VIRAL DELIVERY OF NEOANTIGENS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Nos. 62/425,996 filed November 23, 2016; 62/435,266 filed December 16, 2016; 62/503,196 filed May 8, 2017; and 62/523,212 filed June 21, 2017, each of which is hereby incorporated in its entirety by reference.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted via EFS-Web and is hereby incorporated herein by reference in its entirety. Said ASCII copy, created on December 27, 2017, is named 38804PCT_CRF_sequencelisting.txt, and is 467,920 bytes in size.

BACKGROUND

[0003] Therapeutic vaccines based on tumor-specific neoantigens hold great promise as a next-generation of personalized cancer immunotherapy. [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. [4,5] Early evidence shows that neoantigen-based vaccination can elicit T-cell responses[6] and that neoantigen targeted cell-therapy can cause tumor regression under certain circumstances in selected patients.[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 [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[9]. Consequently, existing methods are likely to suffer from reduced low positive predictive value (PPV). (Figure 1A)

[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

sequence encoded by a wild-type nucleic acid sequence, b. optionally a 5' linker sequence, and c. optionally a 3' linker sequence; (2) at least one promoter sequence operably linked to at least one sequence of the plurality, (3) optionally, at least one MHC class II antigen-encoding nucleic acid sequence; (4) optionally, at least one GPGPG linker sequence (SEQ ID NO:56); and (5) optionally, at least one polyadenylation sequence..

[0012] Also disclosed herein is a A chimpanzee adenovirus vector comprising: a. a modified ChAdV68 sequence comprising the sequence of SEQ ID NO:1 with an E1 (nt 577 to 3403) deletion and an E3 (nt 27,125-31,825) deletion; b. a CMV promoter sequence; c. an SV40 polyadenylation signal nucleotide sequence; and d. a neoantigen cassette, the neoantigen cassette comprising: (1) a plurality of neoantigen-encoding nucleic acid sequences derived from a tumor present within a subject, the plurality comprising: at least 20 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 is a native 5' nucleic acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 5 amino acids in length, (C) a 3' linker sequence, wherein the 3' linker sequence is a native 3' nucleic acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 5 amino acids in length, and wherein each of the MHC class I neoantigen-encoding nucleic acid sequences encodes a polypeptide that is 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 plurality; and (2) at least two MHC class II antigen-encoding nucleic acid sequences comprising: (A) a PADRE MHC class II sequence (SEQ ID NO:48), (B) a Tetanus toxoid MHC class II sequence (SEQ ID NO:46), (C) a first GPGPG linker sequence (SEQ ID NO: 56) linking the PADRE MHC class II sequence and the Tetanus toxoid MHC class II sequence, (D) a second GPGPG 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 plurality of neoantigen-encoding nucleic acid sequences, (E) a third GPGPG linker sequence (SEQ ID NO: 56) linking the 3' end of the at least two MHC class II antigen-encoding nucleic acid sequences to the SV40 polyadenylation signal nucleotide sequence; and

wherein the neoantigen cassette is inserted within the E1 deletion and the CMV promoter sequence is operably linked to the neoantigen cassette.

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

$$P_a\text{-}(L5_b\text{-}N_c\text{-}L3_d)_X\text{-}(G5_e\text{-}U_f)_Y\text{-}G3_g\text{-}A_h$$

wherein P comprises the at least one promoter sequence operably linked to at least one sequence of the plurality, where a chimpanzee adenovirus vector, optionally= 1, N comprises one of the 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 the wild-type nucleic acid sequence, where c = 1, L5 comprises the 5' linker sequence, where b = 0 or 1, L3 comprises the 3' linker sequence, where d = 0 or 1, G5 comprises one of the at least one GPGPG linker sequences (SEQ ID NO: 56), where e = 0 or 1, G3 comprises one of the at least one GPGPG linker sequences (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, A comprises the at least one polyadenylation sequence, where h = 0 or 1, X = 2 to 400, where for each X the corresponding Nc is a C68distinct MHC class I epitope encoding nucleic acid sequence, and Y = 0-2, where for each Y the corresponding Uf MHC class II antigen-encoding nucleic acid sequence. In a particular aspect, b = 1, d = 1, e = 1, g = 1, h = 1, X = 20, Y = 2, P is a CMV promoter sequence, each N encodes a MHC class I epitope 7-15 amino acids in length, L5 is a native 5' nucleic acid sequence of the MHC I epitope, and wherein the 5' linker sequence encodes a peptide that is at least 5 amino acids in length, L3 is a native 3' nucleic acid sequence of the MHC I epitope, and wherein the 3' linker sequence encodes a peptide that is at least 5 amino acids in length, U is each of a PADRE class II sequence and a Tetanus toxoid MHC class II sequence, the chimpanzee adenovirus vector comprises a modified ChAdV68 sequence comprising the sequence of SEQ ID NO:1 with an E1 (nt 577 to 3403) deletion and an E3 (nt 27,125- 31,825) deletion and the neoantigen cassette is inserted within the E1 deletion, and each of the MHC class I neoantigen-encoding nucleic acid sequences encodes a polypeptide that is 25 amino acids in length.

[0014] In some aspects, at least 1, 2, or optionally 3 neoantigen-encoding nucleic acid sequences in the plurality encode polypeptide sequences or portions thereof that is presented by MHC class I on the tumor cell surface.

[0015] In some aspects, each antigen-encoding nucleic acid sequence in the plurality is linked directly to one another. In some aspects, at least one antigen-encoding nucleic acid sequence in the plurality is linked to a distinct antigen-encoding nucleic acid sequence in the plurality with

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).

[0016] In some aspects, at least one sequence in the plurality 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 plurality. 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.

[0017] In some aspects, at least one of the neoantigen-encoding nucleic acid sequences in the plurality 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 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, parental nucleic acid sequence. In some aspects, at least one of the 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, parental nucleic acid sequence.

[0018] In some aspects, at least one alteration comprises a point mutation, a frameshift mutation, a non-frameshift mutation, a deletion mutation, an insertion mutation, a splice variant, a genomic rearrangement, or a proteasome-generated spliced antigen.

sequence and the neoantigen cassette; transfecting the plasmid sequence into one or more host cells; and isolating the vector from the one or more host cells.

[0047] In some aspects, isolating comprises: lysing the host cell to obtain a cell lysate comprising the vector; and purifying the vector from the cell lysate and optionally also from media used to culture the host cell.

[0048] In some aspects, the plasmid sequence is generated using one of the following; DNA recombination or 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 vector from the cell lysate involves one or more of chromatographic separation, centrifugation, virus precipitation, and filtration.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0049] 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:

[0050] **Figure (FIG.) 1A** shows current clinical approaches to neoantigen identification.

[0051] **FIG. 1B** shows that <5% of predicted bound peptides are presented on tumor cells.

[0052] **FIG. 1C** shows the impact of the neoantigen prediction specificity problem.

[0053] **FIG. 1D** shows that binding prediction is not sufficient for neoantigen identification.

[0054] **FIG. 1E** shows probability of MHC-I presentation as a function of peptide length.

[0055] **FIG. 1F** shows an example peptide spectrum generated from Promega's dynamic range standard. Figure discloses SEQ ID NO: 59.

[0056] **FIG. 1G** shows how the addition of features increases the model positive predictive value.

[0057] **FIG. 2A** is an overview of an environment for identifying likelihoods of peptide presentation in patients, in accordance with an embodiment.

[0058] **FIG. 2B** and **FIG. 2C** illustrate a method of obtaining presentation information, in accordance with an embodiment. FIG. 2B discloses SEQ ID NO: 61. FIG. 2C discloses SEQ ID NOS 61-66, respectively, in order of appearance.

[0059] **FIG. 3** is a high-level block diagram illustrating the computer logic components of the presentation identification system, according to one embodiment.

[0060] **FIG. 4** illustrates an example set of training data, according to one embodiment. Figure discloses Peptide Sequences as SEQ ID NOs 68-71 and C-Flanking Sequences as SEQ ID NOs 72 and 154-155, respectively, in order of appearance.

[0061] FIG. 5 illustrates an example network model in association with an MHC allele.

[0062] FIG. 6 illustrates an example network model shared by MHC alleles.

[0063] FIG. 7 illustrates generating a presentation likelihood for a peptide in association with an MHC allele using an example network model.

[0064] FIG. 8 illustrates generating a presentation likelihood for a peptide in association with a MHC allele using example network models.

[0065] FIG. 9 illustrates generating a presentation likelihood for a peptide in association with MHC alleles using example network models.

[0066] FIG. 10 illustrates generating a presentation likelihood for a peptide in association with MHC alleles using example network models.

[0067] FIG. 11 illustrates generating a presentation likelihood for a peptide in association with MHC alleles using example network models.

[0068] FIG. 12 illustrates generating a presentation likelihood for a peptide in association with MHC alleles using example network models.

[0069] FIG. 13 illustrates performance results of various example presentation models.

[0070] FIG. 14 illustrates an example computer for implementing the entities shown in FIGS. 1 and 3.

[0071] 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.

[0072] FIG 16A 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.

[0073] FIG 16B 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 128-129, 132, 131, 130, 49 and 156, respectively, in order of appearance.

[0074] 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) .

[0075] 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.

[0076] FIG. 19A 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.

[0077] FIG. 19B 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 128-129, 132, 131, 130, 157-159, 133, 160-171, respectively, in order of appearance.

[0078] FIG. 20A 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.

[0079] FIG. 20B 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 172-177, respectively, in order of appearance), mouse (SEQ ID NOS 57-58 and 178-184, respectively, in order of appearance) and human origin (SEQ ID NOS 130-132 and 128-129, respectively, in order of appearance), as well as sequences of 2 universal MHC class II epitopes PADRE and Tetanus toxoid (SEQ ID NOS 49 and 47, respectively, in order of appearance).

[0080] FIG. 21A 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 (40x magnification).

[0081] FIG. 21B 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 40x magnification.

("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.

[0090] FIG. 26B 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 ug 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.

[0091] FIG. 27A 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.

[0092] FIG. 27B 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).

[0096] FIG. 28B 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).

[0097] 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 10^6 splenocytes measured in ELISpot assays presented. Error bars represent standard deviation.

[0098] 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 10^6 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.

[0099] 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.

[00100] 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

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.

[00401] 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.

[00402] 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).

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

[00403] Using the peptide YVYYVADVAAK (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 pmol, 100fmol, 10 fmol, 1 fmol, and 100amol. (Table 1) The results are shown in Figure 1F. These results indicate that the lowest limit of detection (LoD) is in the attomol range (10^{-18}), that the dynamic range spans five orders of magnitude, and that the signal to noise appears sufficient for sequencing at low femtomol ranges (10^{-15}).

Table 1

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

IX. Presentation Model

IX.A. System Overview

[00404] FIG. 2A 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.

[00405] 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. 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.

[00406] 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-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.

IX.B. Presentation Information

[00407] 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

[00408] 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. 2B 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. Since in this situation 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.

[00409] The presented peptide sequences may also be collected from cells that express multiple MHC alleles. Typically in humans, 6 different types of MHC 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. 2C shows an example of this, where the six example peptides, YEMFNDKSF (SEQ ID NO: 61), HROEIFSHDFJ (SEQ ID NO: 62), FJIEJFOESS (SEQ ID NO: 63), NEIOREIREI (SEQ ID NO: 64), JFKSIFEMMSJDSSU (SEQ ID NO: 65), and KNFLENFIESOFI (SEQ ID NO: 66), 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 contrast to single-allele cell lines, 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.

[00410] 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.

[00411] Allele-interacting information can also include measurements or predictions of binding affinity between a given MHC allele and a given peptide. One or more affinity models can generate such predictions. For example, going back to the example shown in FIG. 1D, presentation information 165 may include a binding affinity prediction of 1000nM between the peptide YEMFNDKSF (SEQ ID NO: 61) and the allele HLA-A*01:01. Few peptides with IC50 > 1000nm are presented by the MHC, and lower IC50 values increase the probability of presentation.

[00412] 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. 2C, presentation information 165 may include a stability prediction of a half-life of 1h for the molecule HLA-A*01:01.

[00413] 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.

[00414] 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 Figure 5.

[00415] 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.

[00416] 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).

[00417] 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.

[00418] 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.

[00419] 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.

[00420] 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 11.

[00421] Allele-interacting information can also include the protein sequence of the particular MHC allele.

[00422] 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

[00423] Allele-noninteracting information can include C-terminal sequences flanking the neoantigen encoded peptide within its source protein sequence. 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. 2C, presentation information 165 may include the C-terminal flanking sequence FOEIFNDKSLDKFJI (SEQ ID NO: 67) of the presented peptide FJIEJFOESS (SEQ ID NO: 63) identified from the source protein of the peptide.

[00424] Allele-noninteracting information can also include mRNA quantification measurements. For example, mRNA quantification data can be obtained for the same samples

[00457] The data management module 312 may also include allele-noninteracting variables w^i, such as C-terminal flanking sequences and mRNA quantification measurements in conjunction with the peptide sequences p^i.

[00458] 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.

[00459] 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.

[00460] FIG. 4 illustrates an example set of training data 170A, according to one embodiment. Specifically, the first 3 data instances in the training data 170A 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: 68), FIEUHFWI (SEQ ID NO: 69), and FEWRHRJTRUJR (SEQ ID NO: 70). The fourth data instance in the training data 170A indicates peptide information from a multiple-allele cell line involving the alleles HLA-B*07:02, HLA-C*01:03, HLA-A*01:01and a peptide sequence QIEJOEIJE (SEQ ID NO: 71). The first data instance indicates that peptide sequence QCEIOWARE (SEQ ID NO: 68) 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 170A also includes a binding affinity prediction of 1000nM and a stability prediction of a half-life of 1h for the peptide sequence-allele pair. The training data 170A also includes allele-noninteracting variables, such as the C-terminal flanking sequence of the peptide FJELFISBOSJFIE (SEQ ID NO: 72), and a mRNA quantification

measurement of 10^2 FPKM. The fourth data instance indicates that peptide sequence QIEJOEIJE (SEQ ID NO: 71) was presented by one of the alleles HLA-B*07:02, HLA-C*01:03, or HLA-A*01:01. The training data 170A also includes binding affinity predictions and stability predictions for each of the alleles, as well as the C-flanking sequence of the peptide and the mRNA quantification measurement for the peptide.

IX.C.2. Encoding Module

[00461] 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 $\boldsymbol{p}^i$ with k_i amino acids is represented as a row vector of $20 \cdot k_i$ elements, where a single element among $p^i_{20\cdot(j-1)+1}, p^i_{20\cdot(j-1)+2}, \ldots, p^i_{20\cdot j}$ 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 $\boldsymbol{p}^i$=[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 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]. The C-terminal flanking sequence $\boldsymbol{c}^i$ can be similarly encoded as described above, as well as the protein sequence $\boldsymbol{d}_h$ for MHC alleles, and other sequence data in the presentation information.

[00462] 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_{max} amino acids, the encoding module 314 numerically represents each sequence as a row vector of $(20+1)\cdot k_{max}$ 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 $k_{max}=5$, the same example peptide sequence EAF of 3 amino acids may be represented by the row vector of 105 elements $\boldsymbol{p}^i$=[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 $\boldsymbol{c}^i$ or other sequence data can be similarly encoded as described

XI.A Example 5: Prediction Module

[00564] 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 $\boldsymbol{p}^k$ having 8-15 amino acids. For example, the prediction module 320 may process the given sequence "IEFROEIFJEF (SEQ ID NO: 73) into three peptide sequences having 9 amino acids "IEFROEIFJ (SEQ ID NO: 74)," "EFROEIFJE (SEQ ID NO: 75)," and "FROEIFJEF (SEQ ID NO: 76)." 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.

[00565] 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

[00566] 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 $\boldsymbol{p}^k$, $k=1, 2, \ldots, v$ for inclusion in a vaccine of capacity v, the cassette sequence is given by concatenation of a series of therapeutic epitope sequences $\boldsymbol{p'}^k$, $k=1, 2, \ldots, v$ that each include the sequence of a corresponding peptide $\boldsymbol{p}^k$. 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:

$$\mathbf{C} = [\boldsymbol{p'}^{t_1}\ \boldsymbol{p'}^{t_2}\ \cdots\ \boldsymbol{p'}^{t_v}] \tag{24}$$

[00569] 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^{ti} and p^{tj} may include any epitope sequence that overlaps with both p^{ti} or p^{tj} that is different from the sequences of therapeutic epitopes p^{ti} and p^{tj} themselves. Specifically, each junction between epitope p^{ti} and an adjacent epitope p^{tj} of the cassette with or without an optional linker sequence $l^{(ti,tj)}$ may be associated with $n_{(ti,tj)}$ junction epitopes $e_n^{(ti,tj)}$, $n=1, 2, ..., n_{(ti,tj)}$. The junction epitopes may be sequences that at least partially overlap with both epitopes p^{ti} and p^{tj}, or may be sequences that at least partially overlap with linker sequences placed between the epitopes p^{ti} and p^{tj}. Junction epitopes may be presented by MHC class I, MHC class II, or both.

[00570] FIG. 13 shows two example cassette sequences, cassette 1 ($\mathbf{C}_1$) and cassette 2 ($\mathbf{C}_2$). Each cassette has a vaccine capacity of $v=2$, and includes therapeutic epitopes $p^{t1}=p^1$ = SINFEKL (SEQ ID NO: 185) and $p^{t2}=p^2$ = LLLLLVVVV (SEQ ID NO: 77), and a linker sequence $l^{(t1,t2)}$ = AAY in between the two epitopes. Specifically, the sequence of cassette $\mathbf{C}_1$ is given by $[p^1\ l^{(t1,t2)}\ p^2]$, while the sequence of cassette $\mathbf{C}_2$ is given by $[p^2\ l^{(t1,t2)}\ p^1]$. Example junction epitopes $e_n^{(1,2)}$ of cassette $\mathbf{C}_1$ may be sequences such as EKLAAYLLL (SEQ ID NO: 78), KLAAYLLLLL (SEQ ID NO: 79), and FEKLAAYL (SEQ ID NO: 80) that span across both epitopes p^{t1} and p^{t2} in the cassette, and may be sequences such as AAYLLLLL (SEQ ID NO: 81) and YLLLLLVVV (SEQ ID NO: 82) that span across the linker sequence and a single selected epitope in the cassette. Similarly, example junction epitopes $e_m^{(2,1)}$ of cassette $\mathbf{C}_2$ may be sequences such as VVVVAAYSIN (SEQ ID NO: 83), VVVVAAY (SEQ ID NO: 84), and AYSINFEK (SEQ ID NO: 85). Although both cassettes involve the same set of sequences p^1, $l^{(c1,c2)}$, and p^2, the set of junction epitopes that are identified are different depending on the ordered sequence of the therapeutic epitopes within the cassette.

[00571] 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.[76]

(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 x 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.

[00587] Specifically, the v=20 epitopes were given by:

$p^{,1}$ = YNYSYWISIFAHTMWYNIWHVQWNK (SEQ ID NO: 86)
$p^{,2}$ = IEALPYVFLQDQFELRLLKGEQGNN (SEQ ID NO: 87)
$p^{,3}$ = DSEETNTNYLHYCHFHWTWAQQTTV (SEQ ID NO: 88)
$p^{,4}$ = GMLSQYELKDCSLGFSWNDPAKYLR (SEQ ID NO: 89)
$p^{,5}$ = VRIDKFLMYVWYSAPFSAYPLYQDA (SEQ ID NO: 90)
$p^{,6}$ = CVHIYNNYPRMLGIPFSVMVSGFAM (SEQ ID NO: 91)
$p^{,7}$ = FTFKGNIWIEMAGQFERTWNYPLSL (SEQ ID NO: 92)
$p^{,8}$ = ANDDTPDFRKCYIEDHSFRFSQTMN (SEQ ID NO: 93)
$p^{,9}$ = AAQYIACMVNRQMTIVYHLTRWGMK (SEQ ID NO: 94)
$p^{,10}$ = KYLKEFTQLLTFVDCYMWITFCGPD (SEQ ID NO: 95)
$p^{,11}$ = AMHYRTDIHGYWIEYRQVDNQMWNT (SEQ ID NO: 96)
$p^{,12}$ = THVNEHQLEAVYRFHQVHCRFPYEN (SEQ ID NO: 97)
$p^{,13}$ = QTFSECLFFHCLKVWNNVKYAKSLK (SEQ ID NO: 98)
$p^{,14}$ = SFSSWHYKESHIALLMSPKKNHNNT (SEQ ID NO: 99)
$p^{,15}$ = ILDGIMSRWEKVCTRQTRYSYCQCA (SEQ ID NO: 100)
$p^{,16}$ = YRAAQMSKWPNKYFDFPEFMAYMPI (SEQ ID NO: 101)
$p^{,17}$ = PRPGMPCQHHNTHGLNDRQAFDDFV (SEQ ID NO: 102)
$p^{,18}$ = HNIISDETEVWEQAPHITWVYMWCR (SEQ ID NO: 103)
$p^{,19}$ = AYSWPVVPMKWIPYRALCANHPPGT (SEQ ID NO: 104)
$p^{,20}$ = HVMPHVAMNICNWYEFLYRISHIGR (SEQ ID NO: 105).

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:

C_1 = THVNEHQLEAVYRFHQVHCRFPYENAMHYQMWNTYRAAQMSKWPNKYFDFPEFMAYMPICVH
IYNNYPRMLGIPFSVMVSGFAMAYSWPVVPMKWIPYRALCANHPPGTANDDTPDFRKCYIEDHSFRFSQTMN
IEALPYVFLQDQFELRLLKGEQGNNDSEETNTNYLHYCHFHWTWAQQTTVILDGIMSRWEKVCTRQTRYSYC
QCAFTFKGNIWIEMAGQFERTWNYPLSLSFSSWHYKESHIALLMSPKKNHNNTQTFSECLFFHCLKVWNNVK
YAKSLKHVMPHVAMNICNWYEFLYRISHIGRHNIISDETEVWEQAPHITWVYMWCRVRIDKFLMYVWYSAPF
SAYPLYQDAKYLKEFTQLLTFVDCYMWITFCGPDAAQYIACMVNRQMTIVYHLTRWGMKYNYSYWISIFAHT
MWYNIWHVQWNKGMLSQYELKDCSLGFSWNDPAKYLRPRPGMPCQHHNTHGLNDRQAFDDFV (SEQ ID N
O: 106)

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.

[00588] In the second example, a cassette sequence $\mathbf{C_2}$ 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:

C_2 = IEALPYVFLQDQFELRLLKGEQGNNILDGIMSRWEKVCTRQTRYSYCQCAHVMPHVAMNICN
WYEFLYRISHIGRTHVNEHQLEAVYRFHQVHCRFPYENFTFKGNIWIEMAGQFERTWNYPLSLAMHYQMWNT
SFSSWHYKESHIALLMSPKKNHNNTVRIDKFLMYVWYSAPFSAYPLYQDAQTFSECLFFHCLKVWNNVKYAK
SLKYRAAQMSKWPNKYFDFPEFMAYMPIAYSWPVVPMKWIPYRALCANHPPGTCVHIYNNYPRMLGIPFSVM
VSGFAMHNIISDETEVWEQAPHITWVYMWCRAAQYIACMVNRQMTIVYHLTRWGMKYNYSYWISIFAHTMWY
NIWHVQWNKGMLSQYELKDCSLGFSWNDPAKYLRKYLKEFTQLLTFVDCYMWITFCGPDANDDTPDFRKCYI
EDHSFRFSQTMNDSEETNTNYLHYCHFHWTWAQQTTVPRPGMPCQHHNTHGLNDRQAFDDFV (SEQ ID N
O: 107)

with a presentation score of 1.7. The presentation score of cassette sequence $\mathbf{C_2}$ showed a ~4 fold improvement over the presentation score of cassette sequence $\mathbf{C_1}$, 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 $\mathbf{C_1}$ was 20 seconds on a single thread of a 2.30 GHz Intel Xeon E5-2650 CPU. The run-time for generating cassette $\mathbf{C_2}$ 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.

[00589] 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

[00590] In this example, cassette sequences including v=20 therapeutic epitopes were selecte d based off tumor/normal exome sequencing, tumor transcriptome sequencing and HLA typin g of a lung cancer sample were generated by random sampling 1,000,000 permutations, and b

y 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., 50-1000nM, 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 vaccine need not be the same as the criteria used for determining the distance metric $D(k, m)$ used in the cassette design module 324.

[00591] 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.

[00592] Specifically in this example, the v=20 therapuetic epitopes were

SSTPYLYYGTSSVSYQFPMVPGGDR (SEQ ID NO: 108)
EMAGKIDLLRDSYIFQLFWREAAEP (SEQ ID NO: 109)
ALKQRTWQALAHKYNSQPSVSLRDF (SEQ ID NO: 110)
VSSHSSQATKDSAVGLKYSASTPVR (SEQ ID NO: 111)
KEAIDAWAPYLPEYIDHVISPGVTS (SEQ ID NO: 112)
SPVITAPPSSPVFDTSDIRKEPMNI (SEQ ID NO: 113)
PAEVAEQYSEKLVYMPHTFFIGDHA (SEQ ID NO: 114)
MADLDKLNIHSIIQRLLEVRGS (SEQ ID NO: 115)
AAAYNEKSGRITLLSLLFQKVFAQI (SEQ ID NO: 116)
KIEEVRDAMENEIRTQLRRQAAAHT (SEQ ID NO: 117)
DRGHYVLCDFGSTTNKFQNPQTEGV (SEQ ID NO: 118)
QVDNRKAEAEEAIKRLSYISQKVSD (SEQ ID NO: 119)
CLSDAGVRKMTAAVRVMKRGLENLT (SEQ ID NO: 120)
LPPRSLPSDPFSQVPASPQSQSSSQ (SEQ ID NO: 121)
ELVLEDLQDGDVKMGGSFRGAFSNS (SEQ ID NO: 122)
VTMDGVREEDLASFSLRKRWESEPH (SEQ ID NO: 123)
IVGVMFFERAFDEGADAIYDHINEG (SEQ ID NO: 124)
TVTPTPTPTGTQSPTPTPITTTTTV (SEQ ID NO: 125)
QEEMPPRPCGGHTSSSLPKSHLEPS (SEQ ID NO: 126)
PNIQAVLLPKKTDSHHKAKGK (SEQ ID NO: 127)

[00593] 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.

[00648] 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

[00649] The selected TCRs recognize peptides NLVPMVATV (SEQ ID NO: 128) (PDB# 5D2N), CLGGLLTMV (SEQ ID NO: 129) (PDB#3REV), GILGFVFTL (SEQ ID NO: 130) (PDB#1OGA) LLFGYPVYV (SEQ ID NO: 131) (PDB#1AO7) 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

[00650] Peptides were purchased from ProImmune or Genscript diluted to 10mg/mL with 10mM 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 100µg/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.

[00651] 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.

[00652] 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 (5x10E4-1x10E6 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

[00653] 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: 128), CLGGLLTMV (SEQ ID NO: 129), GLCTLVAML (SEQ ID NO: 132), LLFGYPVYV (SEQ ID NO: 131), GILGFVFTL (SEQ ID NO: 130)) and two irrelevant peptides (WLSLLVPFV (SEQ ID NO: 133), FLLTRICT (SEQ ID NO: 134)) were loaded onto T2 cells. Briefly, T2 cells were counted and diluted to 1x106 cells/mL with IMDM plus 1% FBShi. Peptides were added to result in 10 µg peptide/1x106 cells. Cells were then incubated at 37oC for 90 minutes. Cells were washed twice with IMDM plus 20% FBShi, diluted to 5x10E5 cells/mL and 100 µL plated into a 96-well Costar tissue culture plate. Jurkat-Lucia TCR clones were counted and diluted to 5x10E5 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.

[00654] 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 5x10E5 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 5µl/well. Plates were again incubated for approximately 2 hours at 37°C. Jurkat-Lucia TCR clones were counted and diluted to 5x10E5 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

responses were comparable in magnitude and breadth for both long and short vaccine cassettes, demonstrating that (a) the addition of more epitopes did not impact the magnitude of immune response to the original set of epitopes, and (b) the position of an epitope in a cassette did not influence the ensuing T cell response to it (Table 6).

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

* Suspected technical error caused an absence of a T cell response.

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

[00668] In summary, the findings of the model cassette evaluations (Fig. 16-19, Tables 2-6) demonstrated that, for model vaccine cassettes, optimal immunogenicity 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 best 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: 128) is flanked on its 5' end by the native 5' sequence WQAGILAR (SEQ ID NO: 135) and on its 3' end by the native 3' sequence QGQNLKYQ (SEQ ID NO: 136), thus generating the WQAGILARNLVPMVATVQGQNLKYQ (SEQ ID NO: 137) 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 proved to substantially impact T cell recognition or response. Targeting sequences also did not appear to substantially impact the immunogenicity of cassette-derived antigens.

[00669] 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

[00670] 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 Patent US 6083716) with E1 (nt 457 to 3014) and E3 (nt 27,816- 31,332) sequences deleted. Reporter genes under the control of the CMV promoter/enhancer were inserted in place of the deleted E1 sequences. Transfection of this clone into HEK293 cells did not yield infectious virus. To confirm the sequence of the wild-type C68 virus, isolate VR-594 was obtained from the ATCC, passaged, and then independently sequenced (SEQ ID NO:10). When comparing the AC_000011.1 sequence to the ATCC VR-594 sequence (SEQ ID NO:10) of wild-type ChAdV68 virus , 6 nucleotide differences were identified. In one example, a modified ChAdV68 vector was generated based on AC_000011.1, with the corresponding ATCC VR-594 nucleotides substituted at five positions (ChAdV68.5WTnt SEQ ID NO:1).

[00671] 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

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 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 +/-6.5e3
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

[00690] 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 were measured after immunization of mice with ChAdV68.5WTnt.MAG25mer. Mean cellular immune responses of 8957 or 4019 spot forming cells (SFCs) per 10^6 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

[00691] ***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^7G cap analog (Promega) according to manufacturer's protocol. mRNA was purified using the RNeasy kit (Qiagen) according to manufacturer's protocol.

Table 8. qPCR primers/probes

Target			SEQ ID NO:
Luci	Primer1	GTGGTGTGCAGCGAGAATAG	138
	Primer2	CGCTCGTTGTAGATGTCGTTAG	139
	Probe	/56-FAM/TTGCAGTTC/ZEN/TTCATGCCCGTGTTG/3IABkFQ/	140
GusB	Primer1	GTTTTTGATCCAGACCCAGATG	141
	Primer2	GCCCATTATTCAGAGCGAGTA	142
	Probe	/56-FAM/TGCAGGGTT/ZEN/TCACCAGGATCCAC/3IABkFQ/	143
ActB	Primer1	CCTTGCACATGCCGGAG	144
	Primer2	ACAGAGCCTCGCCTTTG	145
	Probe	/56-FAM/TCATCCATG/ZEN/GTGAGCTGGCGG/3IABkFQ/	146
MAG-25mer	Primer1	CTGAAAGCTCGGTTTGCTAATG	147
Set1	Primer2	CCATGCTGGAAGAGACAATCT	148
	Probe	/56-FAM/CGTTTCTGA/ZEN/TGGCGCTGACCGATA/3IABkFQ/	149
MAG-25mer	Primer1	TATGCCTATCCTGTCTCCTCTG	150
Set2	Primer2	GCTAATGCAGCTAAGTCCTCTC	151
	Probe	/56-FAM/TGTTTACCC/ZEN/TGACCGTGCCTTCTG/3IABkFQ/	152

B16-OVA tumor model

[00696] C57BL/6J mice were injected in the lower left abdominal flank with 10^5 B16-OVA cells/animal. Tumors were allowed to grow for 3 days prior to immunization.

CT26 tumor model

[00697] Balb/c mice were injected in the lower left abdominal flank with 10^6 CT26 cells/animal. Tumors were allowed to grow for 7 days prior to immunization.

Immunizations

[00698] For srRNA vaccine, mice were injected with 10 ug of RNA in 100 uL volume, bilateral intramuscular injection (50 uL per leg). For Ad5 vaccine, mice were injected with $5x10^{10}$ 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-11, BioXcell), 250 ug dose, 2 times per week, via intraperitoneal injection.

In vivo bioluminescent imaging

[00699] 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

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

[00705] 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: 153); 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).

[00706] A strong antigen-specific T-cell response targeting SFL 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 10^6 splenocytes was measured after stimulation with the SFL peptide in ELISpot assays (Fig. 26A, Table 12) and 1.8% (median) of CD8 T-cells were SFL antigen-specific as measured by pentamer staining (Fig. 26B, 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 10^6 splenocytes measured in the ELISpot assay (Fig. 26A, Table 12).

CCTTAAAATTTTTATTTTATTTTTCTTTTCTTTTCCGAATCGGATTTTGTTTTTAATATTTCAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAtacgtagttt
aaac

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

ATGggcggcgcatgagagaagcccagaccaattacctacccaaaatggagaaagttcacgttgacatcga
ggaagacagcccattcctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgat
aatgaccatgctaatgccagagcgttttcgcatctggcttcaaaactgatcgaaacggaggtggacccatccgaca
cgatccttgacattggaagtgcgcccgcccgcagaatgtattctaagcacaagtatcattgtatctgtccgatgag
atgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaaaaactgtaaggaaataactgataag
gaattggacaagaaaatgaaggagctcgccgccgtcatgagcgaccctgacctggaaactgagactatgtgcctcc
acgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgacggaccgacaag
tctctatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttttatgtttaag
aacttggctggagcatatccatcatactctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcc
tatgcagctctgacgttatggagcggtcacgtagagggatgtccattcttagaaagaagtatttgaaaccatccaa
caatgttctattctctgttggctcgaccatctaccacgagaagagggacttactgaggagctggcacctgccgtct
gtatttcacttacgtggcaagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggtacgtcgtta
aaagaatagctatcagtccaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattctt
gtgctgcaaagtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagctacattg
tgtgaccaaatgactggcatactggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaacc
agcgtatagtcgtcaacggtcgcacccagagaaacaccaataccatgaaaaattacctttgcccgtagtggccca
ggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatgaaaggccactaggactacgagatagacag
ttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgcccggatacccaaacca
tcatcaaagtgaacagcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctgag
aacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctctcattaccgccgaggacgtacaagaagct
aagtgcgcagccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctaccacctttggcagctg
atgttgaggagcccactctggaagccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacc
tcgtggcttgataaaggttaccagctacgctggcgaggacaagatcggctcttacgctgtgctttctccgcaggct
gtactcaagagtgaaaaattatcttgcatccaccctctcgctgaacaagtcatagtgataacacactctggccgaa
aagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaatacccgtccaggactt
tcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgcc
acacatggaggagcgctgaacactgatgaagaatattacaaaactgtcaagcccagcgagcacgacggcgaatacc
tgtacgacatcgacaggaaacagtgcgtcaagaaagaactagtcactgggctagggctcacaggcgagctggtgga
tcctcccttccatgaattcgcctacgagagtctgagaacacgaccagccgctccttaccaagtaccaaccataggg
gtgtatggcgtgccaggatcaggcaagtctggcatcattaaaagcgcagtcaccaaaaaagatctagtggtgagcg
ccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatgaaagggctggacgtcaatgccagaactgt
ggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagcttttgcttgtcatgca
ggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggtt
tttttaacatgatgtgcctgaaagtgcattttaaccacgagatttgcacacaagtcttccacaaaagcatctctcg
ccgttgcactaaatctgtgacttcggtcgtctcaaccttgttttacgacaaaaaaatgagaacgacgaatccgaaa
gagactaagattgtgattgacactaccggcagtaccaaacctaagcaggacgatctcattctcacttgtttcagag
ggtgggtgaagcagttgcaaatagattacaaaggcaacgaaataatgacggcagctgcctctcaagggctgacccg
taaaggtgtgtatgccgttcggtacaaggtgaatgaaaatcctctgtacgcacccacctcagaacatgtgaacgtc
ctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaacactgactgcca
agtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttgga
gagaccggaccctaccgacgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaag
accgctggcatagacatgaccactgaacaatggaacactgtggattattttgaaacggacaaagctcactcagcag
agatagtattgaaccaactatgcgtgaggttctttggactcgatctggactccggtctattttctgcacccactgt
tccgttatccattaggaataatcactgggataactccccgtcgcctaacatgtacgggctgaataaagaagtggtc
cgtcagctctctcgcaggtacccacaactgcctcgggcagttgccactggaagagtctatgacatgaacactggta
cactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttagtcctccacca
taatgaacacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaa
aagttgtccgtcccaggcaaaatggttgactggttgtcagaccggcctgaggctaccttcagagctcggctggatt
taggcatcccaggtgatgtgcccaaatatgacataatatttgttaatgtgaggaccccatataaataccatcacta
tcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagcttgtctgcatctgaatcccggcgga
acctgtgtcagcataggttatggttacgctgacagggccagcgaaagcatcattggtgctatagcgcggcagttca
agttttcccgggtatgcaaaccgaaatcctcacttgaagagacggaagttctgtttgtattcattgggtacgatcg
caaggcccgtacgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactccacgaa
gccggatgtgcaccctcatatcatgtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctg
ctaacagcaaaggacaacctggcggaggggtgtgcggagcgctgtataagaaattcccggaaagcttcgatttaca
gccgatcgaagtaggaaaagcgcgactggtcaaaggtg