctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgcttaccccgtc
acgcactcctagcaggtcggtctcgagaaccagcctggtctccaacccgccaggcgtaaatagggtgattacaagagaggagtttgaggcgttcgtagcacaac
aacaatgacggtttgatgcgggtgcatacatcttttcctccgacaccggtcaagggcatttacaacaaaaatcagtaaggcaaacggtgctatccgaagtggtgttg
gagaggaccgaattggagatttcgtatgccccgcgcctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagc
agataccagtccaggaaggtggagaacatgaaagccataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgcta
ccgaaccctgcatcctgttcctttgtattcatctagtgtgaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttccg
actgtggcttcttactgtattattccagagtacgatgcctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccctgcaaagctgcgcag
ctttccaaagaaacactcctatttggaacccacaatacgatcggcagtgccttcagcgatccagaacacgctccagaacgtcctggcagctgccacaaaaagaaa
ttgcaatgtcacgcaaatgagagaattgcccgtattggattcggcggcctttaatgtggaatgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaa
gaaaaccccatcaggcttactgaagaaaacgtggtaaattacattaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgc
aggacataccaatggacaggtttgtaatggacttaaagagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatcca
ggctgccgatccgctagcaacagcgtatctgtgcggaatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgtttgatatgtc
ggctgaagactttgacgctattatagccgagcacttccagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaaagtgaggacgacgccatggctc
tgaccgcgttaatgattctggaagacttaggtgtggacgcagagctgttgacgctgattgaggcggctttcggcgaaatttcatcaatacatttgcccactaaaacta
aatttaaattcggagccatgatgaaatctggaatgttcctcacactgtttgtgaacacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccg
gatcaccatgtgcagcattcattggagatgacaatatcgtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaa
gattatagatgctgtggtgggcgagaaagcgccttatttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaa
ggctgtttaagcttggcaaacctctggcagcagacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaaccgagtgggtatt
ctttcagagctgtgcaaggcagtagaatcaaggtatgaaaccgtaggaacttccatcatagttatggccatgactactctagctagcagtgttaaatcattcagctacc
tgagaggggcccctataactctctacggcTAAcctgaatggactacgactatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtggacgtggtta
acatccctgctgggaggatcagccgtaattattataattggcttggtgctggctactattgtggccatgtacgtgctgaccaaccagaaacataattgaatacagcag
caattggcaagctgcttacatagaactcgcggcgattggcatgccgccttaaaatttttattttatttttttcttttcttttccgaatcggattttgtttttaatatttcAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAtacgtagtttaaac TC-83 Production Vector(SEQ ID NO: 9); TC-83 genome with nucleotides 7544-11175 deleted, plus 5′ T7-promoter, plus 3′ restriction sites
TAATACGACTCACTATAGGATAGGCGGCGCATGAGAGAAGCCCAGACCAATTACCTACCCAAAAT
GGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTCAGAGCTTTGCAGCGGAGCTTCC
CGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAATGACCATGCTAATGCCAGAGCGTTTTCG
CATCTGGCTTCAAAACTGATCGAAACGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAG
TGCGCCCGCCCGCAGAATGTATTCTAAGCACAAGTATCATTGTATCTGTCCGATGAGATGTGCGG
AAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGA
TAAGGAATTGGACAAGAAAATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAACT
GAGACTATGTGCCTCCACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACCAGGA
TGTATACGCGGTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCT
ACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATATCCATCATACT

CTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCTGACGTT
ATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCATCCAACAATGT
TCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTGAGGAGCTGGCACCTGCC
GTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGGTGTGAGACTATAGTTAGTTGCGA
CGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTG
CTACGATGCACCGCGAGGGATTCTTGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTC
TCTTTTCCCGTGTGCACGTATGTGCCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACA
GATGTCAGTGCGGACGACGCGCAAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGG
TCGCACCCAGAGAAACACCAATACCATGAAAAATTACCTTTTGCCCGTAGTGGCCCAGGCATTTG
CTAGGTGGGCAAAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAG
ACAGTTAGTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCC
GGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATAGGCA
GTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCACAAGGAGCC
GTCACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGATGAGGCTAAGGAG
GTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTGAGGAGCCCAC
TCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCCGGCTCAGTGGAGACACCTCGTG
GCTTGATAAAGGTTACCAGCTACGATGGCGAGGACAAGATCGGCTCTTACGCTGTGCTTTCTCCGC
AGGCTGTACTCAAGAGTGAAAAATTATCTTGCATCCACCCTCTCGCTGAACAAGTCATAGTGATA
ACACACTCTGGCCGAAAAGGGCGTTATGCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGA
GGGACATGCAATACCCGTCCAGGACTTTCAAGCTCTGAGTGAAAGTGCCACCATTGTGTACAACG
AACGTGAGTTCGTAAACAGGTACCTGCACCATATTGCCACACATGGAGGAGCGCTGAACACTGAT
GAAGAATATTACAAAACTGTCAAGCCCAGCGAGCACGACGGCGAATACCTGTACGACATCGACA
GGAAACAGTGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCC
TCCCTTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCAA
CCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTCACCAAA
AAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGTGCAGAAATTATAAGGGACGTCAAGAAAA
TGAAAGGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAATGGATGCAAACACCCC
GTAGAGACCCTGTATATTGACGAAGCTTTTGCTTGTCATGCAGGTACTCTCAGAGCGCTCATAGCC
ATTATAAGACCTAAAAAGGCAGTGCTCTGCGGGGATCCCAAACAGTGCGGTTTTTTTAACATGAT
GTGCCTGAAAGTGCATTTTAACCACGAGATTTGCACACAAGTCTTCCACAAAAGCATCTCTCGCCG
TTGCACTAAATCTGTGACTTCGGTCGTCTCAACCTTGTTTTACGACAAAAAAATGAGAACGACGA
ATCCGAAAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTC
ATTCTCACTTGTTTCAGAGGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAAT
GACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATG
AAAATCCTCTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGC
ATCGTGTGGAAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGGA
ATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGGAG
AGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAGGCTTTAGTGCC
GGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACTGTGGATTATTTTGAAA
CGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGACTCGAT
CTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTATCCATTAGGAATAATCACTGGGATAAC
TCCCCGTCGCCTAACATGTACGGGCTGAATAAAGAAGTGGTCCGTCAGCTCTCTCGCAGGTACCC
ACAACTGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTGGTACACTGCGCAATT
ATGATCCGCGCATAAACCTAGTACCTGTAAACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATA
ATGAACACCCACAGAGTGACTTTTCTTCATTCGTCAGCAAATTGAAGGGCAGAACTGTCCTGGTG
GTCGGGGAAAAGTTGTCCGTCCCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTAC
CTTCAGAGCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAA
TGTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCA
TGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATAGGTTATGGTT
ACGCTGACAGGGCCAGCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTCAAGTTTTCCCGGGTA
TGCAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTATTCATTGGGTACGATCGCAA
GGCCCGTACGCACAATCCTTACAAGCTTTCATCAACCTTGACCAACATTTATACAGGTTCCAGACT
CCACGAAGCCGGATGTGCACCCTCATATCATGTGGTGCGAGGGGATATTGCCACGGCCACCGAAG
GAGTGATTATAAATGCTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTA
TAAGAAATTCCCGGAAAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAA
GGTGCAGCTAAACATATCATTCATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGG
TGACAAACAGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGT

CAGTAGCGATTCCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCAT
TGAACCATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAG
AAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATAT
CCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAGAGTTCTTTG
GCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTGGAAGGGACCAAGTT
TCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATG
AGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGTATTAGGTCGAAATGCCCCGTCGAA
GAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCTTGCTTGTGCATCCATGCCATGACTCCAGA
AAGAGTACAGCGCCTAAAAGCCTCACGTCCAGAACAAATTACTGTGTGCTCATCCTTTCCATTGCC
GAAGTATAGAATCACTGGTGTGCAGAAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAG
TGCCTGCGTATATTCATCCAAGGAAGTATCTCGTGGAAACACCACCGGTAGACGAGACTCCGGAG
CCATCGGCAGAGAACCAATCCACAGAGGGGACACCTGAACAACCACCACTTATAACCGAGGATG
AGACCAGGACTAGAACGCCTGAGCCGATCATCATCGAAGAGGAAGAAGAGGATAGCATAAGTTT
GCTGTCAGATGGCCCGACCCACCAGGTGCTGCAAGTCGAGGCAGACATTCACGGGCCGCCCTCTG
TATCTAGCTCATCCTGGTCCATTCCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGA
CACCCTGGAGGGAGCTAGCGTGACCAGCGGGGCAACGTCAGCCGAGACTAACTCTTACTTCGCAA
AGAGTATGGAGTTTCTGGCGCGACCGGTGCCTGCGCCTCGAACAGTATTCAGGAACCCTCCACAT
CCCGCTCCGCGCACAAGAACACCGTCACTTGCACCCAGCAGGGCCTGCTCGAGAACCAGCCTAGT
TTCCACCCCGCCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCAC
GCACTCCTAGCAGGTCGGTCTCGAGAACCAGCCTGGTCTCCAACCCGCCAGGCGTAAATAGGGTG
ATTACAAGAGAGGAGTTTGAGGCGTTCGTAGCACAACAACAATGACGGTTTGATGCGGGTGCATA
CATCTTTTCCTCCGACACCGGTCAAGGGCATTTACAACAAAAATCAGTAAGGCAAACGGTGCTAT
CCGAAGTGGTGTTGGAGAGGACCGAATTGGAGATTTCGTATGCCCCGCGCCTCGACCAAGAAAAA
GAAGAATTACTACGCAAGAAATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATACCAGTC
CAGGAAGGTGGAGAACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATT
TGAAGGCAGAAGGAAAAGTGGAGTGCTACCGAACCCTGCATCCTGTTCCTTTGTATTCATCTAGT
GTGAACCGTGCCTTTTCAAGCCCCAAGGTCGCAGTGGAAGCCTGTAACGCCATGTTGAAAGAGAA
CTTTCCGACTGTGGCTTCTTACTGTATTATTCCAGAGTACGATGCCTATTTGGACATGGTTGACGG
AGCTTCATGCTGCTTAGACACTGCCAGTTTTTGCCCTGCAAAGCTGCGCAGCTTTCCAAAGAAACA
CTCCTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAGCGATCCAGAACACGCTCCAGAACG
TCCTGGCAGCTGCCACAAAAAGAAATTGCAATGTCACGCAAATGAGAGAATTGCCCGTATTGGAT
TCGGCGGCCTTTAATGTGGAATGCTTCAAGAAATATGCGTGTAATAATGAATATTGGGAAACGTT
TAAAGAAAACCCCATCAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAATTAAAAGGAC
CAAAAGCTGCTGCTCTTTTTGCGAAGACACATAATTTGAATATGTTGCAGGACATACCAATGGAC
AGGTTTGTAATGGACTTAAAGAGAGACGTGAAAGTGACTCCAGGAACAAAACATACTGAAGAAC
GGCCCAAGGTACAGGTGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCAC
CGAGAGCTGGTTAGGAGATTAAATGCGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCG
GCTGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAAACTGAC
ATCGCGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTCTGGAAGA
CTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAATTTCATCAATAC
ATTTGCCCACTAAAACTAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCCTCACACTGT
TTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGTTGAGAGAACGGCTAACCGGATCA
CCATGTGCAGCATTCATTGGAGATGACAATATCGTGAAAGGAGTCAAATCGGACAAATTAATGGC
AGACAGGTGCGCCACCTGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGGCGAGAAA
GCGCCTTATTTCTGTGGAGGGTTTATTTTGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCA
GACCCCCTAAAAAGGCTGTTTAAGCTTGGCAAACCTCTGGCAGCAGACGATGAACATGATGATGA
CAGGAGAAGGGCATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTCAGAGCTGT
GCAAGGCAGTAGAATCAAGGTATGAAACCGTAGGAACTTCCATCATAGTTATGGCCATGACTACT
CTAGCTAGCAGTGTTAAATCATTCAGCTACCTGAGAGGGGCCCCTATAACTCTCTACGGCTAACCT
GAATGGACTACGACTATCACGCCCAAACATTTACAGCCGCGGTGTCAAAAACCGCGTGGACGTGG
TTAACATCCCTGCTGGGAGGATCAGCCGTAATTATTATAATTGGCTTGGTGCTGGCTACTATTGTG
GCCATGTACGTGCTGACCAACCAGAAACATAATTGAATACAGCAGCAATTGGCAAGCTGCTTACA
TAGAACTCGCGGCGATTGGCATGCCGCCTTAAAATTTTTATTTTATTTTTCTTTTCTTTTCCGAATC
GGATTTTGTTTTTAATATTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAtacgtagtttaaac VEE-UbAAY (SEQ ID NO: 14); VEE delivery vector with MHC class I mouse tumor epitopes SIINFEKL (SEQ ID NO: 57)_and AH1-A5 inserted

ATGggcggcgcatgagagaagcccagaccaattacctacccaaaatggagaaagttcacgttgacatcgaggaagacagcccattcctcagag
ctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttcgcatctggcttcaaaactgatcgaa
acggaggtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtattctaagcacaagtatcattgtatctgtccgatgagatgtgcgg
aagatccggacagattgtataagtatgcaactaagctgaagaaaaactgtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtc
atgagcgaccctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgac
ggaccgacaagtctctatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttatgtttaagaacttggctggagcatatcca
tcatactctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgtagagggatgtccattcttag
aaagaagtatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggacttactgaggagctggcacctgccgtctgtatttcac
ttacgtggcaagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggtacgtcgttaaaagaatagctatcagtccaggcctgtatgggaagccttc
aggctatgctgctacgatgcaccgcgagggattcttgtgctgcaaagtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagct
acattgtgtgaccaaatgactggcatactggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagtcgtcaacggtcgc
acccagagaaacaccaataccatgaaaaattacctttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatgaaag
gccactaggactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgcccggatacccaaaccatcatca
aagtgaacagcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaag
gagccgtcacctctcattaccgccgaggacgtacaagaagctaagtgcgcagccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctcta
ccacctttggcagctgatgttgaggagcccactctggaagccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttgata
aaggttaccagctacgctggcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttgcatccaccctctcgctgaa
caagtcatagtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaatacccgtccagga
ctttcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacacatggaggagcgctgaacactgat
gaagaatattacaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcgacaggaaacagtgcgtcaagaaagaactagtcactgggctag
ggctcacaggcgagctggtggatcctcccttccatgaattcgcctacgagagtctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatg
gcgtgccaggatcaggcaagtctggcatcattaaaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataaggga
cgtcaagaaaatgaaaggggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagctttt
gcttgtcatgcaggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggtttttttaacatgatgtgcctg
aaagtgcattttaaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtgacttcggtcgtctcaaccttgttttacgacaaaaa
aatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtaccaaacctaagcaggacgatctcattctcacttgtttcagagggtgggtg
aagcagttgcaaatagattacaaaggcaacgaaataatgacggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaatgaa
aatcctctgtacgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaac
actgactgccaagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagaccggaccctaccg
acgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatgaccactgaacaatggaacactgtggatt
attttgaaacggacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttctttggactcgatctggactccggtctatttttctgcacccactgttccg
ttatccattaggaataatcactgggataactccccgtcgcctaacatgtacgggctgaataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcg
ggcagttgccactggaagagtctatgacatgaacactggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttag
tcctccaccataatgaacacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaaaagttgtccgtcccaggcaaa
atggttgactggttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatatgacataatatttgttaatgtgagg
accccatataaataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagcttgtctgcatctgaatcccggcggaacctgtgtc
agcataggttatggttacgctgacagggccagcgaaagcatcattggtgctatagcgcggcagttcaagttttcccgggtatgcaaaccgaaatcctcacttgaag
agacggaagttctgtttgtattcattgggtacgatcgcaaggcccgtacgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactcca
cgaagccggatgtgcaccctcatatcatgtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaaggacaacctggcg
gaggggtgtgcggagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggtcaaaggtgcagctaaacatatc
attcatgccgtaggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgctaagattgtcaacgataacaatta
caagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcattgaaccatttgctgacagctttagacaccactgatgc
agatgtagccatatactgcagggacaagaaatgggaaatgactctcaaggaagcagtggctaggagagaagcagtggaggagatatgcatatccgacgactctt
cagtgacagaacctgatgcagagctggtgaggggtgcatccgaagagttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatatttgg
aagggaccaagtttcaccaggcggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcatgtatatcctcgga
gaaagcatgagcagtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtgcatccatgccatgactccagaa
agagtacagcgcctaaaagcctcacgtccagaacaaattactgtgtgctcatcctttccattgccgaagtatagaatcactggtgtgcagaagatccaatgctccca
gcctatattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacaccaccggtagacgagactccggagccatcggcagagaacca
atccacagaggggacacctgaacaaccaccacttataaccgaggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatag
cataagtttgctgtcagatggcccgacccaccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcctcatgcat
ccgactttgatgtggacagtttatccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtat
ggagtttctggcgcgaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccgtcacttgcacccagcagggcct
gctcgagaaccagcctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgcttaccccgtcacgcactcctagcaggtcggtc
tcgagaaccagcctggtctccaacccgccaggcgtaaatagggtgattacaagagaggagtttgaggcgttcgtagcacaacaacaatgacggtttgatgcggg
tgcatacatcttttcctccgacaccggtcaagggcatttacaacaaaaatcagtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagattt
cgtatgccccgcgcctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagataccagtccaggaaggtg
gagaacatgaaagccataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgctaccgaaccctgcatcctgttcctt

tgtattcatctagtgtgaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttccgactgtggcttcttactgtattattc
cagagtacgatgcctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccctgcaaagctgcgcagctttccaaagaaacactcctattt
ggaacccacaatacgatcggcagtgccttcagcgatccagaacacgctccagaacgtcctggcagctgccacaaaaagaaattgcaatgtcacgcaaatgaga
gaattgcccgtattggattcggcggcctttaatgtggaatgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggcttactg
aagaaaacgtggtaaattacattaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcaggacataccaatggacaggt
ttgtaatggacttaaagagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatccaggctgccgatccgctagcaaca
gcgtatctgtgcggaatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgtttgatatgtcggctgaagactttgacgctatta
tagccgagcacttccagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaaagtgaggacgacgccatggctctgaccgcgttaatgattctggaa
gacttaggtgtggacgcagagctgttgacgctgattgaggcggctttcggcgaaatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatg
aaatctggaatgttcctcacactgtttgtgaacacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgcagcattcatt
ggagatgacaatatcgtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaagattatagatgctgtggtgggc
gagaaagcgccttatttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctgtttaagcttggcaaacct
ctggcagcagacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaaccgagtgggtattctttcagagctgtgcaaggcagt
agaatcaaggtatgaaaccgtaggaacttccatcatagttatggccatgactactctagctagcagtgttaaatcattcagctacctgagaggggcccctataactct
ctacggctaacctgaatggactacgactctagaatagtctttaattaaagtccgccatatgaggccaccatgCAGATCTTCGTGAAGACCCTGA
CCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATCGAGAACGTGAAGGCCAAGAT
CCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAAGCAGCTGGAA
GATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCTG
AGAGGTGGCGCTGCTTACAGTATAATCAACTTTGAAAAACTGGCTGCTTACGGCATCCTGGGCTTT
GTGTTTACACTGGCTGCCTACCTGCTGTTTGGCTATCCTGTGTACGTGGCCGCTTATGGACTGTGTA
CCCTGGTGGCCATGCTGGCTGCTTACAATCTGGTGCCTATGGTGGCCACAGTGGCCGCCTATTGTC
TTGGCGGACTGCTGACAATGGTGGCAGCCTACAgcccgagctatgcgtatcatcagtttGCAGCCTACGGCCCAG
GACCAGGCgCTAAATTTGTGGCTGCCTGGACACTGAAAGCCGCCGCTGGACCAGGTCCTGGACAG
TACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTCGGCCCAGGACCAGGCTATCCCTA
CGATGTGCCTGATTACGCCTGATagTGATGATTCGAACGGCCGtatcacgcccaaacatttacagccgcggtgtcaaaaac
cgcgtggacgtggttaacatccctgctgggaggatcagccgtaattattataattggcttggtgctggctactattgtggccatgtacgtgctgaccaaccagaaaca
taattgaatacagcagcaattggcaagctgcttacatagaactcgcggcgattggcatgccgccttaaaatttttattttatttttcttttcttttccgaatcggattttgtttt
taatatttcAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAA VEE-Luciferase (SEQ ID NO: 15); VEE delivery vector with luciferase gene inserted at 7545
ATGggcggcgcatgagagaagcccagaccaattacctacccaaaATGGagaaagttcacgttgacatcgaggaagacagcccattcctcagagctttgca
gcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttcgcatctggcttcaaaactgatcgaaacggag
gtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtattctaagcacaagtatcattgtatctgtccgatgagatgtgcggaagatcc
ggacagattgtataagtatgcaactaagctgaagaaaaactgtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtcatgagcg
accctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgacggaccga
caagtctctatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccaccccttttatgtttaagaacttggctggagcatatccatcatactct
accaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgtagagggatgtccattcttagaaagaag
tatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggacttactgaggagctggcacctgccgtctgtatttcacttacgtgg
caagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggtacgtcgttaaaagaatagctatcagtccaggcctgtatgggaagccttcaggctatg
ctgctacgatgcaccgcgagggattcttgtgctgcaaagtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagctacattgtgt
gaccaaatgactggcatactggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagtcgtcaacggtcgcacccaga
gaaacaccaataccatgaaaaattaccttttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatgaaaggccactag
gactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgcccggatacccaaaccatcatcaaagtgaac
agcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaaggagccgtc
acctctcattaccgccgaggacgtacaagaagctaagtgcgcagccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctaccaccttttg
gcagctgatgttgaggagcccactctggaagccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttgataaaggttacc
agctacgctggcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttgcatccaccctctcgctgaacaagtcata
gtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaatacccgtccaggactttcaagct
ctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacacatggaggagcgctgaacactgatgaagaatatt
acaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcgacaggaaacagtgcgtcaagaaagaactagtcactgggctagggctcacag
gcgagctggtggatcctcccttccatgaattcgcctacgagagtctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatggcgtgccag
gatcaggcaagtctggcatcattaaaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaa
atgaaagggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagcttttgcttgtcatgc
aggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggtttttttaacatgatgtgcctgaaagtgcatttt
aaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtgacttcggtcgtctcaaccttgttttacgacaaaaaaatgagaacg
acgaatccgaaagagactaagattgtgattgacactaccggcagtaccaaacctaagcaggacgatctcattctcacttgtttcagagggtgggtgaagcagttgc
aaatagattacaaaggcaacgaaataatgacggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaatgaaaatcctctgta

cgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaacactgactgcc
aagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagaccggaccctaccgacgtcttcca
gaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatgaccactgaacaatggaacactgtggattattttgaaacg
gacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttctttggactcgatctggactccggtctattttctgcacccactgttccgttatccattag
gaataatcactgggataactccccgtcgcctaacatgtacgggctgaataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcgggcagttgcc
actggaagagtctatgacatgaacactggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttagtcctccaccat
aatgaacacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaaaagttgtccgtcccaggcaaaatggttgactg
gttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatatgacataatatttgttaatgtgaggaccccatataa
ataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagcttgtctgcatctgaatcccggcggaacctgtgtcagcataggtta
tggttacgctgacagggccagcgaaagcatcattggtgctatagcgcggcagttcaagttttcccgggtatgcaaaccgaaatcctcacttgaagagacggaagtt
ctgtttgtattcattgggtacgatcgcaaggcccgtacgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactccacgaagccggat
gtgcaccctcatatcatgtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaaggacaacctggcggaggggtgtgc
ggagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggtcaaaggtgcagctaaacatatcattcatgccgta
ggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgctaagattgtcaacgataacaattacaagtcagtag
cgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcattgaaccatttgctgacagctttagacaccactgatgcagatgtagccat
atactgcagggacaagaaatgggaaatgactctcaaggaagcagtggctaggagagaagcagtggaggagatatgcatatccgacgactcttcagtgacagaa
cctgatgcagagctggtgagggtgcatccgaagagttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatatttggaagggaccaag
tttcaccaggcggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcatgtatatcctcggagaaagcatgagc
agtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtgcatccatgccatgactccagaaagagtacagcgc
ctaaaagcctcacgtccagaacaaattactgtgtgctcatccttccattgccgaagtatagaatcactggtgtgcagaagatccaatgctcccagcctatattgttctc
accgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacaccaccggtagacgagactccggagccatcggcagagaaccaatccacagagggg
acacctgaacaaccaccacttataaccgaggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagtttgctgtc
agatggcccgacccaccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcctcatgcatccgactttgatgtgg
acagtttatccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtatggagtttctggcgc
gaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccgtcacttgcacccagcagggcctgctcgagaaccag
cctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgcttaccccgtcacgcactcctagcaggtcggtctcgagaaccagcc
tggtctccaacccgccaggcgtaaatagggtgattacaagagaggagtttgaggcgttcgtagcacaacaacaatgacggtttgatgcgggtgcatacatcttttcc
tccgacaccggtcaagggcatttacaacaaaaatcagtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagatttcgtatgccccgcgc
ctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagataccagtccaggaaggtggagaacatgaaagc
cataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgctaccgaaccctgcatcctgttccttgtattcatctagtgt
gaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttccgactgtggcttcttactgtattattccagagtacgatgc
ctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccctgcaaagctgcgcagctttccaaagaaacactcctatttggaacccacaata
cgatcggcagtgccttcagcgatccagaacacgctccagaacgtcctggcagctgccacaaaaagaaattgcaatgtcacgcaaatgagagaattgcccgtattg
gattcggcggcctttaatgtggaatgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggcttactgaagaaaacgtggta
aattacattaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcaggacataccaatggacaggtttgtaatggacttaaa
gagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatccaggctgccgatccgctagcaacagcgtatctgtgcgg
aatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgtttgatatgtcggctgaagactttgacgctattatagccgagcacttc
cagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaaagtgaggacgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtgga
cgcagagctgttgacgctgattgaggcggctttcggcgaaatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatgaaatctggaatgttc
ctcacactgtttgtgaacacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgcagcattcattggagatgacaatatc
gtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaagattatagatgctgtggtgggcgagaaagcgccttat
ttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctgtttaagcttggcaaacctctggcagcagacga
tgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaaccgagtgggtattctttcagagctgtgcaaggcagtagaatcaaggtatga
aaccgtaggaacttccatcatagttatggccatgactactctagctagcagtgttaaatcattcagctacctgagaggggcccctataactctctacggcTAAcct
gaatggactacgactctagaatagtctttaattaaagtccgccatatgagatggaagatgccaaaaacattaagaagggcccagcgccattctacccactcgaaga
cgggaccgccggcgagcagctgcacaaagccatgaagcgctacgccctggtgcccggcaccatcgcctttaccgacgcacatatcgaggtggacattacctac
gccgagtacttcgagatgagcgttcggctggcagaagctatgaagcgctatgggctgaatacaaaccatcggatcgtggtgtgcagcgagaatagcttgcagttc
ttcatgcccgtgttgggtgccctgttcatcggtgtggctgtggccccagctaacgacatctacaacgagcgcgagctgctgaacagcatgggcatcagccagccc
accgtcgtattcgtgagcaagaaagggctgcaaaagatcctcaacgtgcaaaagaagctaccgatcatacaaaagatcatcatcatggatagcaagaccgactac
cagggcttccaaagcatgtacaccttcgtgacttcccatttgccacccggcttcaacgagtacgacttcgtgcccgagagcttcgaccgggacaaaaccatcgccc
tgatcatgaacagtagtggcagtaccggattgcccaagggcgtagccctaccgcaccgcaccgcttgtgtccgattcagtcatgcccgcgaccccatcttcggca
accagatcatccccgacaccgctatcctcagcgtggtgccatttcaccacggcttcggcatgttcaccacgctgggctacttgatctgcggctttcgggtcgtgctc
atgtaccgcttcgaggaggagctattcttgcgcagcttgcaagactataagattcaatctgccctgctggtgcccacactatttagcttcttcgctaagagcactctcat
cgacaagtacgacctaagcaacttgcacgagatcgccagcgggggcgccgctcagcaaggaggtaggtgaggccgtggccaaacgcttccacctaccag
gcatccgccagggctacggcctgacagaaacaaccagcgccattctgatcacccccgaaggggacgacaagcctggcgcagtaggcaaggtggtgcccttct
tcgaggctaaggtggtggacttggacaccggtaagacactgggtgtgaaccagcgcggcgagctgtgcgtccgtggccccatgatcatgagcggctacgttaa
caaccccgaggctacaaacgctctcatcgacaaggacggctggctgcacagcggcgacatcgcctactgggacgaggacgagcacttcttcatcgtggaccg

gctgaagagcctgatcaaatacaagggctaccaggtagccccagccgaactggagagcatcctgctgcaacaccccaacatcttcgacgccggggtcgccgg cctgcccgacgacgatgccggcgagctgcccgccgcagtcgtcgtgctggaacacggtaaaaccatgaccgagaaggagatcgtggactatgtggccagcca ggttacaaccgccaagaagctgcgcggtggtgttgtgttcgtggacgaggtgcctaaaggactgaccggcaagttggacgcccgcaagatccgcgagattctca ttaaggccaagaagggggcaagatcgccgtgtaaTTCGAACGGCCGtatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtggacgt ggttaacatccctgctgggaggatcagccgtaattattataattggcttggtgctggctactattgtggccatgtacgtgctgaccaaccagaaacataattgaataca gcagcaattggcaagctgcttacatagaactcgcggcgattggcatgccgccttaaaatttttattttattttttcttttcttttccgaatcggattttgtttttaatatttcAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAA ubiquitin (SEQ ID NO: 38) >UbG76 0-228 ATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACC ATCGAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATC TTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCC TGCACCTGGTCCTTCGCCTGAGAGGTGGC Ubiquitin A76 (SEQ ID NO: 39) >UbA76 0-228 ATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACC ATCGAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATC TTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCC TGCACCTGGTCCTTCGCCTGAGAGGTGCC HLA-A2 (MHC class I) signal peptide (SEQ ID NO: 40) >MHC SignalPep 0-78 atggccgtcatggcgccccgaaccctcgtcctgctactctcgggggctctggccctgacccagacctgggggctct HLA-A2 (MHC class I) Trans Membrane domain (SEQ ID NO: 41) >HLA A2 TM Domain 0-201 CCGtcttcccagcccaccatccCCATCGTGGGCAtcattgctggcctggttctctttggagctgtgatcactggagctgtggtcgctgctgtgatgtggagg aggaagagctcagatagaaaaggagggagctactctcaggctgcaagcagtgacagtgcccagggctctgatgtgtctctcacagcttgtaaagtgtga IgK Leader Seq (SEQ ID NO: 42) >IgK Leader Seq 0-60 atggagaccgatacactgctgctgtgggtgctgctcctgtgggtgccaggaagcacaggc Human DC-Lamp (SEQ ID NO: 43) >HumanDCLAMP 0-3178 ggcaccgattcggggcctgcccggacttcgccgcacgctgcagaacctcgcccagcgcccaccatgccccggcagctcagcgcggcggccgcgctcttcgcgtc cctggccgtaattttgcacgatggcagtcaaatgagagcaaaagcatttccagaaaccagagattattctcaacctactgcagcagcaacagtacaggacataaaaaa acctgtccagcaaccagctaagcaagcacctcaccaaactttagcagcaagattcatggatggtcatatcaccctttcaaacagcggccacagtaaaaattccaacaact accccagcaactacaaaaaacactgcaaccaccagcccaattacctacaccctggtcacaacccaggccacacccaacaactcacacacagctcctccagttactga agttacagtcggccctagcttagccccttattcactgccacccaccatcaccccaccagctcatacagctggaaccagttcatcaaccgtcagccacacaactgggaa caccactcaacccagtaaccagaccacccttccagcaactttatcgatagcactgcacaaaagcacaaccggtcagaagcctgatcaacccacccatgccccaggaa caacggcagctgcccacaataccacccgcacagctgcacctgcctccacggttcctgggcccaccccttgcacctcagccatcgtcagtcaagactggaatttatcag gttctaaacggaagcagactctgtataaaagcagagatggggatacagctgattgttcaagacaaggagtcggtttttcacctcggagatacttcaacatcgaccccaa cgcaacgcaagcctctgggaactgtggcacccgaaaatccaaccttctgttgaattttcagggcggatttgtgaatctcacatttaccaaggatgaagaatcatattatat cagtgaagtgggagcctatttgaccgtctcagatccagagacagtttaccaaggaatcaaacatgcggtggtgatgttccagacagcagtcgggcattccttcaagtgc gtgagtgaacagagcctccagttgtcagcccacctgcaggtgaaaacaaccgatgtccaacttcaagcctttgattttgaagatgaccactttggaaatgtggatgagtg ctcgtctgactacacaattgtgcttcctgtgattggggccatcgtggttggtctctgccttatgggtatgggtgtctataaaatccgcctaaggtgtcaatcatctggatacc agagaatctaattgttgcccgggggaatgaaaataatggaatttagagaactctttcatcccttccaggatggatgttgggaaattccctcagagtgtgggtccttcaaa caatgtaaaccaccatcttctattcaaatgaagtgagtcatgtgtgatttaagttcaggcagcacatcaatttctaaatactttttgtttattttatgaaagatatagtgagctgttt attttctagtttcctttagaatattttagccactcaaagtcaacatttgagatatgttgaattaacataatatatgtaaagtagaataagccttcaaattataaaccaagggtcaat tgtaactaatactactgtgtgtgcattgaagattttattttacccttgatcttaacaaagcctttgctttgttatcaaatggactttcagtgcttttactatctgtgttttatggtttcat gtaacatacatattcctggtgtagcacttaactccttttccactttaaatttgtttttgtttttgagacggagtttcactcttgtcacccaggctggagtacagtggcacgatctc ggcttatggcaacctccgcctcccgggttcaagtgattctcctgcttcagcttcccgagtagctgggattacaggcacacactaccacgcctggctaatttttgtatttttatt atagacgggtttcaccatgttggccagactggtcttgaactcttgacctcaggtgatccacccacctcagcctcccaaagtgctgggattacaggcatgagccattgcg cccggccttaaatgtttttttaatcatcaaaaagaacaacatatctcaggttgtctaagtgtttttatgtaaaaccaacaaaaagaacaaatcagcttatatttttatcttgatg actcctgctccagaattgctagactaagaattaggtggctacagatggtagaactaaacaataagcaagagacaataataatggcccttaattattaacaaagtgccaga gtctaggctaagcactttatctatatctcatttcattctcacaacttataagtgaatgagtaaactgagacttaagggaactgaatcacttaaatgtcacctggctaactgatg gcagagccagagcttgaattcatgttggtctgacatcaaggtctttggtcttctccctacaccaagttacctacaagaacaatgacaccacactctgcctgaaggctcaca cctcataccagcatacgctcaccttacagggaaatgggtttatccaggatcatgagacattagggtagatgaaaggagagctttgcagataacaaaatagcctatcctta ataaatcctccactctctggaaggagactgaggggctttgtaaaacattagtcagttgctcatttttatgggattgcttagctgggctgtaaagatgaaggcatcaaataaa ctcaaagtatttttaaatttttttgataatagagaaacttcgctaaccaactgttctttcttgagtgtatagccccatcttgtggtaacttgctgcttctgcacttcatatccatattt cctattgttcactttattctgtagagcagcctgccaagaattttatttctgctgttttttttgctgctaaagaaaggaactaagtcaggatgttaacagaaaagtccacataacc ctagaattcttagtcaaggaataattcaagtcagcctagagaccatgttgactttcctcatgtgtttccttatgactcagtaagttggcaaggtcctgactttagtcttaataaa acattgaattgtagtaaaggtttttgcaataaaaacttactttgg Mouse LAMP1 (SEQ ID NO: 44) >MouseLamp1 0-1858 attccggaggtgaaaaacaatggcacaacgtgtataatggccagcttctctgcctccttctgaccacctacgagactgcgaatggttctcagatcgtgaacatttccctg ccagcctctgcagaagtactgaaaaatggcagttcttgtggtaaagaaaatgtttctgaccccagcctcacaattacttttggaagaggatatttactgacactcaacttca caaaaaatacaacacgttacagtgtccagcatatgtattttacatataacttgtcagatacagaacattttcccaatgccatcagcaaagagatctacaccatggattccac aactgacatcaaggcagacatcaacaaagcataccggtgtgtcagtgatatccgggtctacatgaagaatgtgaccgttgtgctccgggatgccactatccaggccta

cctgtcgagtggcaacttcagcaaggaagagacacactgcacacaggatggaccttccccaaccactgggccacccagcccctcaccaccacttgtgcccacaaac
cccactgtatccaagtacaatgttactggtaacaacggaacctgcctgctggcctctatggcactgcaactgaatatcacctacctgaaaaaggacaacaagacggtga
ccagagcgttcaacatcagcccaaatgacacatctagtgggagttgcggtatcaacttggtgaccctgaaagtggagaacaagaacagagccctggaattgcagtttg
ggatgaatgccagctctagcctgtttttcttgcaaggagtgcgcttgaatatgactcttcctgatgccctagtgcccacattcagcatctccaaccattcactgaaagctctt
caggccactgtgggaaactcatacaagtgcaacactgaggaacacatctttgtcagcaagatgctctccctcaatgtcttcagtgtgcaggtccaggctttcaaggtgg
acagtgacaggtttgggtctgtggaagagtgtgttcaggatggtaacaacatgttgatccccattgctgtgggcggtgccctggcagggctgatcctcatcgtcctcatt
gcctacctcattggcaggaagaggagtcacgccggctatcagaccatctagcctggtgggcaggtgcaccagagatgcacaggggcctgttctcacatccccaagc
ttagataggtgtggaagggaggcacactttctggcaaactgttttaaaatctgctttatcaaatgtgaagttcatcttgcaacatttactatgcacaaaggaataactattgaa
atgacggtgttaattttgctaactgggttaaatattgatgagaaggctccactgatttgacttttaagacttggtgtttggttcttcattcttttactcagatttaagcctatcaaag
ggatactctggtccagaccttggcctggcaagggtggctgatggttaggctgcacacacttaagaagcaacgggagcagggaaggcttgcacacaggcacgcaca
gggtcaacctctggacacttggcttgggctacctggccttgggggggctgaactctggcatctggctgggtacacacccccccaatttctgtgctctgccacccgtgag
ctgccactttcctaaatagaaaatggcattatttttatttacttttttgtaaagtgatttccagtcttgtgttggcgttcagggtggccctgtctctgcactgtgtacaataatagat
tcacactgctgacgtgtcttgcagcgtaggtgggttgtacactgggcatcagctcacgtaatgcattgcctgtaacgatgctaataaaaa Human Lamp1
cDNA (SEQ ID NO: 45) >Human Lamp1 0-2339
ggcccaaccgccgcccgcgccccgctctccgcaccgtacccggccgcctcgcgccatggcggcccccggcagcgcccggcgacccctgctgctgctactgct
gttgctgctgctcggcctcatgcattgtgcgtcagcagcaatgtttatggtgaaaaatggcaacgggaccgcgtgcataatggccaacttctctgctgccttctcagtgaa
ctacgacaccaagagtggccctaagaacatgacctttgacctgccatcagatgccacagtggtgctcaaccgcagctcctgtggaaaagagaacacttctgacccca
gtctcgtgattgcttttggaagaggacatacactcactctcaatttcacgagaaatgcaacacgttacagcgtccagctcatgagttttgtttataacttgtcagacacacac
cttttccccaatgcgagctccaaagaaatcaagactgtggaatctataactgacatcagggcagatatagataaaaaatacagatgtgttagtggcacccaggtccacat
gaacaacgtgaccgtaacgctccatgatgccaccatccaggcgtacctttccaacagcagcttcagcaggggagagacacgctgtgaacaagacaggccttcccca
accacagcgccccctgcgccacccagcccctcgccctcacccgtgcccaagagcccctctgtggacaagtacaacgtgagcggcaccaacgggacctgcctgct
ggccagcatggggctgcagctgaacctcacctatgagaggaaggacaacacgacggtgacaaggcttctcaacatcaaccccaacaagacctcggccagcgggga
gctgcggcgcccacctggtgactctggagctgcacagcgagggcaccaccgtcctgctcttccagttcgggatgaatgcaagttctagccggtttttcctacaaggaat
ccagttgaatacaattcttcctgacgccagagaccctgcctttaaagctgccaacggctccctgcgagcgctgcaggccacagtcggcaattcctacaagtgcaacgc
ggaggagcacgtccgtgtcacgaaggcgttttcagtcaatatattcaaagtgtgggtccaggctttcaaggtggaaggtggccagtttggctctgtggaggagtgtctg
ctggacgagaacagcatgctgatcccatcgctgtgggtggtgccctggcggggctggtcctcatcgtcctcatcgcctacctcgtcggcaggaagaggagtcacgc
aggctaccagactatctagcctggtgcacgcaggcacagcagctgcaggggcctctgttcctttctctgggcttagggtcctgtcgaaggggaggcacactttctggc
aaacgtttctcaaatctgcttcatccaatgtgaagttcatcttgcagcatttactatgcacaacagagtaactatcgaaatgacggtgttaattttgctaactgggttaaatattt
tgctaactggttaaacattaatatttaccaaagtaggattttgaggggggggtgctctctctgaggggtgggggtgccgctgtctctgaggggggggtgccgctgt
ctctgaggggtgggggtgccgctctctctgaggggtgggggtgccgctttctctgaggggtgggggtgccgctctctctgaggggtgggggtgctgctctctcc
gaggggtggaatgccgctgtctctgaggggtgggggtgccgctctaaattggctccatatcatttgagtttagggttctggtgtttggtttcttcattctttactgcactcaga
tttaagccttacaaagggaaagcctctggccgtcacacgtaggacgcatgaaggtcactcgtggtgaggctgacatgctcacacattacaacagtagagagggaaaa
tcctaagacagaggaactccagagatgagtgtctggagcgcttcagttcagctttaaaggccaggacgggccacacgtggctggcggcctcgttccagtggcggca
cgtccttgggcgtctctaatgtctgcagctcaagggctggcacttttttaaatataaaaatgggtgttatttttatttttttttgtaaagtgatttttggtcttctgttgacattcggg
gtgatcctgttctgcgctgtgtacaatgtgagatcggtgcgttctcctgatgttttgccgtggcttggggattgtacacgggaccagctcacgtaatgcattgcctgtaaca
atgtaataaaaagcctctttctttttaaaaaaaaaaaaaaaaaaaaaaaa Tetanus toxoid nulceic acid sequence (SEQ ID
NO: 46) CAGTACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTC Tetanus toxoid
amino acid sequence (SEQ ID NO: 47) QYIKANSKFIGITEL PADRE nulceotide sequence
(SEQ ID NO: 48) GCTAAATTTGTGGCTGCCTGGACACTGAAAGCCGCCGCT PADRE amino
acid sequence (SEQ ID NO: 49) AKFVAAWTLKAAA WPRE (SEQ ID NO: 50) >WPRE 0-
593
aatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatcatgctattgcttccc
gtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaac
ccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttcccctccctattgccacggcggaactcatcgccgcctgccttgcccgctgct
ggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaagctgacgtcctttccatggctgctcgcctgtgttgccacctggattctgcgcgggac
gtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgcggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtc
ggatctccctttgggccgcctccccgcctgt IRES (SEQ ID NO: 51) >eGFP_IRES_SEAP_Insert 1746-2335
tctccccccccccctctccctcccccccccctaacgttactggccgaagccgcttggaataaggccggtgtgcgtttgtctatatgttattttccaccatattgccgtctttt
ggcaatgtgagggcccggaaacctggccctgtcttcttgacgagcattcctaggggtctttcccctctgccaaaggaatgcaaggtctgttgaatgtcgtgaaggaag
cagttcctctggaagcttcttgaagacaaacaacgtctgtagcgacccttgcaggcagcggaacccccccacctggcgacaggtgcctctgcggccaaaagccacgt
gtataagatacacctgcaaaggcggcacaaccccagtgccacgttgtgagttggatagttgtggaaagagtcaaatggctctcctcaagcgtattcaacaaggggctg
aaggatgcccagaaggtaccccattgtatgggatctgatctggggcctcggtgcacatgctttacatgtgtttagtcgaggttaaaaaaacgtctaggccccccgaacc
acggggacgtggttttcctttgaaaaacacgatgataatatg GFP (SEQ ID NO: 52)
atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcga
gggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcag
tgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaa
ctacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcaca

agctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggca
gcgtgcagctcgccgaccactaccagcagaacaccccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaa
gaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtag SEAP
(SEQ ID NO: 53)
atgctgctgctgctgctgctgggcctgaggctacagctctccctgggcatcatcccagttgaggaggagaacccggacttctggaaccgcgaggcagccgagg
ccctgggtgccgccaagaagctgcagcctgcacagacagccgccaagaacctcatcatcttcctgggcgatgggatgggggtgtctacggtgacagctgccaggat
cctaaaagggcagaagaaggacaaactggggcctgagataccctggccatggaccgcttcccatatgtggctctgtccaagacatacaatgtagacaaacatgtgc
cagacagtggagccacagccacggcctacctgtgcggggtcaagggcaacttccagaccattggcttgagtgcagccgcccgctttaaccagtgcaacacgacac
gcggcaacgaggtcatctccgtgatgaatcgggccaagaaagcagggaagtcagtgggagtggtaaccaccacacgagtgcagcacgcctcgccagccggcac
ctacgcccacacggtgaaccgcaactggtactcggacgccgacgtgcctgcctcggcccgccaggaggggtgccaggacatcgctacgcagctcatctccaacat
ggacattgacgtgatcctaggtggaggccgaaagtacatgtttcgcatgggaaccccagaccctgagtacccagatgactacagccaaggtgggaccaggctggac
gggaagaatctggtgcaggaatggctggcgaagcgccagggtgcccggtatgtgtggaaccgcactgagctcatgcaggcttccctggacccgtctgtgacccatc
tcatgggtctctttgagcctggagacatgaaatacgagatccaccgagactccacactggacccctccctgatggagatgacagaggctgccctgcgcctgctgagc
aggaacccccgcggcttcttcctcttcgtggagggtggtcgcatcgaccatggtcatcatgaaagcagggcttaccgggcactgactgagacgatcatgttcgacgac
gccattgagaggggggccagctcaccagcgaggaggacacgctgagcctcgtcactgccgaccactcccacgtcttctccttcggaggctacccccctgcgaggg
agctccatcttcgggctggcccctggcaaggcccgggacaggaaggcctacacggtcctcctatacggaaacggtccaggctatgtgctcaaggacggcgcccgg
ccggatgttaccgagagcgagagcgggagccccgagtatcggcagcagtcagcagtgcccctggacgaagagacccacgcaggcgaggacgtggcggtgttcg
cgcgcggcccgcaggcgcacctggttcacggcgtgcaggagcagaccttcatagcgcacgtcatggccttcgccgcctgcctggagccctacaccgcctgcgacc
tggcgccccccgccggcaccaccgacgccgcgcacccgggttactctagagtcgggggccggccgcttcgagcagacatgataa Firefly
Luciferase (SEQ ID NO: 54)
atggaagatgccaaaaacattaagaagggcccagcgccattctacccactcgaagacgggaccgccggcgagcagctgcacaaagccatgaagcgctacgccct
ggtgcccggcaccatcgcctttaccgacgcacatatcgaggtggacattacctacgccgagtacttcgagatgagcgttcggctggcagaagctatgaagcgctatg
ggctgaatacaaaccatcggatcgtggtgtgcagcgagaatagcttgcagttcttcatgcccgtgttgggtgccctgttcatcggtgtggctgtggccccagctaacga
catctacaacgagcgcgagctgctgaacagcatgggcatcagccagcccaccgtcgtattcgtgagcaagaaagggctgcaaaagatcctcaacgtgcaaaagaa
gctaccgatcatacaaaagatcatcatcatggatagcaagaccgactaccagggcttccaaagcatgtacaccttcgtgacttcccatttgccacccggcttcaacgagt
acgacttcgtgcccgagagcttcgaccgggacaaaaccatcgccctgatcatgaacagtagtggcagtaccggattgcccaaggggtagccctaccgcaccgcac
cgcttgtgtccgattcagtcatgcccgcgaccccatcttcggcaaccagatcatccccgacaccgctatcctcagcgtggtgccatttcaccacggcttcggcatgttca
ccacgctgggctacttgatctgcggctttcgggtcgtgctcatgtaccgcttcgaggaggagctattcttgcgcagcttgcaagactataagattcaatctgccctgctgg
tgcccacactatttagcttcttcgctaagagcactctcatcgacaagtacgacctaagcaacttgcacgagatcgccagcggcggggcgccgctcagcaaggaggta
ggtgaggccgtggccaaacgcttccacctaccaggcatccgccagggctacggcctgacagaaacaaccagcgccattctgatcacccccgaaggggacgacaa
gcctggcgcagtaggcaaggtggtgcccttcttcgaggctaaggtggtggacttggacaccggtaagacactgggtgtgaaccagcgcggcgagctgtgcgtccgt
ggccccatgatcatgagcggctacgttaacaaccccgaggctacaaacgctctcatcgacaaggacggctggctgcacagcggcgacatcgcctactgggacgag
gacgagcacttcttcatcgtggaccggctgaagagcctgatcaaatacaagggctaccaggtagccccagccgaactggagagcatcctgctgcaacaccccaaca
tcttcgacgccgggggtcgccggcctgcccgacgacgatgccggcgagctgcccgccgcagtcgtcgtgctggaacacggtaaaaccatgaccgagaaggagatc
gtggactatgtggccagccaggttacaaccgccaagaagctgcgcggtggtgttgtgttcgtggacgaggtgcctaaaggactgaccggcaagttggacgcccgca
agatccgcgagattctcattaaggccaagaagggcggcaagatcgccgtgtaa FMDV 2A (SEQ ID NO: 55)
GTAAAGCAAACACTGAACTTTGACCTTCTCAAGTTGGCTGGAGACGTTGAGTCCAATCCTGGGCCC
REFERENCES
[0926] 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 [0927] 2. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* 348, 69-74 (2015). [0928] 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). [0929] 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). [0930] 5. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189-2199 (2014). [0931] 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). [0932] 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). [0933] 8. Hacohen, N. & Wu, C. J.-Y. U. S. patent application Ser. No. 01/029,3637—COMPOSITIONS AND METHODS OF IDENTIFYING TUMOR SPECIFIC NEOANTIGENS. (A1). at appft1.uspto.gov/netacgi/nph-Parser?
Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r-
1&f=G&1=50&s1=20110293637.PGNR. [0934] 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). [0935] 10. Yadav, M. et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* 515, 572-576 (2014). [0936] 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). [0937] 12. Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350, 207-211 (2015). [0938] 13. Yoshida, K. & Ogawa, S. Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* 5, 445-459 (2014). [0939] 14. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543-550 (2014). [0940] 15. Rajasagi, M. et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* 124, 453-462 (2014). [0941] 16. Downing, S. R. et al. United States Patent Application: 0120208706—OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES. (A1). at 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.PGNR. [0942] 17. Target Capture for NextGen Sequencing—IDT. at www.idtdna.com/pages/products/nextgen/target-capture. [0943] 18. Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* 33, 1152-1158 (2015). [0944] 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). [0945] 20. Bodini, M. et al. The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations. *Blood* 125, 600-605 (2015). [0946] 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). [0947] 22. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213-219 (2013). [0948] 23. Wilkerson, M. D. et al. Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* 42, e107 (2014). [0949] 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). [0950] 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). [0951] 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). [0952] 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). [0953] 28. Boegel, S. et al. HLA typing from RNA-Seq sequence reads. *Genome Med.* 4, 102 (2012). [0954] 29. Liu, C. et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res.* 41, e142 (2013). [0955] 30. Mayor, N. P. et al. HLA Typing for the Next Generation. *PloS One* 10, e0127153 (2015). [0956] 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). [0957] 32. Song, L. & Florea, L. CLASS: constrained transcript assembly of RNA-seq reads. *BMC Bioinformatics* 14 Suppl 5, S14 (2013). [0958] 33. Maretty, L., Sibbesen, J. A. & Krogh, A. Bayesian transcriptome assembly. *Genome Biol.* 15, 501 (2014). [0959] 34. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290-295 (2015). [0960] 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 [0961] 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). [0962] 37. Rivas, M. A. et al. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* 348, 666-669 (2015). [0963] 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). [0964] 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). [0965] 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 [0966] 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 [0967] 42. Maguire, S. L. et al. SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* 235, 571-580 (2015). [0968] 43. Carithers, L. J. et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation Biobanking* 13, 311-319 (2015). [0969] 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). [0970] 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 [0971] 46. Jorgensen, 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). [0972] 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). [0973] 48. Nielsen, M., Lundegaard, C., Lund, O. & Ke mir, 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). [0974] 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). [0975] 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). [0976] 51. Janeway's Immunobiology: 9780815345312: Medicine & Health Science Books @Amazon.com. at www.amazon.com/Janeways-Immunobiology-Kenneth-Murphy/dp/0815345313. [0977] 52. Calis, J. J. A. et al. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* 9, e1003266 (2013). [0978] 53. Zhang, J. et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 346, 256-259 (2014) [0979] 54. Walter, M. J. et al. Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366, 1090-1098 (2012). [0980] 55. Hunt D F, Henderson R A, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox A L, Appella E, Engelhard V H. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* 1992. 255: 1261-1263. [0981] 56. Zarling A L, Polefrone J M, Evans A M, Mikesh L M, Shabanowitz J, Lewis S T, Engelhard V H, Hunt D F. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. *Proc Natl Acad Sci USA.* 2006 Oct. 3; 103(40):14889-94. [0982] 57. 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. 2015 March;14(3):658-73. doi: 10.1074/mcp.M114.042812. [0983] 58. Abelin J G, Trantham P D, Penny S A, Patterson A M, Ward S T, Hildebrand W H, Cobbold M, Bai D L, Shabanowitz J, Hunt D F. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. *Nat Protoc.* 2015 September; 10(9):1308-18. doi: 10.1038/nprot.2015.086. Epub 2015 Aug. 6 [0984] 59. Barnstable C J, Bodmer W F, Brown G, Galfre G, Milstein C, Williams A F, 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. [0985] 60. Goldman J M, Hibbin J, Kearney L, Orchard K, Th'ng K H. HLA-D R monoclonal antibodies inhibit the proliferation of normal and chronic granulocytic leukaemia myeloid progenitor cells. *Br J Haematol.* 1982 November; 52(3):411-20. [0986] 61. Eng J K, Jahan T A, Hoopmann M R. Comet: an open-source M S/M S sequence database search tool. Proteomics. 2013 January; 13(1):22-4. doi: 10.1002/pmic.201200439. Epub 2012 Dec. 4. [0987] 62. Eng J K, Hoopmann M R, Jahan T A, Egertson J D, Noble W S, MacCoss M J. A deeper look into Comet—implementation and features. *J Am Soc Mass Spectrom.* 2015 November; 26(11):1865-74. doi: 10.1007/s13361-015-1179-x. Epub 2015 Jun. 27. [0988] 63. Lukas Kall, 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 [0989] 64. Lukas Kall, 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 [0990] 65. Lukas Kall, 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 [0991] 66. Kinney R M, B J Johnson, V L Brown, D W 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. [0992] 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 Oct. 2000, *Pages* 529-538. [0993] 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 USA.;* 93(18): 9730-9735, 1996 Sep. 3. [0994] 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. [0995] 70. Aarnoudse, C. A., KrUse, 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. [0996] 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 D R-restricted helper epitopes by modification of high affinity D R-blocking peptides. *Immunity* 1, 751-761. [0997] 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. [0998] 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. [0999] 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. [1000] 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. [1001] 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. [1002] 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. [1003] 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. [1004] 79. Panina-Bordignon, P., Tan, A., Termijtelen, A., Demotz, S., Corradin, G., and Lanzavecchia, A. (1989). Universally immunogenic T cell epitopes: promiscuous binding to human MHC class II and promiscuous recognition by T cells. *Eur J Immunol* 19, 2237-2242. [1005] 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. [1006] 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. [1007] 82. Pushko P, Parker M, Ludwig G V, Davis N L, Johnston R E, Smith J F. 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. [1008] 83. Strauss, J H and E G Strauss. The alphaviruses: gene expression, replication, and evolution. *Microbiol Rev.* 1994 September; 58(3): 491-562. [1009] 84. Rhême C, Ehrengruber M U, Grandgirard D. Alphaviral cytotoxicity and its implication in vector development. *Exp Physiol.* 2005 January; 90(1):45-52. Epub 2004 Nov. 12. [1010] 85. Riley, Michael K. II, and Wilfred Vermerris. Recent Advances in Nanomaterials for Gene Delivery-A Review. Nanomaterials 2017, 7(5), 94. [1011] 86. Frolov I, Hardy R, Rice C M. Cis-acting RNA elements at the 5′ end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. RNA. 2001 November; 7(11):1638-51. [1012] 87. Jose J, Snyder J E, Kuhn R J. A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol.* 2009 September; 4(7):837-56. [1013] 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 [1014] 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, [1015] 90. Juliane Liepe, Fabio Marino, John Sidney, Anita Jeko, Daniel E. Bunting, Alessandro Sette, Peter M. Kloetzel, 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. [1016] 91. Mommen G P., Marino, F., Meiring H D., Poelen, M C., van Gaans-van den Brink, J A., Mohammed S., Heck A J., and van Els C A. 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. [1017] 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 Tureci, and Ugur Sahin. Mutant MHC class II epitopes drive therapeutic immune responses to caner. *Nature* 520, 692-696, April 2015. [1018] 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. [1019] 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. [1020] 95. Nielsen, M., Lund, O. N N-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics* 10:296, September 2009. [1021] 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. [1022] 97. Zhang, J., et al. PEAKS D B: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Molecular & Cellular Proteomics.* 11(4):1-8.

1/2/2012. [1023] 98. Jensen, Kamilla Kjaergaard, et al. “Improved Methods for Prediting Peptide Binding Affinity to MHC Class II Molecules.” *Immunology*, 2018, doi:10.1111/imm.12889. [1024] 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 [1025] 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. [1026] 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. [1027] 102. Van Loo, P., Nordgard, S. H., Lingjorde, 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. [1028] 103. Van Loo, P., Nordgard, S. H., Lingjorde, 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.

VARIOUS EMBODIMENTS

[1029] 1. Disclosed herein is a viral vector comprising a neoantigen or plurality of neoantigens. In certain embodiments, a neoantigen is identified using a method disclosed herein, e.g., below. In certain embodiments, a neoantigen has at least one characteristic or property as disclosed herein, e.g., below. [1030] 2. Disclosed herein is a method for identifying one or more neoantigens from a tumor cell of a subject that are likely to be presented on the tumor cell surface, comprising the steps of: [1031] obtaining at least one of exome, transcriptome or whole genome tumor nucleotide sequencing data from the tumor cell of the subject, wherein the tumor nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of neoantigens, and wherein the peptide sequence of each neoantigen comprises at least one alteration that makes it distinct from the corresponding wild-type, parental peptide sequence; [1032] inputting the peptide sequence of each neoantigen into one or more presentation models to generate a set of numerical likelihoods that each of the neoantigens is presented by one or more MHC alleles on the tumor cell surface of the tumor cell of the subject, the set of numerical likelihoods having been identified at least based on received mass spectrometry data; and [1033] selecting a subset of the set of neoantigens based on the set of numerical likelihoods to generate a set of selected neoantigens. [1034] 3. In certain embodiments, a number of the set of selected neoantigens is 20. [1035] 4. In certain embodiments, the presentation model represents dependence between: presence of a pair of a particular one of the MHC alleles and a particular amino acid at a particular position of a peptide sequence; and likelihood of presentation on the tumor 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. [1036] 5. In certain embodiments, inputting the peptide sequence comprises: [1037] applying the one or more presentation models to the peptide sequence of the corresponding neoantigen to generate a dependency score for each of the one or more MHC alleles indicating whether the MHC allele will present the corresponding neoantigen based on at least positions of amino acids of the peptide sequence of the corresponding neoantigen. [1038] 6. In certain embodiments, the method further comprises: [1039] transforming the dependency scores to generate a corresponding per-allele likelihood for each MHC allele indicating a likelihood that the corresponding MHC allele will present the corresponding neoantigen; and [1040] combining the per-allele likelihoods to generate the numerical likelihood. [1041] 7. In certain embodiments, the transforming the dependency scores model the presentation of the peptide sequence of the corresponding neoantigen as mutually exclusive. [1042] 8. In certain embodiments, the method further comprises: [1043] transforming a combination of the dependency scores to generate the numerical likelihood. [1044] 9. In certain embodiments, the transforming the combination of the dependency scores models the presentation of the peptide sequence of the corresponding neoantigen as interfering between MHC alleles. [1045] 10. In certain embodiments, the set of numerical likelihoods are further identified by at least an allele noninteracting feature, and further comprising: [1046] applying an allele noninteracting one of the one or more presentation models to the allele noninteracting features to generate a dependency score for the allele noninteracting features indicating whether the peptide sequence of the corresponding neoantigen will be presented based on the allele noninteracting features. [1047] 11. In certain embodiments, the method further comprises: [1048] combining the dependency score for each MHC allele in the one or more MHC alleles with the dependency score for the allele noninteracting feature; [1049] transforming the combined dependency scores for each MHC allele to generate a corresponding per-allele likelihood for the MHC allele indicating a likelihood that the corresponding MHC allele will present the corresponding neoantigen; and [1050] combining the per-allele likelihoods to generate the numerical likelihood. [1051] 12. In certain embodiments, the method further comprises: [1052] transforming a combination of the dependency scores for each of the MHC alleles and the dependency score for the allele noninteracting features to generate the numerical likelihood. [1053] 13. In certain

embodiments, a set of numerical parameters for the presentation model is trained based on a training data set including at least a set of training peptide sequences identified as present in a plurality of samples and one or more MHC alleles associated with each training peptide sequence, wherein the training peptide sequences are identified through mass spectrometry on isolated peptides eluted from MHC alleles derived from the plurality of samples. [1054] 14. In certain embodiments, the training data set further includes data on mRNA expression levels of the tumor cell. [1055] 15. In certain embodiments, the samples comprise cell lines engineered to express a single MHC class I or class II allele. [1056] 16. In certain embodiments, the samples comprise cell lines engineered to express a plurality of MHC class I or class II alleles. [1057] 17. In certain embodiments, the samples comprise human cell lines obtained or derived from a plurality of patients. [1058] 18. In certain embodiments, the samples comprise fresh or frozen tumor samples obtained from a plurality of patients. [1059] 19. In certain embodiments, the samples comprise fresh or frozen tissue samples obtained from a plurality of patients. [1060] 20. In certain embodiments, the samples comprise peptides identified using T-cell assays. [1061] 21. In certain embodiments, the training data set further comprises data associated with: [1062] peptide abundance of the set of training peptides present in the samples; [1063] peptide length of the set of training peptides in the samples. [1064] 22. In certain embodiments, the training data set is generated by comparing the set of training peptide sequences via alignment to a database comprising a set of known protein sequences, wherein the set of training protein sequences are longer than and include the training peptide sequences. [1065] 23. In certain embodiments, the training data set is generated based on performing or having performed mass spectrometry on a cell line to obtain at least one of exome, transcriptome, or whole genome peptide sequencing data from the cell line, the peptide sequencing data including at least one protein sequence including an alteration. [1066] 24. In certain embodiments, the training data set is generated based on obtaining at least one of exome, transcriptome, and whole genome normal nucleotide sequencing data from normal tissue samples. [1067] 25. In certain embodiments, the training data set further comprises data associated with proteome sequences associated with the samples. [1068] 26. In certain embodiments, the training data set further comprises data associated with MHC peptidome sequences associated with the samples. [1069] 27. In certain embodiments, the training data set further comprises data associated with peptide-MHC binding affinity measurements for at least one of the isolated peptides. [1070] 28. In certain embodiments, the training data set further comprises data associated with peptide-MHC binding stability measurements for at least one of the isolated peptides. [1071] 29. In certain embodiments, the training data set further comprises data associated with transcriptomes associated with the samples. [1072] 30. In certain embodiments, the training data set further comprises data associated with genomes associated with the samples. [1073] 31. In certain embodiments, the training peptide sequences are of lengths within a range of k-mers where k is between 8-15, inclusive. [1074] 32. In certain embodiments, the method further comprises encoding the peptide sequence using a one-hot encoding scheme. [1075] 33. In certain embodiments, the method further comprises encoding the training peptide sequences using a left-padded one-hot encoding scheme. [1076] 34. Also disclosed herein is a method of treating a subject having a tumor, comprising performing any of the steps of the methods disclosed herein, and further comprising obtaining a tumor vaccine comprising the set of selected neoantigens, and administering the tumor vaccine to the subject. [1077] 35. Also disclosed herein is a method of manufacturing a tumor vaccine, comprising performing any of the steps a method disclosed herein, and further comprising producing or having produced a tumor vaccine comprising the set of selected neoantigens. [1078] 36. Also disclosed herein is a tumor vaccine comprising a set of selected neoantigens, selected by performing a method disclosed herein. [1079] 37. In certain embodiments, the tumor vaccine comprises one or more of a nucleotide sequence, a polypeptide sequence, RNA, DNA, a cell, a plasmid, or a vector. [1080] 38. In certain embodiments, the tumor vaccine comprises one or more neoantigens presented on the tumor cell surface. [1081] 39. In certain embodiments, the tumor vaccine comprises one or more neoantigens that is immunogenic in the subject. [1082] 40. In certain embodiments, the tumor vaccine does not comprise one or more neoantigens that induce an autoimmune response against normal tissue in the subject. [1083] 41. In certain embodiments, the tumor vaccine further comprises an adjuvant. [1084] 42. In certain embodiments, the tumor vaccine further comprises an excipient. [1085] 43. In certain embodiments, selecting the set of selected neoantigens comprises selecting neoantigens that have an increased likelihood of being presented on the tumor cell surface relative to unselected neoantigens based on the presentation model. [1086] 44. In certain embodiments, selecting the set of selected neoantigens comprises selecting neoantigens that have an increased likelihood of being capable of inducing a tumor-specific immune response in the subject relative to unselected neoantigens based on the presentation model. [1087] 45. In certain embodiments, selecting the set of selected neoantigens comprises selecting neoantigens that have an increased likelihood of being capable of being presented to naïve T cells by professional antigen presenting cells (APCs) relative to unselected neoantigens based on the presentation model, optionally wherein the APC is a dendritic cell (DC). [1088] 46. In certain embodiments, selecting the set of selected neoantigens comprises selecting neoantigens that have a decreased likelihood of being subject to

inhibition via central or peripheral tolerance relative to unselected neoantigens based on the presentation model. [1089] 47. In certain embodiments, selecting the set of selected neoantigens comprises selecting neoantigens that have a decreased likelihood of being capable of inducing an autoimmune response to normal tissue in the subject relative to unselected neoantigens based on the presentation model. [1090] 48. In certain embodiments, exome or transcriptome nucleotide sequencing data is obtained by performing sequencing on the tumor tissue. [1091] 49. In certain embodiments, sequencing is next generation sequencing (NGS) or any massively parallel sequencing approach. [1092] 50. In certain embodiments, the set of numerical likelihoods are further identified by at least MHC-allele interacting features comprising at least one of: [1093] a. The predicted affinity with which the MHC allele and the neoantigen encoded peptide bind. [1094] b. The predicted stability of the neoantigen encoded peptide-MHC complex. [1095] c. The sequence and length of the neoantigen encoded peptide. [1096] d. The probability of presentation of neoantigen encoded peptides with similar sequence in cells from other individuals expressing the particular MHC allele as assessed by mass-spectrometry proteomics or other means. [1097] e. The expression levels of the particular MHC allele in the subject in question (e.g. as measured by RNA-seq or mass spectrometry). [1098] f. The overall neoantigen encoded peptide-sequence-independent probability of presentation by the particular MHC allele in other distinct subjects who express the particular MHC allele. [1099] g. The overall neoantigen encoded peptide-sequence-independent probability of presentation by MHC alleles in the same family of molecules (e.g., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) in other distinct subjects. [1100] 51. In certain embodiments, the set of numerical likelihoods are further identified by at least MHC-allele noninteracting features comprising at least one of: [1101] a. The C- and N-terminal sequences flanking the neoantigen encoded peptide within its source protein sequence. [1102] b. The presence of protease cleavage motifs in the neoantigen encoded peptide, optionally weighted according to the expression of corresponding proteases in the tumor cells (as measured by RNA-seq or mass spectrometry). [1103] c. The turnover rate of the source protein as measured in the appropriate cell type. [1104] d. The length of the source protein, optionally considering the specific splice variants (“isoforms”) most highly expressed in the tumor cells as measured by RNA-seq or proteome mass spectrometry, or as predicted from the annotation of germline or somatic splicing mutations detected in DNA or RNA sequence data. [1105] e. The level of expression of the proteasome, immunoproteasome, thymoproteasome, or other proteases in the tumor cells (which may be measured by RNA-seq, proteome mass spectrometry, or immunohistochemistry). [1106] f. The expression of the source gene of the neoantigen encoded peptide (e.g., as measured by RNA-seq or mass spectrometry). [1107] g. The typical tissue-specific expression of the source gene of the neoantigen encoded peptide during various stages of the cell cycle. [1108] h. A comprehensive catalog of features of the source protein and/or its domains as can be found in e.g. uniProt or PDB www.rcsb.org/pdb/home/home.do. [1109] i. Features describing the properties of the domain of the source protein containing the peptide, for example: secondary or tertiary structure (e.g., alpha helix vs beta sheet); Alternative splicing. [1110] j. The probability of presentation of peptides from the source protein of the neoantigen encoded peptide in question in other distinct subjects. [1111] k. The probability that the peptide will not be detected or over-represented by mass spectrometry due to technical biases. [1112] l. The expression of various gene modules/pathways as measured by RNASeq (which need not contain the source protein of the peptide) that are informative about the state of the tumor cells, stroma, or tumor-infiltrating lymphocytes (TILs). [1113] m. The copy number of the source gene of the neoantigen encoded peptide in the tumor cells. [1114] n. The probability that the peptide binds to the TAP or the measured or predicted binding affinity of the peptide to the TAP. [1115] o. The expression level of TAP in the tumor cells (which may be measured by RNA-seq, proteome mass spectrometry, immunohistochemistry). [1116] p. Presence or absence of tumor mutations, including, but not limited to: [1117] i. Driver mutations in known cancer driver genes such as EGFR, KRAS, ALK, RET, ROS1, TP53, CDKN2A, CDKN2B, NTRK1, NTRK2, NTRK3 [1118] ii. In genes encoding the proteins involved in the antigen presentation machinery (e.g., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 or any of the genes coding for components of the proteasome or immunoproteasome). Peptides whose presentation relies on a component of the antigen-presentation machinery that is subject to loss-of-function mutation in the tumor have reduced probability of presentation. [1119] q. Presence or absence of functional germline polymorphisms, including, but not limited to: [1120] i. In genes encoding the proteins involved in the antigen presentation machinery (e.g., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 or any of the genes coding for components of the proteasome or immunoproteasome) [1121] r. Tumor type (e.g., NSCLC, melanoma). [1122] s. Clinical tumor subtype (e.g., squamous lung cancer vs. non-squamous). [1123] t. Smoking

history. [1124] u. The typical expression of the source gene of the peptide in the relevant tumor type or clinical subtype, optionally stratified by driver mutation. [1125] 52. In certain embodiments, the at least one mutation is a frameshift or nonframeshift indel, missense or nonsense substitution, splice site alteration, genomic rearrangement or gene fusion, or any genomic or expression alteration giving rise to a neoORF. [1126] 53. In certain embodiments, the tumor cell 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, 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. [1127] 54. In certain embodiments, the method further comprises obtaining a tumor vaccine comprising the set of selected neoantigens or a subset thereof, optionally further comprising administering the tumor vaccine to the subject. [1128] 55. In certain embodiments, at least one of neoantigens in the set of selected neoantigens, when in polypeptide form, comprises at least one of: a binding affinity with MHC with an IC50 value of less than 1000 nM, for MHC Class 1 polypeptides a length of 8-15, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids, presence of sequence motifs within or near the polypeptide in the parent protein sequence promoting proteasome cleavage, and presence of sequence motifs promoting TAP transport. [1129] 56. Also disclosed herein is a method for generating a model for identifying one or more neoantigens that are likely to be presented on a tumor cell surface of a tumor cell, comprising executing the steps of: [1130] receiving mass spectrometry data comprising data associated with a plurality of isolated peptides eluted from major histocompatibility complex (MHC) derived from a plurality of samples; [1131] obtaining a training data set by at least identifying a set of training peptide sequences present in the samples and one or more MHCs associated with each training peptide sequence; [1132] training a set of numerical parameters of a presentation model using the training data set comprising the training peptide sequences, the presentation model providing a plurality of numerical likelihoods that peptide sequences from the tumor cell are presented by one or more MHC alleles on the tumor cell surface. [1133] 57. In certain embodiments, the presentation model represents dependence between: presence of a particular amino acid at a particular position of a peptide sequence; and likelihood of presentation, by one of the MHC alleles on the tumor cell, of the peptide sequence containing the particular amino acid at the particular position. [1134] 58. In certain embodiments, the samples comprise cell lines engineered to express a single MHC [1135] class I or class II allele. [1136] 59. In certain embodiments, the samples comprise cell lines engineered to express a plurality of [1137] MHC class I or class II alleles. [1138] 60. In certain embodiments, the samples comprise human cell lines obtained or derived from a [1139] plurality of patients. [1140] 61. In certain embodiments, the samples comprise fresh or frozen tumor samples obtained from [1141] a plurality of patients. [1142] 62. In certain embodiments, the samples comprise peptides identified using T-cell assays. [1143] 63. In certain embodiments, the training data set further comprises data associated with: [1144] peptide abundance of the set of training peptides present in the samples; [1145] peptide length of the set of training peptides in the samples. [1146] 64. In certain embodiments, obtaining the training data set comprises: [1147] obtaining a set of training protein sequences based on the training peptide sequences by comparing the set of training peptide sequences via alignment to a database comprising a set of known protein sequences, wherein the set of training protein sequences are longer than and include the training peptide sequences. [1148] 65. In certain embodiments, obtaining the training data set comprises: [1149] performing or having performed mass spectrometry on a cell line to obtain at least one of exome, transcriptome, or whole genome nucleotide sequencing data from the cell line, the nucelotide sequencing data including at least one protein sequence including a mutation. [1150] 66. In certain embodiments, training the set of parameters of the presentation model comprises: encoding the training peptide sequences using a one-hot encoding scheme. [1151] 67. In certain embodiments, the method further comprises: [1152] obtaining at least one of exome, transcriptome, and whole genome normal nucleotide sequencing data from normal tissue samples; and [1153] training the set of parameters of the presentation model using the normal nucleotide sequencing data. [1154] 68. In certain embodiments, the training data set further comprises data associated with proteome sequences associated with the samples. [1155] 69. In certain embodiments, the training data set further comprises data associated with MHC peptidome sequences associated with the samples. [1156] 70. In certain embodiments, the training data set further comprises data associated with peptide-MHC binding affinity measurements for at least one of the isolated peptides. [1157] 71. In certain embodiments, the training data set further comprises data associated with peptide-MHC binding stability measurements for at least one of the isolated peptides. [1158] 72. In certain embodiments, the training data set further comprises data associated with transcriptomes associated with the samples. [1159] 73. In certain embodiments, the training data set further comprises data associated with genomes associated with the samples. [1160] 74. In certain embodiments, training the set of numerical parameters further comprises: logistically regressing the set of parameters. [1161] 75. In certain embodiments, the training peptide sequences are of lengths within a range of k-mers where k is between 8-15, inclusive. [1162] 76. In certain embodiments, training the set of numerical parameters of the presentation

model comprises: [1163] encoding the training peptide sequences using a left-padded one-hot encoding scheme. [1164] 77. In certain embodiments, training the set of numerical parameters further comprises: determining values for the set of parameters using a deep learning algorithm. [1165] 78. Also disclosed herein is a method for generating a model for identifying one or more neoantigens that are likely to be presented on a tumor cell surface of a tumor cell, comprising executing the steps of: [1166] receiving mass spectrometry data comprising data associated with a plurality of isolated peptides eluted from major histocompatibility complex (MHC) derived from a plurality of fresh or frozen tumor samples; [1167] obtaining a training data set by at least identifying a set of training peptide sequences present in the tumor samples and presented on one or more MHC alleles associated with each training peptide sequence; [1168] obtaining a set of training protein sequences based on the training peptide sequences; and [1169] training a set of numerical parameters of a presentation model using the training protein sequences and the training peptide sequences, the presentation model providing a plurality of numerical likelihoods that peptide sequences from the tumor cell are presented by one or more MHC alleles on the tumor cell surface. [1170] 79. In certain embodiments, the presentation model represents dependence between: [1171] 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 [1172] likelihood of presentation on the tumor 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.

Claims

1-173. (canceled)
174. A method for inducing an immune response in a subject, comprising: (A) administering a first vaccine composition, wherein the first vaccine composition comprises a vector encoding a first neoantigen-encoding nucleic acid sequence; and (B) administering a second vaccine composition, wherein the second vaccine composition comprises a RNA vector encoding a second neoantigen-encoding nucleic acid sequence, wherein the vector encoding the first neoantigen-encoding nucleic acid sequence is not a RNA vector.
175. The method of claim 174, wherein the first neoantigen-encoding nucleic acid sequence and the second neoantigen-encoding nucleic acid sequence comprise the same neoantigen-encoding nucleic acid sequence.
176. The method of claim 174, wherein the second neoantigen-encoding nucleic acid sequence is different than the first neoantigen-encoding nucleic acid sequence.
177. The method of claim 174, wherein the first vaccine composition is administered prior to the second vaccine composition.
178. The method of claim 174, wherein the first vaccine composition is administered is administered as a priming dose.
179. The method of claim 174, wherein the RNA vector comprises a +-stranded RNA vector.
180. The method of claim 174, wherein the RNA vector comprises a RNA alphavirus vector.
181. The method of claim 180, wherein the RNA alphavirus vector comprises a RNA alphavirus backbone comprising: (a) the sequence of SEQ ID NO:3 further comprising a deletion between base pair 7544 and 11175, wherein the second neoantigen cassette is inserted to replace the deletion between base pair 7544 and 11175, and/or (b) the sequence set forth in SEQ ID NO:6, wherein the second neoantigen cassette is inserted at position 7544 of the sequence set forth in SEQ ID NO:6.
182. The method of claim 174, wherein one or both of the first vaccine composition and the second vaccine composition comprises a plurality of neoantigen-encoding nucleic acid sequences comprising junctional epitope sequences encoded by adjacent antigen-encoding nucleic acid sequences in the respective vaccine compositions, wherein at least one or each junctional epitope sequence has an affinity of greater than 500 nM for MHC, and optionally wherein at least one or each junctional epitope sequence is non-self.
183. The method of claim 174, wherein the vector encoding the first neoantigen-encoding nucleic acid sequence comprises a chimpanzee adenovirus vector, optionally wherein the chimpanzee adenovirus vector comprises a ChAdV68-based vector.
184. The method of claim 174, wherein one or both of the first vaccine composition and the second vaccine composition comprise at least two antigen-encoding nucleic acid sequences, wherein each antigen-encoding nucleic acid sequence is linked directly to one another.
185. The method of claim 184, wherein each neoantigen-encoding nucleic acid sequence comprises: (A) a neoepitope-encoding nucleic acid sequence; (B) a 5′ linker sequence that encodes a N-terminal amino acid sequence, and wherein the 5′ linker sequence encodes a peptide that is between 2-20 amino acids in length; and (C) a 3′ linker sequence that encodes a C-terminal amino acid sequence, and wherein the 3′ linker sequence encodes a peptide that is between 2-20 amino acids in length.

186. The method of claim 185, wherein: (A) the N-terminal amino acid sequence comprises a native N-terminal amino acid sequence that flanks the encoded neoepitope in the cognate protein of origin, and (B) the C-terminal amino acid sequence comprises a native C-terminal amino acid sequence that flanks the encoded neoepitope in the cognate protein of origin.
187. The method of claim 174, wherein one or both of the first or the second neoantigen-encoding nucleic acid sequences comprises an MHC class I neoepitope-encoding nucleic acid sequence.
188. The method of claim 174, wherein one or both of the first or the second neoantigen-encoding nucleic acid sequences comprises an neoepitope-encoding nucleic acid sequence comprising an alteration that makes the encoded neoepitope sequence distinct from a corresponding peptide sequence encoded by a wild-type nucleic acid sequence, wherein the alteration comprises one or more alterations selected from the group consisting of: a point mutation, a frameshift mutation, a non-frameshift mutation, a deletion mutation, an insertion mutation, a splice variant, a genomic rearrangement, and a proteasome-generated spliced neoantigen.
189. The method of claim 188, wherein the method further comprises selecting the encoded neoepitope by performing the steps of: obtaining at least one of exome, transcriptome, or whole genome nucleotide sequencing data from both tumor cells and normal cells of a subject, wherein the nucleotide sequencing data is used to obtain data representing peptide sequences of each of a set of MHC neoepitopes identified by comparing the nucleotide sequencing data from the tumor cells and the nucleotide sequencing data from the normal cells, and wherein the peptide sequence of each MHC neoepitope comprises at least one alteration that makes it distinct from the corresponding wild-type peptide sequence identified from the normal cells of the subject; encoding the peptide sequences of each of the MHC neoepitopes into a corresponding numerical vector, each numerical vector including information regarding a plurality of amino acids that make up the peptide sequence and a set of positions of the amino acids in the peptide sequence; inputting the numerical vectors, using a computer processor, into a deep learning presentation model to generate a set of presentation likelihoods for the set of MHC neoepitopes, each presentation likelihood in the set representing the likelihood that a corresponding MHC neoepitope is presented by one or more MHC alleles on the surface of the tumor cells of the subject, the deep learning presentation model comprising: a plurality of parameters identified at least based on a training data set comprising: labels obtained by mass spectrometry measuring presence of peptides bound to at least one MHC allele identified as present in at least one of a plurality of samples; training peptide sequences encoded as numerical vectors including information regarding a plurality of amino acids that make up the peptide sequence and a set of positions of the amino acids in the peptide sequence; and at least one HLA allele associated with the training peptide sequences; and a function representing a relation between the numerical vector received as input and the presentation likelihood generated as output based on the numerical vector and the parameters, the presentation model having a positive predictive value that achieves 0.114 at 10% recall rate.
190. The method of claim 188, wherein the encoded neoepitope is characterized by being predicted to have an increased likelihood of presentation on its corresponding MHC allele relative to the translated, corresponding wild-type nucleic acid sequence.
191. The method of claim 174, wherein one or both of the first or the second neoantigen-encoding nucleic acid sequences is encoded within a cassette, wherein an ordered sequence of each element of the cassette is described in the formula, from 5′ to 3′, comprising:
P.sub.a-(L5.sub.b-N.sub.c-L3.sub.d)X-(G5.sub.e-U.sub.f)Y-G3.sub.g wherein P comprises a second promoter nucleotide sequence, where a=0 or 1, N comprises a neoepitope-encoding nucleic acid sequences, where c=1, L5 comprises a 5′ linker sequence, where b=0 or 1, L3 comprises a 3′ linker sequence, where d=0 or 1, G5 comprises at least one nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56), where e=0 or 1, G3 comprises at least one second nucleic acid sequence encoding a GPGPG amino acid linker sequence (SEQ ID NO:56), where g=0 or 1, U comprises at least one MHC class II antigen-encoding nucleic acid sequence, where f=1, X=2 to 400, and Y=0, 1, or 2; optionally (i) wherein for each X the corresponding N is distinct; and/or (ii) wherein for each Y the corresponding U is distinct.
192. The method of claim 191, wherein: a=0, b=1,d=1, e=1,g=1, X>2, Y>0, the at least one promoter nucleotide sequence is a single 26S promoter nucleotide sequence provided by an RNA alphavirus backbone, the at least one polyadenylation poly(A) sequence is a poly(A) sequence of at least 80 consecutive A nucleotides provided by the RNA alphavirus backbone, the encoded neoepitope comprises an MHC class I epitope, wherein the MHC class I epitope encoded by each N is 7-15 amino acids in length, U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the RNA alphavirus backbone is: (a) the sequence of SEQ ID NO:3 further comprising a deletion between base pair 7544 and 11175, wherein the second neoantigen cassette is inserted to replace the deletion between base pair 7544 and 11175, and/or (b) the sequence set forth in SEQ ID NO:6, wherein the second neoantigen cassette is inserted at position 7544 of the sequence set forth in SEQ ID NO:6 the sequence set forth in SEQ ID NO:6, and each of the neoantigen-encoding nucleic acid sequences encodes a

polypeptide that is between 8 and 35 amino acids in length.

193. The method of claim 174, wherein one or both of the first or the second neoantigen-encoding nucleic acid sequences comprises an neoepitope-encoding nucleic acid sequence that encodes a neoepitope that is predicted to be presented by an MHC allele on the surface of a cell of the subject, optionally wherein the cell is a tumor cell optionally 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.

194. The method of claim 174, wherein one or both of the first vaccine composition and the second vaccine composition comprises at least two neoantigen-encoding nucleic acid sequences, optionally wherein the at least two neoantigen-encoding nucleic acid sequences comprises at least 2-10, 11-20, 15-20, 11-100, 11-200, 11-300, 11-400, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 2, 3, 4, 5, 6, 7, 8, 9, 10 neoantigen-encoding nucleic acid sequences, or up to 400 neoantigen-encoding nucleic acid sequences, optionally wherein at least two of the neoantigen-encoding nucleic acid sequences comprise an neoepitope-encoding nucleic acid sequence that encodes neoepitopes predicted to be presented by an MHC allele on the surface of a cell of the subject.

195. The method of claim 174, wherein the method further comprises administering (1) an adjuvant, and/or (2) an immune modulator, optionally wherein the immune modulator is an anti-CTLA4 antibody or an antigen-binding fragment thereof, an anti-PD-1 antibody or an antigen-binding fragment thereof, an anti-PD-L1 antibody or an antigen-binding fragment thereof, an anti-4-1BB antibody or an antigen-binding fragment thereof, or an anti-OX-40 antibody or an antigen-binding fragment thereof.

196. The method of claim 174, wherein the method comprises treating a subject known or suspected of having cancer, and wherein the method further comprises obtaining or having obtained at least one of exome, transcriptome, or whole genome nucleotide sequencing data from the subject, optionally exome, transcriptome, or whole genome tumor nucleotide sequencing data from a tumor of the subject.