VIRAL PEPTIDES AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This patent application claims the benefit of U.S. Provisional Application No. 63/583,834, filed September 19, 2023, the disclosure of which is incorporated by reference herein in its entirety for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in XML file format and is hereby incorporated by reference in its entirety. Said XML copy, created on August 29, 2024, is named 250298_000682_SL.xml and is 126,068 bytes in size.

TECHNICAL FIELD

[0003] The present disclosure relates to methods and compositions that involve isolated peptides derived from human T-lymphotropic virus type-1 (HTLV-1), and the use of such methods and compositions for treating, preventing, or reducing the likelihood of HTLV-1 infection and/or HTLV-1-induced diseases.

BACKGROUND

[0004] Human T-lymphotropic virus type-1 (HTLV-1) is a complex retrovirus which belongs to the *Deltaretrovirus* genus, infecting an estimated 5-10 million individuals worldwide. Of the four known types of HTLV (HTLV-1, HTLV-2 HTLV-3, and HTLV-4), HTLV-1 has been identified as the most pathogenic and is the primary etiological agent of a malignant lymphoproliferation called adult T-cell leukemia/lymphoma (ATL), as well as an array of neurological pathologies known as HTLV-1-myelopathy/tropical spastic paraparesis (HAM/TSP). HTLV-1 is endemic in areas of Southwestern Japan, sub-Saharan Africa, South America, and the Caribbean, with foci in the Middle-East and Australo-Melanosia. HTLV-1 is predominately found in $CD4^+$ T lymphocytes and can be spread by, e.g., sexual transmission, contaminated blood products, and mother-to-child transmission (parental/vertical). Further, HTLV-1 is a latent virus, and as such is unable to

be cleared from the host's immune system, where 4-5% of HTLV-1 carriers develop ATL in their life-time and another 0.25-4% develop HAM/TSP. Despite scientific advances toward understanding the pathogenic progression of HTLV-1 infection, the prognosis of ATL, as well as the quality of life for those suffering from HAM/TSP, remains poor. There is no vaccine to prevent HTLV-1 infection or HTLV-1-associated disease. Accordingly, identifying HTLV-1-associated epitopes that can be amenable to targeting such as by therapeutic vaccines, immunotherapy, or other therapeutics that target these HTLV-1-associated antigens would be useful to combat HTLV-1-mediated infections and diseases.

SUMMARY

[0005] As specified in the Background section above, there is a need in the art for development of methods to identify HTLV-1 epitopes that could be helpful for patients infected with HTLV-1. The present application addresses these and other needs.

[0006] In one aspect, provided herein is an isolated peptide comprising an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof, wherein the isolated peptide is 5-20 amino acids in length.

[0007] In some embodiments, the isolated peptide comprises an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146.

[0008] In some embodiments, the isolated peptide consists essentially of an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146.

[0009] In some embodiments, the isolated peptide consists of an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146.

[0010] In another aspect, provided herein is an isolated peptide comprising two or more amino acid sequences selected from any one of SEQ ID NO: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof.

[0011] In some embodiments, the isolated peptide comprises one or more reverse peptide bonds, one or more non-peptide bonds, one or more D-isomers of amino acids, one or more chemical modifications, or any combination thereof.

[0012] In some embodiments, the isolated peptide is produced by expression in a heterologous host cell.

[0013] In some embodiments, the isolated peptide is produced synthetically.

[0014] In some embodiments, the isolated peptide, or pharmaceutically acceptable salt thereof, or fragment or derivative thereof induces a Human T-Lymphotropic Virus Type-1 (HTLV-1)-specific immune response in a subject when presented in a complex with a major histocompatibility complex (MHC) molecule on the surface of an antigen presenting cell (APC).

[0015] In another aspect, provided herein is a fusion protein comprising one or more isolated peptides described herein fused to one or more heterologous molecules.

[0016] In some embodiments, the one or more heterologous molecules enhance a peptide-specific immune response in a subject.

[0017] In some embodiments, the one or more heterologous molecules mediate peptide delivery to a specific site within a subject.

[0018] In some embodiments, the one or more heterologous molecules are a MHC molecule, or a fragment or derivative thereof.

[0019] In another aspect, provided herein is a conjugate comprising one or more isolated peptides described herein conjugated to one or more heterologous molecules.

[0020] In some embodiments, the one or more heterologous molecules enhance a peptide-specific immune response in a subject.

[0021] In some embodiments, the one or more heterologous molecules mediate peptide delivery to a specific site within a subject.

[0022] In some embodiments, the one or more heterologous molecules are an MHC molecule, or a fragment or derivative thereof.

[0023] In some embodiments, the one or more peptides are conjugated to a particle.

[0024] In another aspect, provided herein is an oligomeric complex comprising two or more isolated peptides described herein.

[0025] In another aspect, provided herein is a non-covalent complex comprising an isolated peptide described herein and an MHC molecule, or a fragment or derivative thereof.

[0026] In some embodiments, the MHC molecule, or the fragment thereof, is a class I MHC molecule.

[0027] In some embodiments, the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

[0028] In some embodiments, the MHC molecule, or the fragment thereof, is a class II MHC molecule.

[0029] In some embodiments, the class II MHC molecule is a class II HLA molecule.

[0030] In another aspect, provided herein is a fusion protein comprising an isolated peptide described herein and an MHC molecule, or a fragment or derivative thereof.

[0031] In some embodiments, the MHC molecule, or the fragment thereof, is a class I MHC molecule.

[0032] In some embodiments, the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

[0033] In some embodiments, the MHC molecule, or the fragment thereof, is a class II MHC molecule.

[0034] In some embodiments, the class II MHC molecule is a class II HLA molecule.

[0035] In another aspect, provided herein is a conjugate comprising an isolated peptide described herein and a MHC molecule, or a fragment or derivative thereof.

[0036] In some embodiments, the MHC molecule, or the fragment thereof, is a class I MHC molecule.

[0037] In some embodiments, the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

[0038] In some embodiments, the MHC molecule, or the fragment thereof, is a class II MHC molecule.

[0039] In some embodiments, the class II MHC molecule is a class II HLA molecule.

[0040] In another aspect, provided herein is a pharmaceutical composition comprising (i) one or more isolated peptides described herein, one or more fusion proteins described herein, one or more conjugates described herein, one or more oligomeric complexes described herein, or one or more non-covalent complexes described herein, or any combination thereof; and (ii) a pharmaceutically acceptable carrier or excipient.

[0041] In some embodiments, the pharmaceutical composition can further comprise an adjuvant.

[0042] In another aspect, provided herein is an isolated molecule that binds an isolated peptide described herein, a fusion protein described herein, a conjugate described herein, an oligomeric complex described herein, or a non-covalent complex described herein.

[0043] In some embodiments, the molecule is an antibody or an antigen-binding fragment thereof.

[0044] In some embodiments, the antibody is a bispecific antibody.

[0045] In some embodiments, the molecule is an alternative scaffold.

[0046] In some embodiments, the molecule is a chimeric antigen receptor (CAR).

[0047] In some embodiments, the molecule is a T cell receptor (TCR).

[0048] In another aspect, provided herein is an isolated cell comprising a CAR described herein.

[0049] In some embodiments, the isolated cell is an immune cell.

[0050] In some embodiments, the immune cell is a T cell, an NK cell, or a macrophage.

[0051] In another aspect, provided herein is an isolated cell comprising a TCR described herein.

[0052] In some embodiments, the isolated cell is an immune cell.

[0053] In some embodiments, the immune cell is a T cell, an NK cell, or a macrophage.

[0054] In another aspect, provided herein is a pharmaceutical composition comprising (i) an isolated molecule described herein, or an isolated cell described herein; and (ii) a pharmaceutically acceptable carrier or excipient.

[0055] In another aspect, provided herein is an isolated polynucleotide comprising a nucleotide sequence encoding one or more isolated peptides described herein or a fusion protein described herein.

[0056] In some embodiments, the nucleotide sequence can be operably linked to a promoter.

[0057] In some embodiments, the isolated polynucleotide comprises DNA.

[0058] In some embodiments, the isolated polynucleotide comprises RNA.

[0059] In some embodiments, the RNA is mRNA.

[0060] In some embodiments, the RNA is self-replicating RNA.

[0061] In another aspect, provided herein is a vector comprising an isolated polynucleotide described herein.

[0062] In some embodiments, the vector is an expression vector.

[0063] In some embodiments, the vector is a viral vector.

[0064] In another aspect, provided herein is a host cell comprising an isolated polynucleotide described herein or a vector described herein.

[0065] In some embodiments, the host cell is a prokaryotic cell.

[0066] In some embodiments, the host cell is a eukaryotic cell.

[0067] In some embodiments, the host cell is an APC.

[0068] In another aspect, provided herein is a pharmaceutical composition comprising (i) an isolated polynucleotide described herein, or a vector described herein; and (ii) a pharmaceutically acceptable carrier or excipient.

[0069] In some embodiments, the pharmaceutically acceptable carrier is a lipid nanoparticle carrier.

[0070] In another aspect, provided herein is a method of inducing an immune response against HTLV-1 infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of:

a) one or more isolated peptides described herein;

b) a fusion protein described herein;

c) a conjugate described herein;

d) a oligomeric complex described herein;

e) a non-covalent complex described herein;

f) a pharmaceutical composition described herein;

g) a molecule described herein;

h) an isolated cell described herein;

i) an isolated polynucleotide described herein; or

j) a vector described herein.

[0071] In another aspect, provided herein is a method of inducing an immune response against HTLV-1 infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of one or more isolated peptides described herein.

[0072] In another aspect, provided herein is a method of inducing an immune response against a HTLV-1infection in a subject in need thereof, comprising administering to the

subject an activated T cell that is produced by contacting a T cell with an APC that presents an isolated peptide described herein in complex with an MHC molecule.

[0073] In another aspect, provided herein is a method of treating a HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of:

a) one or more isolated peptides described herein;

b) a fusion protein described herein;

c) a conjugate described herein;

d) an oligomeric complex described herein;

e) a non-covalent complex described herein;

f) a pharmaceutical composition described herein;

g) a molecule described herein;

h) an isolated cell described herein;

i) an isolated polynucleotide described herein; or

j) a vector described herein.

[0074] In another aspect, provided herein is a method of preventing or reducing the likelihood of a HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of:

a) one or more isolated peptides described herein;

b) a fusion protein described herein;

c) a conjugate described herein;

d) an oligomeric complex described herein;

e) a non-covalent complex described herein;

f) a pharmaceutical composition described herein;

g) a molecule described herein;

h) an isolated cell described herein;

i) an isolated polynucleotide described herein; or

j) a vector described herein.

[0075] In another aspect, provided herein is a method of treating a HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of one or more isolated peptides described herein.

[0076] In another aspect, provided herein is a method of preventing or reducing the likelihood of a HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of one or more isolated peptides described herein.

[0077] In some embodiments, the HTLV-1-induced disease or disorder is adult T-cell leukemia/lymphoma (ATL) or HTLV-1-myelopathy/tropical spastic paraparesis (HAM/TSP).

[0078] In another aspect, provided herein is a kit comprising:

(i) a) one or more isolated peptides described herein;

b) a fusion protein described herein;

c) a conjugate described herein;

d) an oligomeric complex described herein;

e) a non-covalent complex described herein;

f) a pharmaceutical composition described herein;

g) a molecule described herein;

h) an isolated cell described herein;

i) an isolated polynucleotide described herein; or

j) a vector described herein; and

(ii) packaging and/or instructions for use for the same.

[0079] These and other aspects of the present invention will be apparent to those of ordinary skill in the art in the following description, claims and drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0080] **Figures 1A-1E** illustrate affinity purification mass spectrometry for HTLV-transformed T-lymphoblast cell lines. An example of an anti-HLA-I W6/32 affinity purification workflow and subsequent peptide identifications by mass spectrometry is illustrated in **Figure 1A**. A description of peptide lengths and counts detected by mass spectrometry analysis for cell lines C8166 (**Figure 1B**), MT-2 (**Figure 1C**), MT-4 (**Figure 1D**), and C5/MJ (**Figure 1E**) is shown in **Figures 1B-1E**, respectively.

[0081] **Figure 2** shows that eleven (11) HTLV peptides were detected from the C8166 cell line with good reproducibility, six (6) of which were predicted HLA binders to known

C8166 HLA alleles by NetMHCpan 4.0 (see, e.g., **Table 3**). Figure discloses SEQ ID NOs: 85, 63, 74, 83, 78, 81, 51, 24, 29, 70 and 31, respectively, in order of columns.

[0082] **Figure 3** shows that thirty-two (32) HTLV peptides were detected from the MT-2 cell line, twenty-one (21) of which were found in all three replicates, and a majority of which were predicted binders to known MT-2 alleles by NetMHCpan 4.0 (see, e.g., **Table 4**). Figure discloses SEQ ID NOs: 55, 46, 86, 80, 21 and 5, respectively, in order of appearance and SEQ ID NOs: 84, 89, 88, 87, 72, 11, 9, 52, 67, 20, 44, 51, 57, 4, 45, 17, 60, 62, 54, 35, 77, 49, 76, 71, 82, and 73, respectively, in order of columns.

[0083] **Figure 4** shows that nineteen (19) HTLV peptides were detected from the MT-4 cell line, with sixteen (16) peptides overlapping between replicates. NetMHCpan 4.0 predicted a few of these peptides to be binders to common HLA alleles, e.g., HLA-A02, HLA-A11, and HLA-A24, in the patient population (see, e.g., **Table 5**). Figure discloses SEQ ID NOs: 2, 1, 65, 66, 70, 68, 18, 56, 41, 64, 59, 47, 27, 31, 26, 42, 75, 15, and 53, respectively, in order of columns.

[0084] **Figure 5** shows that HLA-I immunopeptidomics from C5/MJ cells with interferon (IFN) and without IFN (no IFN) treatment yielded nine (9) overlapping HTLV peptides, all of which were predicted binders to C5/MJ HLA alleles by NetMHCpan 4.0 (see, e.g., **Table 6**). Figure discloses SEQ ID NOs: 79, 50, 20, 58, 64, 36, 69, 9, 38, 8, 14, 89, and 37, respectively, in order of columns.

[0085] **Figures 6A-6E** show that sixteen (16) and twenty-one (21) HTLV peptides were detected from no IFN- and IFN-treated ED-41214(-) primary cells by mass spectrometry, with good reproducibility among biological replicates (see, e.g., **Figure 6B** and **Figure 6D**, respectively). Fifteen (15) HTLV peptides were shared between no IFN- and IFN-treated ED-41214(-) cells (**Figure 6E**). Only a few HTLV peptides were predicted binders to the common alleles by NetMHCpan 4.0 (see, e.g., **Table 7**). Descriptions of peptide lengths and counts detected by mass spectrometry analysis for no IFN- and IFN-treated ED-41214(-) cells are shown in **Figure 6A** and **Figure 6C**, respectively. Figure discloses SEQ ID NOs: 32, 7, 3, 25, 30, 52, 33, 61, 39, 19, 6, 28, 40, 23, 10, 43, 48, 32, 3, 7, 30, 34, 13, 52, 12, 33, 61, 39, 6, 19, 28, 40, 23, 10, 43, 16, and 22, respectively, order of appearance.

[0086] **Figures 7A-7D** illustrate peptide distributions were obtained for all HTLV primary samples tested ((see, e.g., **Figure 6A** for ED41214(-), no IFN 1; ED40515(-), no IFN 1

(**Figure 7A**); ED40515(+), no IFN 1 (**Figure 7B**); ATL43T(+), no IFN 1 (**Figure 7C**); and ATL43Tb(-) IFN 1 (**Figure 7E**)); however, HTLV peptides were detected only for ED-41214(-) (see, e.g., **Figure 6A**) (combined search against human UniProt + HTLV sequence).

[0087] **Figures 8A-8D** show examples of peptide distributions from HLA-I peptidome analysis of ATL patient samples by Thermo Orbitrap (**Figures 8A-8B**) and Bruker timsTOF SCP (**Figures 8C-8D**) mass spectrometers.

[0088] **Figures 9A-9B** show an example of a workflow used to reconstruct patient-specific HTLV genomes from RNA sequencing. RNA sequencing and RNA contig reconstruction steps are illustrated in **Figure 9A**. HTLV genome reconstruction, including BLAST analysis against the virus reference genome, viral genome reconstruction, and coding sequence extraction steps, are illustrated in **Figure 9B**.

[0089] **Figure 10** shows RNA sequencing read coverage of patient-specific reconstructed genomes.

DETAILED DESCRIPTION

[0090] The present disclosure provides, among other things, isolated peptides derived from human T-lymphotropic virus type-1 (HTLV-1), and fragments or derivatives thereof. Various peptide-based molecules including complexes (e.g., peptide-MHC (pMHC) complexes), fusion proteins, and conjugates comprising the peptides are also provided. Further provided herein are polynucleotides and vectors encoding the peptides or peptide-based molecules described herein. Binding moieties (e.g., antibodies, alternative scaffolds, T-cell receptors (TCRs) or chimeric antigen receptors (CARs)) that bind to the peptides or peptide-based molecules are also provided. The compositions of the present disclosure can be used to induce an immune response against HTLV infection (e.g., HTLV-1, HTLV-2, HTLV-3, and/or HTLV-4 infection) and/or for treating, preventing, and/or reducing the likelihood of an HTLV-induced disease or disorder (e.g., an HTLV-1-, HTLV-2-, HTLV-3-, and/or HTLV-4-induced disease or disorder). In one aspect, the present disclosure also provides methods for identifying an immunogenic virus-derived peptide.

[0091] Lymphocytes, such as T cells, play important roles in adaptive anti-infection, antitumor, autoimmune, and transplant rejection responses. Generally, a T cell mediated

immune response involves close contact, e.g., an immunological synapse, between a T cell and an antigen presenting cell (APC). The pairing of several molecules is involved in the formation of the immunological synapse, including, but not limited to: (a) a T-cell receptor (TCR) on a T cell, which specifically binds to a peptide presented in the peptide binding groove of a major histocompatibility complex (MHC) molecule on an APC; and (b) CD28 (on the T cell), which pairs with a B7 molecule on the APC. A TCR, together with CD3 molecules, form a TCR complex, and upon pairing of the TCR to the peptide-MHC (pMHC) complex, a signal is sent through CD3. Signaling through both the TCR complex and CD28 on the T cell results in activation of the T cell.

[0092] T cell receptors are heterodimeric structures composed of two types of chains (an α (alpha) and β (beta) chain, or a γ (gamma) and δ (delta) chain). The α chain is encoded by the nucleic acid sequence located within the α locus (on human or mouse chromosome 14), which also encompasses the entire δ locus that encodes the δ chain, and the β chain is encoded by the nucleic acid sequence located within the β locus (on mouse chromosome 6 or human chromosome 7). The majority of T cells have an αβ TCR, while a minority of T cells bear a γδ TCR. T cell receptor α and β polypeptides (and similarly γ and δ polypeptides) are linked to each other via a disulfide bond. Each of the two polypeptides that make up the TCR contains an extracellular domain comprising constant and variable regions, a transmembrane domain, and a cytoplasmic tail (the transmembrane domain and the cytoplasmic tail also being a part of the constant region).

[0093] The variable region of each TCR comprises a unique and characteristic structure, i.e., an idiotope or idiotype, that determines the specificity of the TCR. Generally, a TCR will bind to a pMHC complex only if the TCR comprises an idiotype that recognizes the peptide being presented in the context of MHC, e.g., the unique conformation of a particular pMHC complex.

[0094] Immunotherapeutic approaches to treating disease work to regulate T cell activity *in vivo*, e.g., to enhance anti-infection and anti-tumor responses, or, for example, to downregulate autoimmune and transplant rejection responses. However, such methods can lack specificity since immunotherapies can target signaling by the TCR complex by binding CD3 and/or the pairing of costimulatory molecules. Such approaches can result in undesirable side effects, e.g., a hyperactive immune response or generalized immune

suppression. Accordingly, therapies that take advantage of the uniquely specific interaction between a TCR and pMHC complex may provide the ability to specifically modulate the activity of specific T cells *in vivo*, and provide treatments based on T cell modulation.

[0095] The presentation of virus-derived peptides by MHC molecules on the surface of an infected cell and the recognition of these pMHC complexes by, and subsequent activation of, CD8+ cytotoxic T cells provides an important mechanism for immunity-based protection against viruses. Cells infected with HTLV-1, including those that developed into cancer cells, can express various HTLV-1-associated antigens. Peptides derived from these antigens may be displayed on the cell surface in complex with MHC molecules. Detection of an MHC-presented HTLV-1-derived peptide by a T cell bearing the corresponding TCR, leads to targeted killing of the infected cell. However, because of the selection processes which occur during T cell maturation in the thymus, there is often a scarcity of T cells in the circulating repertoire, which recognize HTLV-1-derived peptides with a sufficiently high level of affinity. As a result, infected cells often escape elimination by the immune system.

[0096] The identification of HTLV-1-derived peptides presented on infected cells (e.g., HTLV-1-induced cancer cells) can allow for the development of immunotherapeutic reagents designed to specifically target and destroy HTLV-1-infected cells (e.g., HTLV-1-induced cancer cells). Such reagents may be moieties that bind to the HTLV-1-derived peptide and/or pMHC complexes and the reagents, e.g., moieties that can function by inducing a T cell response. For example, such reagents may be based on antibodies, TCRs, and/or CARs.

[0097] The present disclosure, in part, is based on a proteogenomic approach that detects MHC-associated HTLV-1 peptides from HTLV-1 infected cells. The repertoire of HLA-I-HTLV-1 peptides that are expressed in HTLV-1-transformed cells and ATL primary cells characterized herein can provide an accurate representation of HTLV-1 epitopes in a population.

[0098] HLA-restricted viral peptides as potential targets may be leveraged for delivering immunotherapeutics (such as antibodies (e.g., bispecific antibodies), engineered TCR- or CAR- based cellular therapies) to infected tissues. However, genomic variability between strains of a virus in combination with differences in patient HLA alleles can be a challenge

in developing therapeutics against these peptide targets. To address this challenge, in one aspect the present disclosure provides a proteogenomics approach for generating patient-specific databases that allow for the comprehensive identification of viral peptides, e.g., HTLV-1-derived peptides, based on the viral transcriptomes sequenced from individual patient samples. The HTLV-1-HLA-associated peptides disclosed herein may be used in the development of immunotherapeutics (such as antibodies (e.g., bispecific antibodies), engineered TCR- or CAR- based cellular therapies) for the treatment of HTLV-1-related diseases or disorders including T-cell leukemia/lymphoma (ATL), HTLV-1-myelopathy/tropical spastic paraparesis (HAM/TSP), and hypersensitivity reactions, e.g., arthritis, uveitis, and HTLV-1-associated infective dermatitis (IDH). The proteogenomic discovery platform described herein provides a method for identifying viral-derived peptides as targets for anti-viral related immunotherapy.

Definitions

[0099] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs.

[00100] Singular forms "a", "an", and "the" include plural references unless the context clearly dictates otherwise. Thus, for example, a reference to "a method" includes one or more methods, and/or steps of the type described herein and/or which will become apparent to those persons skilled in the art upon reading this disclosure.

[00101] The term "about" or "approximately" includes being within a statistically meaningful range of a value. Such a range can be within an order of magnitude, preferably within 50%, more preferably within 20%, still more preferably within 10%, and even more preferably within 5% of a given value or range. The allowable variation encompassed by the term "about" or "approximately" depends on the particular system under study, and can be readily appreciated by one of ordinary skill in the art.

[00102] The term "antigen" encompasses any agent (e.g., protein, peptide, polysaccharide, glycoprotein, glycolipid, nucleotide, portions thereof, or combinations thereof) that, when introduced into an immunocompetent host (directly or upon expression as in, e.g., DNA or RNA vaccines) is recognized by the immune system of the host and is

capable of eliciting an immune response by the host. The T-cell receptor (TCR) recognizes a peptide presented in the context of a major histocompatibility complex (MHC) as part of an immunological synapse. The peptide-MHC (pMHC) complex is recognized by TCR, with the peptide (antigenic determinant) and the TCR idiotype providing the specificity of the interaction. Accordingly, the term "antigen" encompasses peptides presented in the context of MHCs, e.g., pMHC complexes. The peptide displayed on MHC may also be referred to as an "epitope" or an "antigenic determinant". The terms "peptide," "antigenic determinant," "epitopes," etc., encompass not only those presented naturally by antigen-presenting cells (APCs), but may be any desired peptide so long as it is recognized by an immune cell, e.g., when presented appropriately to the cells of an immune system. For example, a peptide having an artificially prepared amino acid sequence may also be used as the epitope.

[00103] A single antigen (such as an antigenic polypeptide) may have more than one epitope. Epitopes may be defined as structural or functional. Functional epitopes are generally a subset of structural epitopes and are defined as those residues that directly contribute to the affinity of the interaction between an MHC molecule and the antigen. Epitopes may also be conformational, that is, composed of non-linear amino acids. In certain embodiments, epitopes may include determinants that are chemically active surface groupings of molecules such as amino acids, sugar side chains, phosphoryl groups, or sulfonyl groups, and, in certain embodiments, may have specific three-dimensional structural characteristics, and/or specific charge characteristics. Epitopes formed from contiguous amino acids are typically retained on exposure to denaturing solvents, whereas epitopes formed by tertiary folding are typically lost on treatment with denaturing solvents.

[00104] The terms "major histocompatibility complex," and "MHC" encompass the terms "human leukocyte antigen" or "HLA" (the latter two of which are generally reserved for human MHC molecules), naturally occurring MHC molecules (e.g., MHC class I molecule comprising MHC class I α (heavy) chain and β2 microglobulin; MHC class II molecule comprising MHC class II α chain and MHC class II β chain), individual chains of MHC molecules (e.g., MHC class I α (heavy) chain, MHC class II α chain, and MHC class II β chain), individual subunits of such chains of MHC molecules (e.g., α1, α2, and/or α3 subunits of MHC class I α chain, α1-α2 subunits of MHC class II α chain, β1-β2 subunits

of MHC class II β chain) as well as portions (e.g., the peptide-binding portions, e.g., the peptide-binding grooves), mutants and various derivatives thereof (including fusions proteins), wherein such portion, mutants and derivatives retain the ability to display an antigenic peptide for recognition by a TCR, e.g., an antigen-specific TCR. An MHC class I molecule comprises a peptide binding groove formed by the α1 and α2 domains of the heavy a chain that can stow a peptide of around 8-10 amino acids. Despite the fact that both classes of MHC bind a core of about 9 amino acids (e.g., 5 to 17 amino acids) within peptides, the open-ended nature of MHC class II peptide binding groove (the α1 domain of a class II MHC α polypeptide in association with the β1 domain of a class II MHC β polypeptide) allows for a wider range of peptide lengths. Peptides binding MHC class II usually vary between 13 and 17 amino acids in length, though shorter or longer lengths are not uncommon. As a result, peptides may shift within the MHC class II peptide binding groove, changing which 9-mer sits directly within the groove at any given time. In some embodiments, the peptide-MHC complex described herein may be a peptide-MHC complex from a non-human animal. In other embodiments, the peptide-MHC complex described herein may include an peptide-HLA complex, i.e., a peptide-MHC complex from a human. Conventional identifications of particular MHC variants are used herein. For example, HLA-A11 refers to a human leucocyte antigen from the A gene group (hence a class I type MHC) gene position (known as a gene locus) number 11; gene HLA-DR11, refers to a human leucocyte antigen coded by a gene from the DR region (hence a class II type MHC) locus number 11.

[00105] "MHC-peptide complex," "peptide-MHC complex," "pMHC complex," "peptide-in-groove," and the like include (i) an MHC molecule, e.g., a human and/or non-human animal MHC molecule, or portion thereof (e.g., the peptide-binding groove thereof, and e.g., the extracellular portion thereof), and (ii) an antigenic peptide (e.g., a HTLV-1-derived peptide), where the MHC molecule and the antigenic peptide are complexed in such a manner that the pMHC complex can specifically bind a T-cell receptor. A pMHC complex encompasses cell surface expressed pMHC complexes and soluble pMHC complexes.

[00106] "HLA-peptide complex," "peptide-HLA complex," "pHLA complex," and the like refer to an MHC-peptide complex wherein the MHC molecule is a Human Leukocyte Antigen (HLA) molecule.

[00107] The term "T cell" or "T lymphocyte" is used herein in its broadest sense to refer to all types of immune cells expressing CD3, including, but not limited to, T-helper cells (CD4+ cells), cytotoxic T-cells (CD8+ cells), tumor infiltrating cytotoxic T cells (TIL; CD8+ T cell), CD4+CD8+ T cells, T-regulatory cells (Treg), and NK-T cells. T cells can include thymocytes, naïve T cells, memory T cells, immature T cells, mature T cells, resting T cells, or activated T cells. T cells may also include "gamma-delta T cells (γδ T cells)," which refer to a specialized population that to a small subset of T cells possessing a distinct TCR on their surface, and unlike the majority of T cells in which the TCR is composed of two glycoprotein chains designated α- and β-TCR chains, the TCR in γδ T cells is made up of a γ-chain and a δ-chain.

[00108] The term "antigen presenting cell" or "APC" refers to any cell that presents on the surface of the cell an antigen in association with a major histocompatibility complex molecule, either MHC class I or MHC class II molecule, or both.

[00109] Terms "antibody," "antibodies," "immunoglobulin", and the like refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that specifically binds an antigen, whether natural or partly or wholly synthetically produced. The terms include monoclonal antibodies, multispecific antibodies (e.g., bispecific antibodies), human antibodies, humanized antibodies, chimeric antibodies, single-chain Fvs (scFv), single chain antibodies, Fab fragments, F(ab′) fragments, disulfide-linked Fvs (sdFv), intrabodies, minibodies, diabodies and anti-idiotypic (anti-Id) antibodies (including, e.g., anti-Id antibodies to antigen-specific TCR), and epitope-binding fragments of any of the above. The terms "antibody" and "antibodies" also refer to covalent diabodies such as those disclosed in U.S. Pat. Appl. Pub. 2007/0004909, incorporated herein by reference in its entirety, and Ig-DARTS such as those disclosed in U.S. Pat. Appl. Pub. 2009/0060910, incorporated herein by reference in its entirety. Antibodies useful in the present disclosure include immunoglobulin molecules and immunologically active fragments of immunoglobulin molecules, i.e., molecules that contain an antigen binding site.

Immunoglobulin molecules can be of any type (e.g., IgG, IgE, IgM, IgD, IgA and IgY), class (e.g., IgG1, IgG2, IgG3, IgG4, IgA1 and IgA2) or subclass.

[00110] The term "specifically binds," "binds in a specific manner," "antigen-specific" or the like, indicates that the molecules involved in the specific binding are able to form a complex with each other that is relatively stable under physiological conditions, and are unable to form stable complexes non-specifically with other molecules outside the specified binding pair. Accordingly, a peptide binding moiety (e.g., an antibody, an alternative scaffold, a CAR, or a TCR) that binds in a specific manner to an HTLV-1-derived peptide, or a peptide-based molecule (such as a complex (e.g., a pMHC complex), fusion protein, or conjugate comprising the described peptide) indicates that the peptide binding moiety forms a stable intermolecular non-covalent bonds with the HTLV-1-derived peptide or peptide-based molecule (such as a complex (e.g., a pMHC complex), fusion protein, or conjugate comprising the described peptide). Specific binding can be characterized by an equilibrium dissociation constant (K_D) in the low micromolar to picomolar range (i.e., a smaller K_D denotes a tighter binding). High specificity may be in the low nanomolar range, with very high specificity being in the picomolar range. For example, a peptide binding moiety may exhibit binding to an HTLV-1-derived peptide or peptide-based molecule (such as a complex (e.g., a pMHC complex), fusion protein, or conjugate comprising the described peptide) with a K_D of about 3000 nM or less, about 2000 nM or less, about 1000 nM or less, about 500 nM or less, about 300 nM or less, about 200 nM or less, about 100 nM or less, about 50 nM or less, about 1 nM or less, or about 0.5 nM or less. Methods for determining whether two molecules specifically bind to one another are well known in the art and include, for example, equilibrium dialysis, surface plasmon resonance, and the like.

[00111] The terms "protein" and "polypeptide", used interchangeably herein, encompass all kinds of naturally occurring and synthetic proteins, including protein fragments of all lengths, fusion proteins and modified proteins, including without limitation, glycoproteins, as well as all other types of modified proteins (e.g., proteins resulting from phosphorylation, acetylation, myristoylation, palmitoylation, glycosylation, oxidation, formylation, amidation, polyglutamylation, ADP-ribosylation, PEGylation,

biotinylation, etc.). Small polypeptides of less than 100 amino acids, preferably less than 50 amino acids, may be referred to as "peptides".

[00112] The terms "polynucleotide" and "nucleic acid", used interchangeably herein, include polymeric forms of nucleotides of any length, including ribonucleotides (RNA), deoxyribonucleotides (DNA), or analogs or modified versions thereof. They include single-, double-, and multi-stranded DNA or RNA, genomic DNA, complementary DNA (cDNA), DNA-RNA hybrids, and polymers comprising purine bases, pyrimidine bases, or other natural, chemically modified, biochemically modified, non-natural, or derivatized nucleotide bases.

[00113] The term "operably linked" or the like refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. For example, a control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. "Operably linked" sequences include both expression control sequences that are contiguous with a gene of interest and expression control sequences that act in trans or at a distance to control a gene of interest (or sequence of interest). The term "expression control sequence" includes polynucleotide sequences, which are necessary to affect the expression and processing of coding sequences to which they are ligated. "Expression control sequences" include: appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (i.e., Kozak consensus sequence); sequences that enhance polypeptide stability; and when desired, sequences that enhance polypeptide secretion. The nature of such control sequences differs depending upon the host organism. For example, in prokaryotes, such control sequences generally include promoter, ribosomal binding site and transcription termination sequence, while in eukaryotes typically such control sequences include promoters and transcription termination sequence. The term "control sequences" is intended to include components whose presence is essential for expression and processing and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

[00114] The term "isolated" refers to a homogenous population of molecules (such as polynucleotides or polypeptides) which have been substantially separated and/or purified away from other components of the system the molecules are produced in, such as a recombinant cell, as well as a protein that has been subjected to at least one purification or isolation step. "Isolated" refers to a molecule that is substantially free of other cellular material and/or chemicals and encompasses molecules that are isolated to a higher purity, such as to 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% purity.

[00115] The term "derivative" as used herein refers to a peptide, polypeptide, or polynucleotide, or a variant or analog thereof, comprising one or more mutations and/or chemical modifications as compared to a reference peptide, polypeptide or polynucleotide. Mutations and/or chemical modifications are further detailed below and can include, for example, insertions, substitutions, deletions, transversions, and/or inversions at one or more locations in the amino acid or nucleotide sequence.

[00116] The terms "treat" or "treatment" of a state, disorder, disease, or condition include: (1) preventing, delaying, or reducing the incidence and/or likelihood of the appearance of at least one clinical or sub-clinical symptom of the state, disorder, disease, or condition developing in a subject that may be afflicted with or predisposed to the state, disorder, disease, or condition, but does not yet experience or display clinical or subclinical symptoms of the state, disorder, disease, or condition; or (2) inhibiting the state, disorder, disease, or condition, i.e., arresting, reducing or delaying the development of the disease or a relapse thereof or at least one clinical or sub-clinical symptom thereof; or (3) relieving the state, disorder, disease, or condition, i.e., causing regression of the state, disorder, disease, or condition or at least one of the clinical or sub-clinical symptoms of the state, disorder, disease, or condition. The benefit to a subject to be treated is either statistically significant or at least perceptible to the patient or to the physician.

[00117] An "individual" or "subject" or "animal" refers to humans, veterinary animals (e.g., cats, dogs, cows, horses, sheep, pigs, etc.) and experimental animal models of diseases (e.g., mice, rats). In a preferred embodiment, the subject is a human.

[00118] The term "effective" applied to dose or amount refers to that quantity of a compound or pharmaceutical composition that is sufficient to result in a desired activity

upon administration to a subject in need thereof. Note that when a combination of active ingredients is administered, the effective amount of the combination may or may not include amounts of each ingredient that would have been effective if administered individually. The exact amount required will vary from subject to subject, depending on the species, age, and general condition of the subject, the severity of the condition being treated, the particular drug or drugs employed, the mode of administration, and the like.

[00119] The phrase "pharmaceutically acceptable", as used in connection with compositions described herein, refers to molecular entities and other ingredients of such compositions that are physiologically tolerable and do not typically produce untoward reactions when administered to a mammal (e.g., a human). Preferably, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in mammals, and more particularly in humans.

[00120] The term "administration" and the like refers to and includes the administration of a composition to a subject or system (e.g., to a cell, organ, tissue, organism, or relevant component or set of components thereof). The skilled artisan will appreciate that route of administration may vary depending, for example, on the subject or system to which the composition is being administered, the nature of the composition, the purpose of the administration, etc. For example, in certain embodiments, administration to an animal subject (e.g., to a human or a rodent) may be bronchial (including by bronchial instillation), buccal, enteral, interdermal, intra-arterial, intradermal, intragastric, intramedullary, intramuscular, intranasal, intraperitoneal, intrathecal, intravenous, intraventricular, mucosal, nasal, oral, rectal, subcutaneous, sublingual, topical, tracheal (including by intratracheal instillation), transdermal, vaginal and/or vitreal. In some embodiments, administration may involve intermittent dosing. In some embodiments, administration may involve continuous dosing (e.g., perfusion) for at least a selected period of time.

[00121] In accordance with the disclosure herein, there may be employed conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature. See, e.g., Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition.

Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989 (herein "Sambrook et al., 1989"); DNA Cloning: A Practical Approach, Volumes I and II (D.N. Glover ed. 1985); Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization [B.D. Hames & S.J. Higgins eds. (1985)]; Transcription And Translation [B.D. Hames & S.J. Higgins, eds. (1984)]; Animal Cell Culture [R.I. Freshney, ed. (1986)]; Immobilized Cells And Enzymes [IRL Press, (1986)]; B. Perbal, A Practical Guide To Molecular Cloning (1984); Ausubel, F.M. et al. (eds.). Current Protocols in Molecular Biology. John Wiley & Sons, Inc., 1994. These techniques include site directed mutagenesis as described in Kunkel, Proc. Natl. Acad. Sci. USA 82: 488- 492 (1985), U. S. Patent No. 5,071, 743, Fukuoka et al., Biochem. Biophys. Res. Commun. 263: 357-360 (1999); Kim and Maas, BioTech. 28: 196-198 (2000); Parikh and Guengerich, BioTech. 24: 4 28-431 (1998); Ray and Nickoloff, BioTech. 13: 342-346 (1992); Wang et al., BioTech. 19: 556-559 (1995); Wang and Malcolm, BioTech. 26: 680-682 (1999); Xu and Gong, BioTech. 26: 639-641 (1999), U.S. Patents Nos. 5,789, 166 and 5,932, 419, Hogrefe, Strategies l4. 3: 74-75 (2001), U. S. Patents Nos. 5,702,931, 5,780,270, and 6,242,222, Angag and Schutz, Biotech. 30: 486-488 (2001), Wang and Wilkinson, Biotech. 29: 976-978 (2000), Kang et al., Biotech. 20: 44-46 (1996), Ogel and McPherson, Protein Engineer. 5: 467-468 (1992), Kirsch and Joly, Nucl. Acids. Res. 26: 1848-1850 (1998), Rhem and Hancock, J. Bacteriol. 178: 3346-3349 (1996), Boles and Miogsa, Curr. Genet. 28: 197-198 (1995), Barrenttino et al., Nuc. Acids. Res. 22: 541-542 (1993), Tessier and Thomas, Meths. Molec. Biol. 57: 229-237, and Pons et al., Meth. Molec. Biol. 67: 209-218.

Peptides Disclosed Herein

[00122] In one aspect, the present disclosure provides isolated peptides comprising an amino acid sequence derived from human T-lymphotropic virus type-1 (HTLV-1).

[00123] The HTLV genome follows a canonical structure of replication competent retroviruses comprising GAG, POL, and ENV domains flanked by two long terminal repeat (LTR) domains on either end of the provirus. In particular, HTLV-1 has a monopartite, linear, dimeric single strand RNA (+) genome of 8.5 kb, comprising a 5'-cap and a 3' poly-A tail. The two LTRs are about 600 nucleotide residues in length and are located at the 5' and 3' ends of the genome. The LTRs comprise U3, R, and U5 regions, which together can

function as a transcription unit. The 5' end of the genome further comprises a primer binding site and the 3' end further comprises a polypurin tract. The integrated virus utilizes promoter elements in the 5' LTR to trigger and drive transcription, resulting in an unspliced full-length mRNA that can serve, e.g., as genomic RNA for packaging into virions. While similar to other retroviruses, the genomic RNA of HTLV-1 encodes the structural and enzymatic proteins Gag (an internal group specific antigen), Pol (a polymerase, reverse transcriptase and integrase enzymes required for viral replication and maturation), and Env (envelope glycoprotein), HTLV-1 differs from other retroviruses due to a pX region located toward the 3' end between the ENV and 3' LTR. The pX region encodes several alternatively spliced regulatory genes including, without limitation, Tax and HTLV-I basic leucine zipper factor (HBZ), both of which can be implicated in viral pathogenesis.

[00124] The viral gene product Tax is a 40 kDa protein which mainly localizes to the nucleus but can also be within the cytoplasm of infected cells. Tax interacts with a variety of host proteins and plays a key role in transactivating proviral transcription from the 5' LTR. Tax can functionally inactivate p53 and target pRB for degradation, allowing for survival of infected cells. Tax can also dysregulate several signaling pathways such as but not limited to activator protein 1 (AP-1), nuclear factor kappa B (NF-κB), serum responsive factor (SRF), and cyclic AMP response element-binding protein (CREB) pathways. The pleiotropic functions of Tax can contribute to viral pathogenicity and transformation of HTLV-1 infected cells.

[00125] HBZ is a nuclear protein but can also be localized to the cytoplasm. HBZ comprises three domains: an activation domain, a central domain, and a basic leucine zipper domain. HBZ antagonizes several Tax-mediated functions and participates in viral persistence and immune evasion. The activation domain of HBZ comprises two LXXLL -like motifs (where L is leucine and X is any amino acid) which bind can bind to a kinase-inducible domain interacting (KIX) domain of p300-CBP transcription coactivators. The LXXLL-like motifs and KIX domain can be a requisite for HBZ activation of TGF-β (Transforming growth factor beta)/Smad signaling.

[00126] After infection of a host, HTLV-1 infects cells via its glycoprotein gp62. HTLV-1 can then achieve latency by integrating into the host genome and increasing proviral load by proliferation of infected cells.

[00127] The HTLV-1 life cycle initiates when the HTLV-1 virion attaches to a host cell such as by way of a surface receptor(s), e.g., GLUT1/ HSPG/NRP-1, via the HTLV-1 envelope surface and/or transmembrane domains of the Env protein. After the outer membrane of the viral particle fuses with the host cell membrane, the contents of the virions can enter the cytoplasm of the host cell and the viral genomic RNA (gRNA) can undergo reverse transcription to convert gRNA into double stranded DNA (dsDNA) using the virally encoded reverse transcriptase (RNA-dependent/DNA-dependent polymerase). Viral dsDNA traffics to the nucleus and stochastically integrates into the host genome using a viral encoded integrase at which point the viral genome is now a provirus. The host cell machinery can then drive transcription of the proviral DNA using host cell RNA polymerase II to produce viral mRNA. The unspliced full-length viral genomic RNA (~9 kb) can be used to produce progeny virions and to produce, e.g., Gag protein, Pol protein, and Pro protein (a protease enzyme required for maturation of structural and enzymatic proteins). The singly spliced subgenomic mRNA (~4.3 kb) can be used to produce Env protein, and the doubly-spliced RNA (~ 2.1 kb) can be used to translate the regulatory proteins Tax (open reading frame (ORF) IV) and Rex (ORF III). Alternative splicing can also lead to the production of a number of accessory proteins from pX ORF I and II. Immature virions can be assembled at the plasma membrane and are considered mature following viral budding.

[00128] Amino acid sequences derived from HTLV-1 disclosed herein may include naturally occurring proteogenic amino acids as well as non-proteogenic amino acids and non-naturally occurring amino acids such as amino acid analogs. In some embodiments, the amino acids that may be used in the practice of the present disclosure may include, for example, without limitation, naturally occurring proteogenic (L)-amino acids, their optical (D)-isomers, chemically modified amino acids, including, e.g., amino acid analogs such as, e.g., selenocysteine (Sec), penicillamine (3-mercapto-D-valine), pyroglutamic acid (5-oxoproline), etc., naturally occurring non-proteogenic amino acids such as norleucine, and chemically synthesized amino acids that have properties known in the art to be characteristic of an amino acid, and amino acid equivalents.

[00129] In some embodiments, an isolated peptide of the present disclosure comprises an amino acid sequence that is at least about 60%, at least about 65%, at least

about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% identical to the amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof. In some embodiments, an isolated peptide of the present disclosure comprises an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146. In some embodiments, an isolated peptide of the present disclosure consists essentially of an amino acid sequence of any of SEQ ID NOs: 1-89 and 143-146. In some embodiments, the isolated peptide consists of an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146. In some embodiments, the isolated peptide comprises two or more sequences selected from any one of SEQ ID NOs: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof.

[00130] A list of non-limiting examples of HTLV-1-derived peptides are provided in **Table 1** below.

Table 1. Examples of HTLV-1-derived Peptides

SEQ ID NO:	Peptide Sequence
1	ATMASLISH
2	LPFRTTPIVL
3	QTQAQIPSR
4	RQWKGPQEAL
5	WAIIGRDAL
6	YFKLPGLNSR
7	AETRGITGY
8	ATCPEHQITW
9	DPASILPVI
10	DPIDGRVIGSA
11	DPPDSDPQI
12	DTFSGAISA
13	EAISSLLQA
14	FPDPGLRPQNL
15	GPSSYDFHQLK
16	GYNPLAGPL
17	GYPGRVNEIL
18	IPISLLFNE
19	IPRPPRGLAA
20	ITNSHVSIL
21	KECQKLLQA
22	KNTSVLGAGGQ

23	KTRWQLHHSPR
24	LENRVLTGW
25	LNFLQAAYR
26	LPAPHLTLPF
27	LPEDCLPTTL
28	LPMDNALSIA
29	LSFPDPGLRP
30	MGQIFSRSA
31	NVPYKRIEEL
32	PMDNALSIAL
33	PRPPRGLAA
34	QAQIPSRPA
35	QFLIPRLPSF
36	RLHVWVDTF
37	RSRRAVPVAVW
38	RVLTGWGLNW
39	SASPIPRPPR
40	SEAETRGITGY
41	SKCGFPFSL
42	SLYLFPHWIKK
43	SPIPRPPRGLAA
44	SYSSYHATY
45	TEASNILRSC
46	VSSPCHNSL
47	YPVYVFGDCV
48	LPVMHPHGA
49	AAHHWLNFL
50	EALQTGITL
51	EELLYKISL
52	GYPGRVNEI
53	LPAPHLTLP
54	LYQLSPPITW
55	WQLHHSPRL
56	YLFPHWIKK
57	YQLSPPITW
58	VPSSSSTPL
59	GKFLATLIL
60	HEVDKDISQL
61	LPTTLFQPA
62	QEVSRLNINL
63	SARLHRHAL

64	SPCHNSLIL
65	VVCMYLYQL
66	KVLTPPITH
67	LPMDNALSI
68	IQYSSFHSL
69	RVNEILHIL
70	VPYKRIEEL
71	AGPLRVQANNPQQ
72	ALPELQAL
73	EELLYKIS
74	ELLYKISL
75	EPGPSSYDFHQLK
76	GPLRVQANNPQQQ
77	HQLKKFLKIALETP
78	LLPFHSTL
79	LPAPHLTL
80	LPFRTTPI
81	LSFPDPGLRPQNLY
82	NANKECQKLLQARG
83	PYKRIEEL
84	REYQQLWL
85	RLHRHALL
86	RQLRHLPSRVRYPH
87	TAPQVLPV
88	VERLNIAL
89	YPGRVNEI
143	QSSSFIFHK
144	EYTNIPISLL
145	EYTNIPISL
146	LTAFAALPGS

[00131] A peptide of the disclosure may be synthetically produced or produced by hydrolysis. Synthetically produced peptides can include randomly generated peptides, specifically designed peptides, and peptides where at least some of the amino acid positions are conserved among several peptides and the remaining positions are random. Alternatively, a peptide of the present disclosure may be produced by expression in a heterologous host cell.

[00132] In nature, peptides that are produced by hydrolysis undergo hydrolysis prior to binding of the antigen to an MHC molecule. Class I MHC typically present peptides

derived from proteins actively synthesized in the cytoplasm of the cell. In contrast, class II MHC typically present peptides derived either from exogenous proteins that enter a cell's endocytic pathway or from proteins synthesized in the endoplasmic reticulum (ER). Intracellular trafficking permits a peptide to become associated with an MHC molecule.

[00133] The binding of a peptide to an MHC peptide binding groove can control the spatial arrangement of MHC and/or peptide amino acid residues recognized by a TCR. Such spatial control is due in part to hydrogen bonds formed between a peptide and an MHC molecule. Based on the knowledge on how peptides bind to various MHC molecules, the major MHC anchor amino acids and the surface exposed amino acids that are varied among different peptides can be determined.

[00134] Preferably, the length of an MHC-binding peptide is from about 5 to about 40 amino acid residues, more preferably from about 6 to about 30 amino acid residues, and even more preferably from about 8 to about 20 amino acid residues, and even more preferably between about 9 and 11 amino acid residues, including any size peptide between 5 and 40 amino acids in length, in whole integer increments (i.e., 5, 6, 7, 8, 9 . . . 40). While naturally MHC class II-bound peptides vary from about 9-40 amino acids, in nearly all cases the peptide can be truncated to an about 9-11 amino acid core without loss of MHC binding activity or T cell recognition.

[00135] In some embodiments, the isolated peptides of the disclosure may be about 8-12 amino acids in length. For example, a peptide disclosed herein may be 8 amino acids, 9 amino acids, 10 amino acids, 11 amino acids, or 12 amino acids in length.

[00136] The peptides of the disclosure may comprise one or more reverse peptide bonds, one or more non-peptide bonds, one or more chemical modifications, one or more D-isomers of amino acids, or any combination thereof.

[00137] In some embodiments, the peptide may be modified to comprise one or more reverse peptide bonds or non-peptide bonds. Such modification may improve stability and/or binding of the peptide to MHC molecules to elicit a stronger immune response. In a reverse peptide bond, amino acid residues are not joined by peptide (—CO—NH—) linkages but the peptide bond is reversed. Such retro-inverso peptidomimetics may be made using methods known in the art, for example such as those described in Meziere et al., 1997 (Meziere C., et al. J Immunol 1997). This approach involves making

pseudopeptides containing changes involving the backbone, and not the orientation of side chains. Such pseudopeptides may be useful, for example, for desired MHC binding and/or T helper cell responses. Retro-inverse peptides, which contain NH—CO bonds instead of CO—NH peptide bonds, are much more resistant to proteolysis. Additional non-peptide bond that may be used are, for example, —CH_2—NH, —CH_2S—, —CH_2CH_2—, —CH=CH—, —$COCH_2$—, —$CH(OH)CH_2$—, and —CH_2SO—.

[00138] The amino acid residues comprising the peptides of the disclosure may be chemically modified. Non-limiting examples of chemical modifications include, for example, phosphorylation, acetylation, deamidation acylation, amidination, pyridoxylation of lysine, reductive alkylation, trinitrobenzylation of amino groups with 2,4,6-trinitrobenzene sulphonic acid (TNBS), amide modification of carboxyl groups and sulphydryl modification by performic acid oxidation of cysteine to cysteic acid, formation of mercurial derivatives, formation of mixed disulfides with other thiol compounds, reaction with maleimide, carboxymethylation with iodoacetic acid or iodoacetamide and carbamoylation with cyanate at alkaline pH. Chemical modifications may not correspond to those that may be present *in vivo*.

[00139] For example, modification of, for example, arginyl residues in proteins may be based on the reaction of vicinal dicarbonyl compounds such as phenylglyoxal, 2,3-butanedione, and 1,2-cyclohexanedione to form an adduct. Another example is the reaction of methylglyoxal with arginine residues. Cysteine can be modified without concomitant modification of other nucleophilic sites such as lysine and histidine. Selective reduction of disulfide bonds in proteins can also be performed. Disulfide bonds can be formed and oxidized during the heat treatment of biopharmaceuticals. Woodward's Reagent K may be used to modify specific glutamic acid residues. N-(3-(dimethylamino)propyl)-N′-ethylcarbodiimide can be used to form intra-molecular crosslinks between a lysine residue and a glutamic acid residue. For example, diethylpyrocarbonate and 4-hydroxy-2-nonenal can be used to modify histidyl residues in proteins. The reaction of lysine residues and other α-amino groups is, for example, useful in binding of peptides to surfaces or the cross-linking of proteins/peptides. Lysine is the site of attachment of poly(ethylene)glycol and the major site of modification in the glycosylation of proteins. Methionine residues in proteins can be modified with e.g. iodoacetamide, bromoethylamine, and chloramine T.

Tetranitromethane and N-acetylimidazole can be used for the modification of tyrosyl residues. Cross-linking via the formation of dityrosine can be accomplished with hydrogen peroxide/copper ions. N-bromosuccinimide, 2-hydroxy-5-nitrobenzyl bromide or 3-bromo-3-methyl-2-(2-nitrophenylmercapto)-3H-indole (BPNS-skatole) have been used in recent studies for the modification of tryptophan. Successful modification of therapeutic proteins and peptides with PEG can lead to an extension of circulatory half-life while cross-linking of proteins/ peptides with glutaraldehyde, polyethylene glycol diacrylate and formaldehyde can be used for the preparation of hydrogels. Chemical modification of allergens for immunotherapy can be achieved by carbamylation with potassium cyanate.

[00140] Peptides of the present disclosure may also be synthesized with additional chemical groups present at their N- and/or C-termini, to enhance the stability, bioavailability, and/or affinity of the peptides.

[00141] N-terminal modifications can include methylation (e.g., —$NHCH_3$ or —$N(CH_3)_2$), acetylation (e.g., with acetic acid or a halogenated derivative thereof such as α-chloroacetic acid, α-bromoacetic acid, or α-iodoacetic acid), adding a benzyloxycarbonyl (Cbz) group, or blocking the amino terminus with any blocking group containing a carboxylate functionality defined by RCOO— or sulfonyl functionality defined by R—SO_2—, where R is selected from alkyl, aryl, heteroaryl, alkyl aryl, and the like, and similar groups. One can also incorporate a desamino acid at the N-terminus (so that there is no N-terminal amino group) to decrease susceptibility to proteases or to restrict the conformation of the peptide. Additionally, hydrophobic groups such as carbobenzoxyl, dansyl, or t-butyloxycarbonyl groups may be added to the N-terminus. Likewise, an acetyl group or a 9-fluorenylmethoxy-carbonyl group may be placed at the N-terminus.

[00142] C-terminal modifications can include replacing the free acid with a carboxamide group or forming a cyclic lactam at the carboxy terminus to introduce structural constraints. One can also cyclize the peptides of the disclosure, or incorporate a desamino or descarboxy residue at the termini of the peptide, so that there is no terminal amino or carboxyl group, to decrease susceptibility to proteases or to restrict the conformation of the peptide. C-terminal functional groups of the compounds of the present disclosure include amide, amide lower alkyl, amide di(lower alkyl), lower alkoxy, hydroxy, and carboxy, and the lower ester derivatives thereof, and the pharmaceutically acceptable

salts thereof. Additionally, the hydrophobic group, t-butyloxycarbonyl, or an amido group may be added to the C-terminus.

[00143] Further examples of non-natural modifications include incorporation of non-encoded α-amino acids, photoreactive cross-linking amino acids, N-methylated amino acids, and β-amino acids, backbone reduction, retroinversion by using D-amino acids, and C-terminal amidation and PEGylation.

[00144] Peptides described herein may comprise one or more (e.g., 1, 2, 3, or 4) amino acid substitutions and/or insertions and/or deletions. Amino acid substitution means that an amino acid residue is substituted for a replacement amino acid residue at the same position. Inserted amino acid residues may be inserted at any position and may be inserted such that some or all of the inserted amino acid residues are immediately adjacent one another or may be inserted such that none of the inserted amino acid residues is immediately adjacent another inserted amino acid residue. One or more (e.g., 1, 2, 3 or 4) amino acids may be substituted and/or inserted and/or deleted from the sequence of any one of SEQ ID NOs: 1-89 and 143-146. Each substitution and/or insertion and/or deletion can take place at any position of any one of SEQ ID NOs: 1-89 and 143-146.

[00145] In some embodiments, the peptides of the disclosure may comprise additional amino acids (e.g., 1, 2, 3 or 4) at the C-terminal end and/or at the N-terminal end of the sequence of any one SEQ ID NOs: 1-89 and 143-146. A peptide of the disclosure may comprise the amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146 except for one or more (e.g., 1, 2, 3, or 4) amino acid substitutions, insertions or deletions.

[00146] Inserted amino acids and substituted amino acids may be naturally occurring amino acids or may be non-naturally occurring amino acids and, for example, may contain a non-natural side chain, and/or be linked together via non-native peptide bonds. Such altered peptide ligands are discussed further in Douat-Casassus et al., J. Med. Chem, 2007; 50(7):1598-609 and Hoppes et al., J. Immunol 2014; 193(10):4803-13 and references therein. If more than one amino acid residue is substituted and/or inserted, the replacement/inserted amino acid residues may be the same as each other or different from one another. Each replacement amino acid may have a different side chain to the amino acid being replaced.

[00147] D-amino acids may be substituted for the L-amino acids in the antigenic peptides of the disclosure. In addition, non-standard amino acids (i.e., other than the common naturally occurring proteinogenic amino acids such as β-γ-δ-amino acids, as well as many derivatives of L-α-amino acids) may also be used for substitutions or additions to produce peptides of the present disclosure.

[00148] Amino acid substitutions may be conservative, by which it is meant the substituted amino acid has similar chemical properties to the original amino acid. For example, the following groups of amino acids share similar chemical properties such as size, charge, and polarity: Group 1 - Ala, Ser, Thr, Pro, Gly; Group 2 - Asp, Asn, Glu, Gln; Group 3 - His, Arg, Lys; Group 4 - Met, Leu, Ile, Val, Cys; Group 5 - Phe, Thy, Trp.

[00149] Substantial changes in function (e.g., affinity for MHC molecules and/or TCRs) can be made by selecting substitutions that are less conservative than those described above, in other words, selecting residues that differ more significantly in their effect on maintaining the structure of the peptide backbone in the area of the substitution (e.g., as a sheet or helical conformation), the bulk of the side chain, or the charge or hydrophobicity of the peptide at the positions involved for MHC or TCR binding. The substitutions which in general are expected to produce the greatest changes in peptide properties will be those in which (a) a hydrophilic residue, e.g. Ser, is substituted for (or by) a hydrophobic residue, e.g. Leu, Ile, Phe, Val or Ala; (b) a residue having an electropositive side chain, e.g., Lys, Arg, or His, is substituted for (or by) an electronegative residue, e.g. Glu or Asp; or (c) a residue having a bulky side chain, e.g. Phe, is substituted for (or by) a residue not having a side chain, e.g., Gly.

[00150] One can also replace the naturally occurring side chains of the 20 genetically encoded amino acids (or the stereoisomeric D-amino acids) with other side chains, for instance with groups such as alkyl, lower alkyl, cyclic 4-, 5-, 6-, to 7-membered alkyl, amide, amide lower alkyl, amide di(lower alkyl), lower alkoxy, hydroxy, carboxy and the lower ester derivatives thereof, and with 4-, 5-, 6-, to 7-membered heterocyclic. For example, proline analogues in which the ring size of the proline residue is changed from 5 members to 4, 6, or 7 members can be employed. Cyclic groups can be saturated or unsaturated, and if unsaturated, can be aromatic or non-aromatic. Heterocyclic groups preferably contain one or more nitrogen, oxygen, and/or sulfur heteroatoms. Examples of

such groups include the furazanyl, furyl, imidazolidinyl, imidazolyl, imidazolinyl, isothiazolyl, isoxazolyl, morpholinyl (e.g. morpholino), oxazolyl, piperazinyl (*e.g.*, 1-piperazinyl), piperidyl (e.g., 1-piperidyl, piperidino), pyranyl, pyrazinyl, pyrazolidinyl, pyrazolinyl, pyrazolyl, pyridazinyl, pyridyl, pyrimidinyl, pyrrolidinyl (e.g., 1-pyrrolidinyl), pyrrolinyl, pyrrolyl, thiadiazolyl, thiazolyl, thienyl, thiomorpholinyl (e.g., thiomorpholino), and triazolyl. These heterocyclic groups can be substituted or unsubstituted. Where a group is substituted, the substituent can be alkyl, alkoxy, halogen, oxygen, or substituted or unsubstituted phenyl.

[00151] Other examples of amino acid replacements include stereoisomers (e.g., D-amino acids) and unnatural amino acids such as, for example, L-ornithine, L-homocysteine, L-homoserine, L-citrulline, 3-sulfino-L-alanine, N-(L-arginino)succinate, 3,4-dihydroxy-L-phenylalanine, 3-iodo-L-tyrosine, 3,5-diiodo-L-tyrosine, triiodothyronine, L-thyroxine, L-selenocysteine, N-(L-arginino)taurine, 4-aminobutylate, (R,S)-3-amino-2-methylpropanoate, a,a-disubstituted amino acids, N-alkyl amino acids, lactic acid, β-alanine, 3-pyridylalanine, 4-hydroxyproline, O-phosphoserine, N-methylglycine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, nor-leucine, and other similar amino acids and imino acids.

[00152] The amino acid residues that do not substantially contribute to interactions with the T-cell receptor can be modified by replacement with other amino acid whose incorporation does not substantially affect T-cell reactivity and does not eliminate binding to the relevant MHC.

[00153] The peptides may also comprise isosteres of two or more residues. An "isostere" as used here refers to a sequence of two or more residues that can be substituted for a second sequence because the steric conformation of the first sequence fits a binding site specific for the second sequence. The term specifically includes peptide backbone modifications well known to those skilled in the art. Such modifications include modifications of the amide nitrogen, the α-carbon, amide carbonyl, complete replacement of the amide bond, extensions, deletions or backbone crosslinks.

[00154] Combinations of several substitutions/additions/deletions at more than one position can be developed and tested to determine if the combination results in an additive

or synergistic effects on the immunogenicity of the peptide. In some embodiments, no more than 4 positions within the peptide are simultaneously altered.

[00155] Preferably, peptides of the disclosure bind to an MHC molecule in the peptide binding groove of the MHC molecule. Generally, the amino acid modifications described above will not impair the ability of the peptide to bind to the MHC molecule. In some embodiments, the amino acid modifications improve the ability of the peptide to bind to the MHC molecule. For example, mutations may be made at positions which anchor the peptide to the MHC molecule. Such anchor positions and the preferred residues at these locations for peptides which bind, in particular, HLA-A*02 may comprise, e.g., amino acids residues at position 2, and/or at the C-terminus of the peptide, which may be considered primary anchor positions. Preferred anchor residues may be different for each HLA type. As a non-limiting example, the preferred amino acids in position 2 for HLA-A*02 are Leu, lie, Val, or Met and at the C-terminus are Val or Leu. Multiple positions may be important for stable peptide binding to HLA-A*02, including positions 2, 3, 5-7, and 9. The anchor residues at position 2 and 9 may be of prime importance for peptide binding to HLA-A2. However, other peptide side chains, e.g., at position 3, may contribute to the stability of the interaction. In certain cases, the optimal length for peptide binding can be longer than 9 residues.

[00156] The immunologic properties of peptides can be described as a function of binding to MHC molecules (K_{on} and K_{off}) and TCR (affinity of interaction between TCR and MHC-peptide complexes). Modifications of primary MHC anchor residues exhibit a significant degree of predictability about overall impact on binding to MHC molecules. Modifications of secondary MHC anchor residues can impact the affinity of interaction of the MHC-peptide complex to TCR as well as with the K_{on} and K_{off} relative to peptide-MHC interaction.

[00157] When the HTLV-1 peptide is a mutant peptide, T cell lines against a natural (non-mutated) epitope are generated, and an immunization strategy potent enough to generate a useful response in transgenic mice carrying human MHC (such as the A2 allele) is used. HTLV-1 peptides are interrogated *ex vivo* in the presence of competent APCs and the functional impact of T cells specific for natural (non-mutated) epitopes is measured. The evaluation is done at various concentrations of HTLV-1 peptide because the expected

effect is biphasic in the instance of cross reactive peptides (activating at limited concentrations and inhibiting at higher concentrations due to antigen-induced cell death [AICD]). Measurement of the following three parameters can be used to describe characteristics of the disclosed HTLV-1 peptides:

[00158] 1. Minimal required concentration of HTLV-1 metrics to induce effects indicative of T cell activation (e.g., cytokine [e.g., IFN-γ] production);

2. Maximal (peak value) effect (e.g., cytokine [e.g., IFN-γ] production) at any HTLV-1 peptide concentration; and

3. HTLV-1 peptide concentration at peak value of activating effect (e.g., cytokine [e.g., IFN-γ] concentration).

[00159] By way of a non-limiting example, HTLV-1 peptides that result in reduced values associated with parameters number 1 and number 3, but increased number 2, can be useful. Use of natural epitope and/or unrelated non-cross-reactive peptide as references is valuable for identifying classes of peptides of possible value. Peptides possessing properties quantitatively comparable to or even moderately attenuated from those of natural epitopes are still considered useful since, while they retain cross-reactivity, they may exhibit immunologic properties that are distinct from those of the natural peptide, e.g., reduced capacity to break tolerance or reestablish responsiveness *in vivo* or lower propensity to induce AICD.

[00160] In addition to practicality and rapidity, additional advantages of this screening approach include, but are not limited to, use of more relevant polyclonal T cell lines instead of potentially biased T cell clones as a read out, and the composite value, integrating parameters such as K_{on}, K_{off}, and TCR affinity that can translate into cross-reactivity and functional avidity of peptide-MHC complexes relative to TCR. These parameters can be predictive of the in vivo immunologic properties and thus can define useful panels of peptides eligible for further evaluation, optimization, and practical applications. Peptides that bind to MHC and retain cross-reactivity against TCR specific for the nominal wild-type peptide are predicted to elicit a measurable effect in this assay.

[00161] A peptide of the disclosure, or a pharmaceutically acceptable salt thereof, or fragment or derivative thereof, may be used to induce an immune response. If this is the case, it is important that the immune response is specific to the intended target (e.g., HTLV-

1, HTLV-2, HTLV-3, and/or HTLV-4) to avoid the risk of unwanted side effects that may be associated with an "off target" immune response. Therefore, it is preferred that the amino acid sequence of a peptide of the disclosure does not match the amino acid sequence of a peptide from any other endogenous protein(s), particularly that of another human protein. Also, the amino acid modifications described herein should not impair the ability of the peptide inducing an antigen-specific immune response when presented in a complex with an MHC molecule on the surface of an antigen presenting cell (APC).

[00162] The peptides may be also modified to improve half-life and/or bioavailability, for example, by PEGylation, glycosylation, polysialylation, HESylation, recombinant PEG mimetics, Fc fusion, albumin fusion, nanoparticle attachment, nanoparticulate encapsulation, cholesterol fusion, iron fusion, or acylation.

[00163] The peptides of the disclosure can also serve as structural models for non-peptidic compounds with similar biological activity. A variety of techniques can be used to construct compounds with the same or similar desired biological activity as the lead peptide compound, but with more favorable activity than the lead with respect to solubility, stability, and susceptibility to hydrolysis and proteolysis. These techniques include replacing the peptide backbone with a backbone composed of amidates, phosphonates, carbamates, sulfonamides, secondary amines, and N-methylamino acids.

[00164] Multiple peptides described herein may be operably linked together. Accordingly, in one aspect, the present disclosure provides an isolated peptide or polypeptide comprising two or more amino acid sequences selected from SEQ ID NO: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof. For example, such a multi-epitope peptide or polypeptide may comprise include 2 to 50, 2 to 40, 2 to 30, 5 to 25, 5 to 20, or 10 to 15 single-epitope peptides as described herein (e.g., SEQ ID NO: 1-89 and 143-146). The single-epitope peptides (e.g., SEQ ID NO: 1-89 and 143-146) may be arranged in any order, and may be identical or different.

[00165] In some embodiments, the present disclosure provides an isolated peptide or polypeptide comprising 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,

or 93 amino acid sequences selected from SEQ ID NO: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof.

[00166] The single-epitope peptides may be linked via a linker. The linker may comprise relatively small, neutral molecules, such as amino acids or amino acid mimetics, which are substantially uncharged under physiological conditions. The linker can be selected from, e.g., Table 2, or other neutral spacers of nonpolar amino acids or neutral polar amino acids. It will be understood that the optionally present linker need not be comprised of the same residues and thus may be a hetero- or homo-oligomer. When present, the linker will usually be at least one or two residues, more usually three to six residues.

[00167] Peptides of the disclosure can be synthesized by e.g., solid phase synthesis. As such, the peptides may be immobilized, for example to a solid support such as a bead. Peptides of the disclosure may be synthesized by the Fmoc-polyamide mode of solid-phase peptide synthesis. Temporary N-amino group protection is afforded by the 9-fluorenylmethyloxycarbonyl (Fmoc) group. Repetitive cleavage of this highly base-labile protecting group is done using 20% piperidine in N, N-dimethylformamide. Side-chain functionalities may be protected as their butyl ethers (in the case of serine threonine and tyrosine), butyl esters (in the case of glutamic acid and aspartic acid), butyloxycarbonyl derivative (in the case of lysine and histidine), trityl derivative (in the case of cysteine) and 4-methoxy-2,3,6-trimethylbenzenesulphonyl derivative (in the case of arginine). Where glutamine or asparagine are C-terminal residues, use is made of the 4,4′-dimethoxybenzhydryl group for protection of the side chain amido functionalities. The solid-phase support is based on a polydimethyl-acrylamide polymer constituted from the three monomers dimethylacrylamide (backbone-monomer), bisacryloylethylene diamine (cross linker) and acryloylsarcosine methyl ester (functionalizing agent). The peptide-to-resin cleavable linked agent used is the acid-labile 4-hydroxymethyl-phenoxyacetic acid derivative. All amino acid derivatives are added as their preformed symmetrical anhydride derivatives except for asparagine and glutamine, which are added using a reversed N, N-dicyclohexyl-carbodiimide/1-hydroxybenzotriazole mediated coupling procedure. All coupling and deprotection reactions are monitored using ninhydrin, trinitrobenzene sulphonic acid or isotin test procedures. Upon completion of synthesis, peptides are cleaved

from the resin support with concomitant removal of side-chain protecting groups by treatment with 95% trifluoroacetic acid containing a 50% scavenger mix. Scavengers commonly used include ethanedithiol, phenol, anisole and water, the exact choice depending on the constituent amino acids of the peptide being synthesized. Also, a combination of solid phase and solution phase methodologies for the synthesis of peptides is possible.

[00168] Trifluoroacetic acid is removed by evaporation in vacuo, with subsequent trituration with diethyl ether affording the crude peptide. Any scavengers present are removed by a simple extraction procedure which on lyophilization of the aqueous phase affords the crude peptide free of scavengers.

[00169] Purification may be performed by techniques such as re-crystallization, ion-exchange chromatography, size exclusion chromatography, hydrophobic interaction chromatography and reverse-phase high performance liquid chromatography using e.g. acetonitrile/water gradient separation, or a combination thereof.

[00170] Peptides may be analyzed using thin layer chromatography, electrophoresis, in particular capillary electrophoresis, solid phase extraction (CSPE), reverse-phase high performance liquid chromatography, amino-acid analysis after acid hydrolysis and by fast atom bombardment (FAB) mass spectrometric analysis, as well as MALDI and ESI-Q-TOF mass spectrometric analysis.

[00171] Alternatively, the peptide may be produced by recombinant expression in a heterologous host cell. Such methods typically involve the use of a vector comprising a nucleic acid sequence encoding the peptide to be expressed, to express the polypeptide *in vivo*; for example, in bacteria, yeast, insect or mammalian cells.

[00172] In further embodiments, *in vitro* cell-free systems may be used. The peptides may be isolated and/or may be provided in substantially pure form. For example, they may be provided in a form which is substantially free of other peptides or proteins.

Peptide-MHC (pMHC) Complexes Disclosed Herein

[00173] In another aspect, the disclosure provides a complex of a peptide of the disclosure and an MHC molecule. Preferably, the peptide is bound to the peptide binding groove of the MHC molecule. In some embodiments, the peptide and the MHC molecule

form a non-covalent complex. In other embodiments, the peptide and the MHC molecule may be covalently linked, for example, via a linker.

[00174] MHC molecules are generally classified into two categories: class I and class II MHC molecules. An MHC class I molecule is an integral membrane protein comprising a glycoprotein heavy chain, also referred to herein as the α chain, which has three extracellular domains (i.e., α1, α2 and α3) and two intracellular domains (i.e., a transmembrane domain (TM) and a cytoplasmic domain (CYT)). The heavy chain is noncovalently associated with a soluble subunit called β2 microglobulin (β2m or β2M). An MHC class II molecule or MHC class II protein is a heterodimeric integral membrane protein comprising one α chain and one β chain in noncovalent association. The α chain has two extracellular domains (α1 and α2), and two intracellular domains (a TM domain and a CYT domain). The β chain contains two extracellular domains (β1 and β2), and two intracellular domains (a TM domain and CYT domain).

[00175] The domain organization of class I and class II MHC molecules forms the antigenic determinant binding site, e.g., the peptide-binding portion or peptide binding groove, of the MHC molecule. A peptide binding groove refers to a portion of an MHC molecule that forms a cavity in which a peptide, e.g., antigenic determinant, can bind. The conformation of a peptide binding groove is capable of being altered upon binding of a peptide to enable proper alignment of amino acid residues important for TCR binding to the peptide-MHC (pMHC) complex.

[00176] In some embodiments, MHC molecules include fragments of MHC chains that are sufficient to form a peptide binding groove. For example, a peptide binding groove of a class I protein can comprise portions of the α1 and α2 domains of the heavy chain capable of forming two β-pleated sheets and two α helices. Inclusion of a portion of the β2 microglobulin chain stabilizes the MHC class I molecule. While for most versions of MHC class II molecules, interaction of the α and β chains can occur in the absence of a peptide, the two-chain molecule of MHC class II is unstable until the binding groove is filled with a peptide. A peptide binding groove of a class II protein can comprise portions of the α1 and β1 domains capable of forming two β-pleated sheets and two α helices. A first portion of the α1 domain forms a first β-pleated sheet and a second portion of the α1 domain forms a first a helix. A first portion of the β1 domain forms a second β-pleated sheet and a second

portion of the β1 domain forms a second a helix. The X-ray crystallographic structure of class II protein with a peptide engaged in the binding groove of the protein shows that one or both ends of the engaged peptide can project beyond the MHC protein. Thus, the ends of the α1 and β1 α helices of class II form an open cavity such that the ends of the peptide bound to the binding groove are not buried in the cavity. Moreover, the X-ray crystallographic structure of class II proteins shows that the N-terminal end of the MHC β chain apparently projects from the side of the MHC protein in an unstructured manner since the first 4 amino acid residues of the β chain could not be assigned by X-ray crystallography.

[00177] The peptides of the present disclosure can bind to an MHC molecule in a manner such that the pMHC complex can bind to a TCR, preferably in a specific manner. In certain embodiments, binding of the pMHC complex to the TCR may induce a T cell response.

[00178] Whether or not a given peptide will form a complex with an MHC molecule can be determined by assessing whether the MHC can be refolded in the presence of the peptide using the process described in, for example, PCT Application WO2018/083505, which is incorporated herein in its entirety for all purposes. If the peptide does not form a complex with MHC then MHC will not refold. Refolding can be confirmed using an antibody that recognizes MHC in a folded state only. Alternatively, the ability of a peptide to stabilize MHC on the surface of transporter associated with antigen processing (TAP)-deficient cell lines such as T2 cells, which lack the capacity for TAP-mediated translocation of cytosolic peptides into the endoplasmic reticulum (ER) for peptide loading onto MHC class I molecules, or other biophysical methods to determine interaction parameters can be determined.

[00179] The peptides according to the present disclosure may be provided as an MHC groove-binding peptide. In some embodiments, the MHC groove-binding peptide can be designed such that the peptide may vary in some or all the positions involved in MHC binding. For example, the MHCBN is a comprehensive database of MHC binding and non-binding peptides compiled from published literature and existing databases. The latest version of the database has 25,860 entries including 20,717 MHC binders and 4,022 MHC non-binders for more than 450 MHCs. The database has sequence and structure data

of (a) source proteins of peptides and (b) MHCs. MHCBN has a number of web tools that include: (i) mapping of peptide on query sequence; (ii) search on any field; (iii) creation of data sets; and (iv) online data submission.

[00180] In some cases, a peptide binding tool for prediction of binding to MHC-I or MHC-II can be, for example, Antibody Epitope Prediction, ANTIGENIC, BepiPred, CTLPred, DiscoTope, EPIPREDICT, Epitope Cluster Analysis, Epitope Conservancy Analysis, EUiPro, HLA Peptide Binding Predictions, HLABinding, MAPPP, MHCBench, MHC-I processing predictions, Mosaic Vaccine Tool Suite, NetChop, NetCTL, NetMHC, NetMHCII, NetMHCpan, nHLAPred-I, OptiTope, PAProC, POPI, PREDEP, Prediction of Antigenic Determinants, ProPred, ProPred-1, RankPep, SMM, SVMHC, TAPPred, VaxiJen, or combinations thereof. Additional example programs are used such as BIMAS, SYFPEITHI, or Rankpep.

[00181] In one specific embodiment, a library of altered peptides is produced by genetically engineering the library using polymerase chain reaction (PCR) or any other suitable technique to construct a DNA fragment encoding the peptide. With PCR techniques, by using oligonucleotides that are randomly mutated within particular triplet codons, the resultant fragment pool encodes all possible combination of codons at these positions. Preferably, certain of the amino acid positions are maintained constant, which are the conserved amino acids that are required for binding to the MHC peptide binding groove, and which do not contact the T cell receptor (TCR).

[00182] In some embodiments, when a library of altered peptides is produced by genetically engineering the library using polymerase chain reaction (PCR) or any other suitable technique to construct a DNA fragment encoding the peptide, the target TCR is a TCR for which it is desired to identify the peptide epitope recognized by the receptor. In some embodiments, the target TCR is from a patient with an HTLV-1 infection and/or an HTLV-1-induced disease or disorder). In some embodiments, the TCR includes an α-chain and a β-chain.

[00183] MHC molecules used in pMHC complexes described herein include naturally occurring full-length MHC molecules as well as individual chains of MHC molecules (e.g., MHC class I α (heavy) chain, β2-microglobulin, MHC class II α chain, and MHC class II β chain), individual subunits of such chains of MHCs (e.g., α1, α2 and/or

α3 subunits of MHC class I α chain, α1 and/or α2 subunits of MHC class II α chain, β1 and/or β2 subunits of MHC class II β chain) as well as fragments, mutants, and various derivatives thereof (including fusion proteins, e.g., fusions with viral envelope proteins or fusogens), wherein such fragments, mutants, and derivatives retain the ability to display an antigenic determinant for recognition by an antigen-specific TCR. In one specific embodiment, the MHC comprises a transmembrane domain embedded in the lipid envelope of a liposome, a recombinant viral particle, or a virus-like particle (VLP).

[00184] Naturally-occurring MHC molecules are encoded by a cluster of genes on human chromosome 6 or mouse chromosome 17. MHCs are also referred to as H-2 in mice and Human Leucocyte Antigen (HLA) in humans. MHC class I molecules specifically bind CD8 molecules expressed on cytotoxic T lymphocytes (CD8+ T cells), whereas MHC class II molecules specifically bind CD4 molecules expressed on helper T lymphocytes (CD4+ T cells). MHCs include, but are not limited to, HLA specificities such as A (e.g. A1-A74), B (e.g., B 1-B77), C (e.g., C1-C11), D (e.g., D1-D26), E, G, DR (e.g., DR1-DR8), DQ (e.g., DQ1-DQ9) and DP (e.g. DP1-DP6). More preferably, HLA specificities include A1, A2, A3, A11, A23, A24, A28, A30, A33, B7, B8, B35, B44, B53, B60, B62, DR1, DR2, DR3, DR4, DR7, DR8, and DR-11.

[00185] In some embodiments, the MHC molecule in a pMHC complex of the present disclosure is a human leukocyte antigen (HLA) molecule. The MHC molecule may be a human HLA molecule selected from the group consisting of HLA-A, HLA-B, HLA-C, HLA-E, HLA-F, and HLA-G. In some embodiments, the MHC class I or MHC II polypeptides may be derived from any functional human HLA-A, B, C, DR, or DQ molecules. Non-limiting examples of HLA-A alleles comprise, without limitation, A*0101, A*0201, A*0202, A*0301, A*1101, A*2301, A*2402, A*2501, A*2601, A*2901 , A*2902, A*3101, A*3201, A*3301, A*3401, A*3601, A*4301, A*6601, A*6801, A*6901, A*7401, and A*8001. Non-limiting examples of HLA-B alleles comprise, without limitation, B*0702. B*0801, B*1301, B*1401, B*1402, B*1501, B*1801, B*1802, B*2701, B*2702, B*3501, B*3502, B*3701, B*3801, B*3901, B*4001, B*4101, B*4201, B*4402, B*4501, B*4601, B*4701, B*4801, B*4901, B*5001, B*5101, B*5201, B*5301, B*5401, B*5501, B*5502, B*5601, B*5701, B*5801, B*5901, B*6701, B*7301, B*1517, B*8101, B*8201, and B*8301. Non-limiting examples of HLA-C alleles

comprise, without limitation, Cw*0101, Cw*0202, Cw*0303, Cw*0401, Cw*0501, Cw*0602, Cw*0701, Cw*0702, Cw*0802, Cw*1203, Cw*1401, Cw*1502, Cw*1601, Cw*1701, and. Cw*1801. Non-limiting examples of HLA-DR alleles comprise, without limitation, DRB1*0101, DRB1*0103, DRB1*1501, DRB1*1502, DRB1*1601, DRB1*1602, DRB1*0301, DRB1*0401, DRB1*0404, DRB1*1101, DRB1*1201, DRB1*1301 , DRB1*1302, DRB1*1401, DRB1*1402, DRB1*0701, DRB1*0801, DRB1*0802, DRB1*0803, DRB1*0901, and DRB1*1001.

[00186] In some embodiments, the MHC class I molecule may be selected from HLA-A*02, HLA-A*01, HLA-A*03, HLA-A*11, HLA-A*23, HLA-A*24, HLA-B*07, HLA-B*08, HLA-B*40, HLA-B*44, HLA-B*15, HLA-C*04, HLA*C*03 HLA-C*07. There are also allelic variants of the above HLA types, all of which are encompassed by the present disclosure. In some embodiments, the MHC molecule may be HLA-A*02 or HLA-A*11.

[00187] The MHC molecules used herein may also be from any other mammalian or avian species, for example, non-human primates, rodents (e.g., mice), rabbits, equines, bovines, canines, felines, pigs, etc.

[00188] Naturally occurring MHC class I molecules bind peptides derived from proteolytically degraded proteins, especially endogenously synthesized proteins, by a cell. Small peptides obtained accordingly are transported into the endoplasmic reticulum where they associate with nascent MHC class I molecules before being routed through the Golgi apparatus and displayed on the cell surface for recognition by cytotoxic T lymphocytes.

[00189] Naturally occurring MHC class I molecules consist of an α (heavy) chain associated with β2-microglobulin. The heavy chain consists of subunits α1-α3. The β2-microglobulin protein and α3 subunit of the heavy chain are associated. In certain embodiments, β2-microglobulin and α3 subunit are covalently bound. In certain embodiments, β2-microglobulin and α3 subunit are non-covalently bound. The α1 and α2 subunits of the heavy chain fold to form a groove for a peptide, e.g., antigenic determinant, to be displayed and recognized by TCR.

[00190] Class I molecules can bind peptides of about 8-10 amino acids in length. All humans have between three and six different class I molecules, which can each bind many different types of peptides.

[00191] In some embodiments, the MHC contained in the pMHC complexes of the disclosure comprises (i) a class I MHC polypeptide or a fragment, mutant or derivative thereof, and, optionally, (ii) a β2 microglobulin polypeptide or a fragment, mutant or derivative thereof. In one specific embodiment, the class I MHC polypeptide is linked to the β2 microglobulin polypeptide by a peptide linker.

[00192] In one specific embodiment, the class I MHC polypeptide is a human class I MHC polypeptide selected from the group consisting of HLA-A, HLA-B, HLA-C, HLA-E, HLA-F, and HLA-G. In another specific embodiment, the class I MHC polypeptide is a murine class I MHC polypeptide selected from the group consisting of H-2K, H-2D, H-2L, H2-IA, H2-IB, H2-IJ, H2-IE, and H2-IC.

[00193] In some embodiments, the peptide disclosed herein forms a complex with one or more MHC class I α heavy chains. In some embodiments, the MHC class I α heavy chain is fully human. In some embodiments, the MHC class I α heavy chain is humanized. Humanized MHC class I α heavy chains are described, e.g., in U.S. Pat. Pub. Nos. 2013/0111617, 2013/0185819 and 2014/0245467. In some embodiments, the MHC class I α heavy chain comprises a human extracellular domain (human α1, α2, and/or α3 domains) and a cytoplasmic domain of another species. In some embodiments, the class I α heavy chain polypeptide is HLA-A, HLA-B, HLA-C, HLA-E, HLA-F, HLA-G, HLA-K, or HLA-L. In some embodiments, the HLA-A sequence can be an HLA-A*0201 sequence. In various aspects, the peptide-MHC can include all the domains of an MHC class I heavy chain.

[00194] In some embodiments, the MHC molecule comprises a β2-microglobulin. In some embodiments, the β2-microglobulin is fully human. In some embodiments, the β2-microglobulin is humanized.

[00195] In some embodiments, the MHC class I molecule comprises a mutation in a β2-microglobulin (β2m or B2M) polypeptide and in the Heavy Chain sequence to effect a disulfide bond between the B2M and the Heavy Chain. In some cases, the Heavy Chain is an HLA and wherein the disulfide bond links one of the following pairs of residues: B2M residue 12, HLA residue 236; B2M residue 12, HLA residue 237; B2M residue 8, HLA residue 234; B2M residue 10, HLA residue 235; B2M residue 24, HLA residue 236; B2M residue 28, HLA residue 232; B2M residue 98, HLA residue 192; B2M residue 99, HLA

residue 234; B2M residue 3, HLA residue 120; B2M residue 31, HLA residue 96; B2M residue 53, HLA residue 35; B2M residue 60, HLA residue 96; B2M residue 60, HLA residue 122; B2M residue 63, HLA residue 27; B2M residue Arg3, HLA residue Gly120; B2M residue His31, HLA residue Gln96; B2M residue Asp53, HLA residue Arg35; B2M residue Trp60, HLA residue Gln96; B2M residue Trp60, HLA residue Asp122; B2M residue Tyr63, HLA residue Tyr27; B2M residue Lys6, HLA residue Glu232; B2M residue Gln8, HLA residue Arg234; B2M residue Tyr10, HLA residue Pro235; B2M residue Ser11, HLA residue Gln242; B2M residue Asn24, HLA residue Ala236; B2M residue Ser28, HLA residue Glu232; B2M residue Asp98, HLA residue His192; and B2M residue Met99, HLA residue Arg234, first linker position Gly2, Heavy Chain (HLA) position Tyr84; Light Chain (B2M) position Arg12, HLA Ala236; and/or B2M residue Arg12, HLA residue Gly237.

[00196] In some embodiments, the antigenic determinant amino acid sequence can be that of a peptide described herein, which can be presented by an MHC class I molecule. In certain embodiments, the sequence can comprise from about 8 to about 15 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 12 contiguous amino acids.

[00197] In some embodiments, at least one chain of the MHC and the peptide are comprised within a fusion protein. In one specific embodiment, the MHC and the peptide are separated by a linker sequence. For example, the single chain molecule can comprise, from amino to carboxy terminal, an antigenic determinant, a β2-microglobulin sequence, and a class I α (heavy) chain sequence. Alternatively, the single chain molecule can comprise, from amino to carboxy terminal, an antigenic determinant, a class I α (heavy) chain sequence, and a β2-microglobulin sequence. The single-chain molecule can further comprise a signal peptide sequence at the amino terminal. In certain embodiments, there can be a linker sequence between the peptide sequence and the β2-microglobulin sequence. In certain embodiments, there can be a linker sequence between the β2-microglobulin sequence and the class I α (heavy) chain sequence. A single-chain molecule can further comprise a signal peptide sequence at the amino terminal, as well as first linker sequence extending between the peptide sequence and the β2-microglobulin sequence, and/or a second linker sequence extending between the β2-microglobulin sequence and the class I

heavy chain sequence. In certain embodiments, the β2-microglobulin and the class I α (heavy) chain sequences can be human, murine, or porcine.

[00198] In some embodiments, a single-chain molecule can comprise a first flexible linker between the peptide ligand segment and the β2-microglobulin segment. For example, linkers can extend from and connect the carboxy terminal of the peptide ligand segment to the amino terminal of the β2-microglobulin segment. Preferably, the linkers are structured to allow the linked peptide ligand to fold into the binding groove resulting in a functional MHC-antigen peptide. In some embodiments, this linker can comprise at least about 10 amino acids, up to about 15 amino acids. In some embodiments, a single-chain molecule can comprise a second flexible linker inserted between the β2-microglobulin and heavy chain segments. For example, linkers can extend from and connect the carboxy terminal of the β2-microglobulin segment to the amino terminal of the heavy chain segment. In certain embodiments, the β2-microglobulin and the heavy chain can fold into the binding groove resulting in a molecule which can function in promoting T cell expansion.

[00199] Suitable linkers used in the MHCs can be of any of a number of suitable lengths, such as from 1 amino acid (e.g., Gly) to 20 amino acids, from 2 amino acids to 15 amino acids, from 3 amino acids to 12 amino acids, including 4 amino acids to 10 amino acids, 5 amino acids to 9 amino acids, 6 amino acids to 8 amino acids, or 7 amino acids to 8 amino acids, and can be 1, 2, 3, 4, 5, 6, or 7 amino acids. Non-limiting examples of linkers include, e.g., glycine polymers (G)n, glycine-serine polymers (including, for example, (GS)n, (GSGGS)n (SEQ ID NO: 92) and (GGGS)n (SEQ ID NO: 93), where n is an integer of at least one), glycine-alanine polymers, alanine-serine polymers, and other flexible linkers. Glycine and glycine-serine polymers can be used; both Gly and Ser are relatively unstructured, and therefore can serve as a neutral tether between components. Glycine polymers can be used; glycine accesses significantly more phi-psi space than even alanine, and is much less restricted than residues with longer side chains). Exemplary linkers can comprise amino acid sequences including, but not limited to, those listed in **Table 2**. In some embodiments, a linker peptide includes a cysteine residue that can form a disulfide bond with a cysteine residue present in a second polypeptide.

Table 2. Examples of Linker Sequences

Linker amino acid sequence	SEQ ID NO:	Linker codon-optimized nucleotide sequence (human)	SEQ ID NO:	Linker codon-optimized nucleotide sequence (mouse)	SEQ ID NO:
GGGS	94	GGGGGTGGTTCC	110	GGTGGCGGTAGT	126
GGSG	95	GGTGGGTCTGGG	111	GGGGGATCTGGT	127
GSGGS	96	GGGTCCGGGGGCTCC	112	GGCAGTGGCGGTAGC	128
GGSGG	97	GGTGGGAGCGGTGGT	113	GGAGGGAGTGGAGGG	129
GSGSG	98	GGCAGCGGAAGCGGA	114	GGGTCTGGCTCAGGC	130
GSGGG	99	GGGAGTGGGGGAGGT	115	GGTTCTGGCGGAGGT	131
GGGSG	100	GGTGGGGGAAGTGGA	116	GGTGGTGGGAGTGGA	132
GSSSG	101	GGCAGCTCATCTGGT	117	GGCTCAAGCAGTGGA	133
GCGASGGGGSGGGGS	102	GGATGTGGTGCATCTGGAGGGGGAGGCTCTGGGGGGGGTGGATCT	118	GGCTGTGGGGCTAGTGGGGGAGGTGGTAGTGGTGGTGGCGGTTCC	134
GCGASGGGGSGGGGS	103	GGGTGTGGTGCTAGTGGGGGTGGCGGATCAGGTGGAGGCGGGAGC	119	GGATGTGGGGCCTCAGGTGGGGGTGGCAGCGGTGGTGGAGGGTCA	135
GGGGSGGGGS	104	GGGGGCGGAGGATCTGGGGGAGGGGGATCA	120	GGTGGCGGGGGCTCTGGTGGAGGAGGATCT	136
GGGASGGGGSGGGGS	105	GGGGGGGGCGCTTCAGGCGGAGGTGGAAGTGGTGGAGGAGGT	121	GGAGGCGGCGCTTCTGGGGGCGGGGGTAGTGGGGGTGGAGGT	137
GGGGSGGGGSGGGGS	106	GGAGGGGGAGGTTCTGGCGGCGGGGGATCAGGAGGCGGTGGGAGC	122	GGTGGAGGTGGAAGTGGAGGAGGGGGATCAGGCGGAGGCGGGAGC	138
GGGASGGGGS	107	GGTGGGGGGGCGTCAGGTGGAGGCGGAAGT	123	GGAGGGGGAGCCTCTGGCGGTGGAGGATCA	139
GGGGSGGGGSGGGGS	108	GGCGGCGGAGGTTCTGGTGGGGGTGGCAGTGGAGGAGGAGGCAGC	124	GGGGGAGGAGGCAGTGGAGGTGGGGGAAGTGGTGGAGGGGGTCT	140

GGGGSGGGGS GGGGSGGGGS	109	GGAGGTGGAGGT AGTGGCGGTGGT GGGTCAGGGGGA GGCGGGTCCGGT GGCGGTGGGAGT	125	GGGGGTGGAGGAT CAGGAGGCGGTGG TTCTGGGGGAGGT GGATCCGGCGGGG GTGGTAGT	141

[00200] In certain embodiments, the single-chain molecule can comprise a peptide covalently attached to an MHC class I α (heavy) chain via a disulfide bridge (i.e., a disulfide bond between two cystines). In certain embodiments, the disulfide bond comprises a first cysteine, comprising a linker extending from the carboxy terminal of an antigen peptide, and a second cysteine comprising an MHC class I heavy chain (e.g., an MHC class I α (heavy) chain which has a non-covalent binding site for the antigen peptide). In certain embodiments, the second cysteine can be a mutation (addition or substitution) in the MHC class I α (heavy) chain. In certain embodiments, the single-chain molecule can comprise one contiguous polypeptide chain as well as a disulfide bridge. In certain embodiments, the single-chain molecule can comprise two contiguous polypeptide chains which are attached via the disulfide bridge as the only covalent linkage. In some embodiments, the linking sequences can comprise at least one amino acid in addition to the Cys residues, including one or more Gly residues, one or more Ala residues, and/or one or more Ser residues.

[00201] In certain embodiments, the disulfide bridge can link an antigen peptide described herein in the class I groove of the pMHC complex if the pMHC complex comprises a first cysteine in a Gly-Ser linker extending between the C-terminus of the peptide and the β2-microglobulin, and a second cysteine in a proximal heavy chain position.

[00202] Attaching the peptide to the MHC class I or MHC class II molecule via a flexible linker has the can help ensure that the peptide will occupy and stay associated with the MHC molecule during biosynthesis, transport, and display. However, there may be situations in which this linker can interfere with peptide binding to the MHC molecule or with TCR recognition of the complex. As an alternate approach, in some embodiments, the MHC molecule and the peptide are expressed separately.

[00203] In some embodiments, the β2-microglobulin sequence can comprise a full-length β2-microglobulin sequence. In certain embodiments, the β2-microglobulin

sequence lacks the leader peptide sequence. As such, in some configurations, the β2-microglobulin sequence can comprise about 99 amino acids, and can be a mouse β2-microglobulin sequence (e.g., GenBank Accession No. X01838). In some other configurations, the β2-microglobulin sequence can comprise about 99 amino acids, and can be a human β2-microglobulin sequence (e.g., GenBank Accession No. AF072097.1).

[00204] In some embodiments, the pMHC complex can contain MHC sequences as disclosed in U.S. Patent Nos. 4,478,823; 6,011,146; 8,518,697; 8,895,020; 8,992,937; WO 96/04314; Mottez et al. J. Exp. Med. 181: 493-502, 1995; Madden et al. Cell 70: 1035-1048, 1992; Matsumura et al., Science 257: 927-934, 1992; Mage et al., Proc. Natl. Acad. Sci. USA 89: 10658-10662, 1992; Toshitani et al, Proc. Nat'l Acad. Sci. 93: 236-240, 1996; Chung et al, J. Immunol. 163:3699-3708, 1999; Uger and Barber, J. Immunol. 160: 1598-1605, 1998; Uger et al., J. Immunol. 162, pp. 6024-6028, 1999; White et al., J. Immunol. 162: 2671-2676, 1999; Yu et al., J. Immunol. 168:3145-3149, 2002; Truscott et al., J. Immunol. 178: 6280–6289, 2007, all of which are incorporated by reference in their entireties.

[00205] In some embodiments, the MHC comprises a class II MHC polypeptide or a fragment, mutant or derivative thereof. In one specific embodiment, the MHC comprises α and β polypeptides of a class II MHC complex or a fragment, mutant or derivative thereof. In one specific embodiment, the α and β polypeptides are linked by a peptide linker. In one specific embodiment, the MHC comprises α and β polypeptides of a human class II MHC complex selected from the group consisting of HLA-DP, HLA-DR, HLA-DQ, HLA-DM and HLA-DO. In another specific embodiment, the MHC comprises α and β polypeptides of a murine H-2A or H-2E class II MHC complex.

[00206] Naturally occurring MHC class II molecules can contain two polypeptide chains, α and β. The chains may come from the DP, DQ, or DR gene groups. There are about 40 known different human MHC class II molecules. All have the same basic structure but can vary subtly in their molecular structure. MHC class II molecules can bind peptides of 13-18 amino acids in length.

[00207] In some embodiments, the MHC class II α chain is fully human. In some embodiments, the MHC class II α chain is humanized. Humanized MHC class II α chains are described, e.g., in U.S. Pat. Nos. 8,847,005, 9,043,996, and 10,154,658, which are

incorporated herein by reference in their entireties. In some embodiments, the humanized MHC class II α chain polypeptide comprises a human extracellular domain and a cytoplasmic domain of another species. In some embodiments, the class II α chain is HLA-DMA, HLA-DOA, HLA-DPA, HLA-DQA or HLA-DRA. In some embodiments, the class II α chain polypeptide is humanized HLA-DMA, HLA-DOA, HLA-DPA, HLA-DQA and/or HLA-DRA.

[00208] In some embodiments, the peptide of the present disclosure forms a complex with one or more MHC class II β chains. In some embodiments, the MHC class II β chain is fully human. In some embodiments, the MHC class II β chain polypeptide is humanized. Humanized MHC class II β chain polypeptides are described, e.g., in U.S. Pat. Nos. 8,847,005, 9,043,996, and 10,154,658, which are incorporated herein by reference in their entireties. In some embodiments, the humanized MHC class II β chain comprises a human extracellular domain and a cytoplasmic domain of another species. In some embodiments, the class II β chain is HLA-DMB, HLA-DOB, HLA-DPB, HLA-DQB or HLA-DRB. In some embodiments, the class II β chain is humanized HLA-DMB, HLA-DOB, HLA-DPB, HLA-DQB and/or HLA-DRB.

[00209] The pMHC complexes of the disclosure may be isolated and/or in a substantially pure form. For example, the complex may be provided in a form which is substantially free of other peptides or proteins. MHC molecules as disclosed herein can include recombinant MHC molecules, non-naturally occurring MHC molecules, and functionally equivalent fragments of MHC, including derivatives or variants thereof, provided that peptide binding is retained. For example, MHC molecules may be fused to a therapeutic moiety, attached to a solid support, in soluble form, attached to a tag, biotinylated and/or in multimeric form. A peptide disclosed herein may be covalently attached to the MHC.

[00210] Methods to produce soluble recombinant MHC molecules with which peptides disclosed herein can form a complex include, but are not limited to, expression and purification from *E. coli* cells or insect cells. Alternatively, MHC molecules may be produced synthetically, or using cell free systems.

[00211] The peptides disclosed herein may be presented on the surface of a cell in complex with MHC. Thus, the present disclosure also provides a cell presenting on its

surface a pMHC complex disclosed herein. Such a cell may be a mammalian cell, preferably a cell of the immune system, and a specialized antigen-presenting cell (APC) such as a dendritic cell or a B cell. Other preferred cells include T2 cells. Cells presenting the peptide or pMHC complex of the disclosure may be isolated, preferably in the form of a homogenous population, or provided in a substantially pure form. Such cells may not naturally present the complex of the disclosure, or alternatively said cells may present the complex at a level higher than they would in nature. Such cells may be obtained by pulsing said cells with one or more peptides (e.g., 2 to 50, 2 to 40, 2 to 30, 5 to 25, 5 to 20, or 10 to 15 peptides) of the disclosure, or genetically modifying the cells (via DNA or RNA transfer) to express one or more peptides (e.g., 2 to 50, 2 to 40, 2 to 30, 5 to 25, 5 to 20, or 10 to 15 peptides) of the disclosure. Pulsing involves incubating the cells with the peptide for several hours using peptide concentrations typically ranging from 10^{-5} to 10^{-12} M. Such cells may additionally be transduced with HLA molecules, such as HLA-A*02 to further induce presentation of the peptide(s). Cells may be produced recombinantly. Cells presenting peptides of the disclosure may be used to isolate T cells and TCRs which are activated by, or bind to, the cells.

Fusion Proteins, Conjugates and Oligomeric Complexes Disclosed Herein

[00212] Peptides or pMHC complexes disclosed herein may be fused or conjugated to one or more heterologous molecules. Peptides or pMHC complexes of the disclosed herein may also be in multimeric form. Accordingly, the present disclosure also provides fusion proteins, conjugates, and oligomeric complexes comprising a peptide or a pMHC complex of the disclosure.

[00213] In some embodiments, peptides are fused or conjugated to one or more heterologous molecules which includes an MHC molecule (or fragments thereof).

[00214] Heterologous molecules suitable for genetical fusion and/or chemical conjugation with the peptides or the pMHC complexes of the disclosure include, but are not limited to, peptides, polypeptides, small molecules, polymers, nucleic acids, lipids, sugars, etc. The heterologous molecule(s) may be fused at the N- and/or C-terminus of the peptide and/or another polypeptide chain in the pMHC complex.

[00215] Heterologous peptides and polypeptides include, but are not limited to, an epitope (e.g., FLAG) or a tag sequence (e.g., His_6 (SEQ ID NO: 147), and the like) to allow for the detection and/or isolation of a fusion protein; a transmembrane receptor protein or a portion thereof, such as an extracellular domain or a transmembrane and intracellular domain; a ligand or a portion thereof which binds to a transmembrane receptor protein; an enzyme or portion thereof which is catalytically active; a polypeptide or peptide which promotes oligomerization, such as a leucine zipper domain; a polypeptide or peptide which increases stability, such as an immunoglobulin constant region (e.g., an Fc domain); a half-life-extending sequence comprising a combination of two or more (e.g., 2, 5, 10, 15, 20, 25, etc.) naturally occurring or non-naturally occurring charged and/or uncharged amino acids (e.g., Ser, Gly, Glu or Asp) designed to form a predominantly hydrophilic or predominantly hydrophobic fusion partner for a fusion protein; a functional or non-functional antibody (e.g., an antibody that is specific for dendritic cells), or a heavy or light chain thereof; and a polypeptide which has an activity, such as a therapeutic activity, different from fusion proteins of the present disclosure. In some embodiments, the one or more heterologous molecules enhances a peptide-specific immune response in a subject. In some embodiments, the one or more heterologous molecules mediates peptide delivery to a specific site within a subject.

[00216] In some embodiments, fusion proteins of the disclosure may comprise one or more affinity tags, e.g., to allow for affinity purification or coupling to another molecule. Examples of affinity tags include, but are not limited to, a His_6 tag (SEQ ID NO: 147), an Avi-tag, a biotin, a hemagglutinin (HA) tag, a FLAG tag, a Myc tag, a GST tag, a MBP tag, a chitin binding protein tag, a calmodulin tag, a V5 tag, a streptavidin binding tag, a green fluorescent protein (GFP), YFP, RFP, CFP, mCherry, tdTomato, SUMO tag, and Ubiquitin tag.

[00217] In some embodiments, fusion proteins of the disclosure may comprise one or more epitopes that is not present in the antigen. One such example is the use of fusion peptides where a promiscuous T helper epitope is covalently linked (e.g., via a polypeptide linker or spacer) to the peptide sequence. Non-limiting examples of promiscuous T helper epitopes include the PADRE peptide, tetanus toxoid peptide (830-843), or influenza haemagglutinin, HA(307-319).

[00218] Peptides or pMHC complexes of the disclosure may be conjugated to additional moieties such as carrier molecules or adjuvants for use as vaccines. Examples of adjuvants used in vaccines include microbes, such as the bacterium Bacillus Calmette-Guérin (BCG), and/or substances produced by bacteria, such as Detox B (an oil droplet emulsion of monophosphoryl lipid A and mycobacterial cell wall skeleton). KLH (keyhole limpet hemocyanin), bovine serum albumin (BSA), the E2 core protein of the pyruvate dehydrogenase complex are examples of suitable carrier proteins used in vaccine compositions. Additional examples of carrier proteins suitable for use in the compositions of the present disclosure include, but are not limited to, ovalbumin (OVA), blue carrier protein (BCP), thyroglobulin (THY), a soybean trypsin inhibitor (STI), and multiple attachment peptide (MAP), albumin, serum albumin, c-reactive protein, conalbumin, lactalbumin, ion carrier protein, acyl carrier protein, signal transduction adapter protein, androgen binding protein, calcium binding protein, calmodulin binding protein, ceruloplasmin, cholesterol Ester transfer protein, f box protein, fatty acid binding protein, follistatin, follistatin related protein, GTP binding protein, insulin-like growth factor binding protein, iron binding protein, latent TGFbeta binding protein, light-harvesting protein complex, lymph Sphere antigen, membrane transport protein, neurophysin, periplasmic binding protein, phosphate binding protein, phosphatidylethanolamine binding protein, phospholipid transport protein, retinol binding protein, RNA binding protein, s-phase kinase related protein, sex hormone binding globulin, Thyroxine binding protein, transcobalamin, transcortin, transferrin binding protein, and/or vitamin D binding protein.

[00219] As a further example, a peptide or pMHC complex of the present disclosure may be fused to, for example, the 80 N-terminal amino acids of the HLA-DR antigen-associated invariant chain (p33 or Ii) as derived from the NCBI, GenBank Accession-number X00497). The Ii fragment may facilitate an efficient introduction of the peptide or pMHC complex into the cells.

[00220] Peptides or pMHC complexes of the present disclosure may also be attached, covalently (e.g., via a linker) or non-covalently, to a moiety capable of eliciting a therapeutic effect, such as antibodies, or cytokines, such as interleukin 2, interferon-α, and granulocyte-macrophage colony-stimulating factor. Alternatively or additionally, the peptides or pMHC complexes may be encapsulated into liposomes.

[00221] Other suitable heterologous molecules include, but are not limited to, fluorescent, or luminescent labels, radiolabels, nucleic acid probes, and contrast reagents, antibodies, or enzymes that produce a detectable product. Methods for detecting heterologous molecules may include flow cytometry, microscopy, electrophoresis, or scintillation counting.

[00222] In some embodiments, peptides or pMHC complexes of the disclosure may be conjugated with fluorocarbon to increase cellular immunogenicity. Where the peptide or another polypeptide chain of the pMHC complex is linked to a fluorocarbon, the terminus of the peptide or polypeptide chain, such as the terminus that is not conjugated to the fluorocarbon, or other attachment, can be altered, for example to promote solubility of the fluorocarbon-peptide/polypeptide construct via the formation of micelles. To facilitate large-scale synthesis of the construct, the N- or C-terminal amino acid residues of the peptide or another polypeptide chain of the pMHC complex can be modified. When the desired peptide or another polypeptide chain of the pMHC complex is particularly sensitive to cleavage by peptidases, the normal peptide bond can be replaced by a non-cleavable peptide mimetic. Such bonds and methods of synthesis are well known in the art.

[00223] Peptides or pMHC complexes of the disclosure may be provided in soluble form, or may be immobilized by attachment to a suitable solid support. Examples of solid supports include, but are not limited to, a bead, a membrane, sepharose, a magnetic bead, a plate, a tube, a column. pMHC complexes may be attached to an ELISA plate, a magnetic bead, or a surface plasmon resonance biosensor chip. Methods of attaching peptides or pMHC complexes to a solid support are known to the skilled person, and include, for example, using an affinity binding pair, e.g. biotin and streptavidin, or antibodies and antigens. In some embodiments, peptides or pMHC complexes are labeled with biotin and attached to streptavidin-coated surfaces.

[00224] Peptides or pMHC complexes of the disclosure may be in multimeric form, for example, dimeric, or tetrameric, or pentameric, or octameric, or greater. Accordingly, in some aspects, the present disclosure provides oligomeric complexes comprising the peptides or pMHC complexes of the present disclosure. As used herein, the terms “oligomer”, “oligomeric”, “oligomerize” and “oligomerization” or the like encompass a dimer, trimer, tetramer, pentamer, hexamer, heptamer, octamer, or higher species of

polymerized monomers that comprise the peptide or pMHC complex. Having multiple copies of the peptides or pMHC complexes in a large complex may enhance their biological activity, e.g., immunogenic activity.

[00225] For example, the peptides of the disclosure may be oligomerized using the biotin/streptavidin system. Biotinylated analogs of peptide monomers may be synthesized by standard techniques. For example, the peptide may be C-terminally biotinylated. These biotinylated peptide monomers are then oligomerized by incubation with streptavidin [e.g., at a 4:1 molar ratio at room temperature in phosphate buffered saline (PBS) or HEPES-buffered RPMI medium for 1 hour]. In a variation of this embodiment, biotinylated peptide monomers may be oligomerized by incubation with anti-biotin antibodies [e.g., goat anti-biotin IgG].

[00226] In general, oligomeric pMHC complexes may be produced using pMHC tagged with a biotin residue and complexed through fluorescently labeled streptavidin. A biotinylation site may be introduced to the pMHC complex to which biotin can be added, for example, using the BirA enzyme. Alternatively, oligomeric pMHC complexes may be formed by using immunoglobulin as a molecular scaffold. In this system, the extracellular domains of MHC molecules are fused with the constant region of an immunoglobulin heavy chain separated by a short amino acid linker. Oligomeric pMHC complexes have also been produced using carrier molecules such as dextran. Oligomeric pMHC complexes can be useful for improving the detection of binding moieties, such as T cell receptors, which bind said complex, because of avidity effects.

[00227] In other embodiments, the peptides or pMHC complexes of the disclosure can be oligomerized by covalent attachment to at least one linker. The linker moiety can be a peptide linker, such as those described herein (e.g., in Table 2). In some embodiments, polyethylene glycol (PEG) may serve as the linker that oligomerizes the peptide monomers. For example, a single PEG moiety may be simultaneously attached to the N-termini of both peptide chains of a peptide dimer.

[00228] Alternatively, oligomeric peptide or pMHC complexes may also contain one or more intramolecular disulfide bonds between cysteine residues of the peptide or pMHC monomers. Preferably, the two monomers contain at least one intramolecular disulfide bond. Most preferably, both monomers contain an intramolecular disulfide bond,

such that each monomer contains a cyclic group. Such disulfide bonds may be formed by oxidation of the cysteine residues of the peptide core sequence. In one embodiment the control of cysteine bond formation is exercised by choosing an oxidizing agent of the type and concentration effective to optimize formation of the desired isomer. For example, oxidation of a peptide dimer to form two intramolecular disulfide bonds (one on each peptide chain) is preferentially achieved (over formation of intermolecular disulfide bonds) when the oxidizing agent is DMSO. The formation of cysteine bonds can be controlled by the selective use of thiol-protecting groups during peptide synthesis.

[00229] In some embodiments, peptides or pMHC complexes described herein may be fused or conjugated to a dimerization moiety. The dimerization moiety may contain, for example, an immunoglobulin domain, such as from an IgG antibody (e.g., human IgG), which connects two monomers generating a homodimer or heterodimer molecule. As a non-limiting example, the dimerization motif in the proteins according to the present disclosure may be constructed to include a hinge region and an immunoglobulin domain (e.g. Cy3 domain), e.g., carboxyterminal C domain (CH3 domain), or a sequence that is substantially identical to the C domain. The hinge region may be Ig derived and contributes to the dimerization through the formation of an interchain covalent bond(s), e.g. disulfide bridge(s). In addition, such homodimer or heterodimer molecules may further comprise one or more targeting moieties that bind to target molecules present on, for example, antigen-presenting cells (APCs) such as dendritic cells or B cells. In such instances, the hinge region may function as a flexible spacer between the domains allowing the two targeting units to bind simultaneously to two target molecules on the APC expressed with variable distances. The immunoglobulin domains contribute to dimerization through non-covalent interactions, e.g. hydrophobic interactions. In a preferred embodiment the CH3 domain is derived from IgG. These dimerization moieties may be exchanged with other multimerization moieties from e.g., other Ig isotypes/subclasses. Preferably the dimerization motif is derived from native human proteins, such as human IgG. Examples of such homodimer protein construct are described in US 10,590,195, which is incorporated herein by reference in its entirety.

Nucleic Acids and Vectors

[00230] In another aspect, the disclosure provides an isolated polynucleotide comprising a nucleic acid sequence encoding one or more peptide(s) and/or peptide-based molecules (such as complexes (e.g., pMHC complexes), fusion proteins, or conjugates comprising the described peptides) of the disclosure. The polynucleotide may be, for example, DNA, cDNA, PNA, RNA or combinations thereof, either single- and/or double-stranded, or native or stabilized forms of polynucleotides, such as, for example, polynucleotides with a phosphorothioate backbone and it may or may not contain introns so long as it codes for the peptide.

[00231] In some embodiments, the polynucleotide described herein encodes a peptide comprising an amino acid sequence that is at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99% identical to the amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146, or a fragment or derivative thereof. In some embodiments, the polynucleotide described herein encodes a peptide comprising an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146, or a fragment or derivative thereof.

[00232] In some embodiments, the polynucleotide described herein encodes more than one peptide selected from any one of SEQ ID NOs: 1-89 and 143-146, or a fragment or derivative thereof . For example, the polynucleotide described herein may encodes 2 to 50, 2 to 40, 2 to 30, 5 to 25, 5 to 20, or 10 to 15 peptides as described herein (e.g., SEQ ID NO: 1-89 and 143-146), or a fragment or derivative thereof. The peptides may be arranged in any order, and may be identical or different.

[00233] In some embodiment, the polynucleotide described herein encodes 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, or 93 amino acid sequences selected from SEQ ID NO: 1-89 and 143-146, or a fragment or derivative thereof.

[00234] In some embodiment, the polynucleotide described herein is a DNA molecule.

[00235] Methods to deliver DNA to a subject include, for example, direct delivery, as naked DNA. Delivery may also be achieved by nanoparticles; gene gun, microneedle array and *in situ* electroporation. The nucleic acids can also be administered using ballistic delivery. Particles comprised solely of DNA can be administered. Alternatively, DNA can be adhered to particles, such as gold particles.

[00236] In some embodiments, the polynucleotide described herein is an RNA molecule. For example, the RNA molecule may be mRNA or a self-replicating RNA.

[00237] A polynucleotide encoding RNA disclosed herein can be used to make a vaccine. RNA cannot integrate into the genome and has therefore little oncogenic potential; thus, RNA can be useful for making a vaccine. Also, RNA only needs to enter the cytoplasm, contrary to DNA which needs to enter the nucleus. An RNA molecule disclosed herein may be chemically modified and/or incorporate modified nucleosides to overcome susceptibility to degradation. RNA vaccines may comprise mRNA and/or self-replicating RNA (also known as RNA replicons). Delivery techniques for RNA vaccines may also encompass, for example, condensation with protamine and encapsulation into liposomes or nanoparticles..

[00238] The nucleic acids (either DNA or RNA) can also be delivered complexed to cationic compounds, such as cationic lipids. Lipid-mediated gene delivery methods are described, for example, in WO 91/06309; WO 93/24640; WO 96/18372; U.S. Pat No. 5,279,833, which are incorporated herein by reference in their entireties.

[00239] A nucleic acid molecule described herein may be generated synthetically. One method is the phosphoramidite method. Without wishing to be bound by theory, in this chemistry, a phosphoramidite (a nucleoside with side protecting groups that preserve the integrity of the sugar, the phosphodiester linkage, and the base during chain extension steps) is coupled through its reactive 3' phosphorous group to the 5' hydroxyl group of a nucleoside immobilized on a solid support column. The steps of oligonucleotide synthesis can include the following: (1) Detritylation, in which the dimethoxytrityl (DMT or trityl) group on the 5' hydroxyl of the support nucleoside is removed by treatment with trichloroacetic acid (TCA). (2) In the coupling step, a phosphoramidite, made reactive by tetrazole (a weak acid), is chemically coupled to the last base added to the column support material. (3) In the capping step, any free 5' hydroxyl groups of unreacted column

nucleotides are acetylated by treatment with acetic anhydride and N-methylimidazole. (4) In the oxidation step, the unstable internucleotide phosphate linkage between the previously coupled base and the most recently added base is oxidized by treatment with iodine and water to a more stable phosphotriester linkage. Following coupling of all bases in the oligonucleotide's sequence, the completed nucleic acid chain may be cleaved from the column by treatment with ammonium hydroxide, and the base protecting groups are removed by heating in the ammonium hydroxide solution.

[00240] By way of a non-limiting example, a synthesis cycle may comprise growth of the nucleotide chain from an initial protected nucleoside derivatized via its terminal 3′ hydroxyl to a solid support. Reagents and solvents can be pumped through the support to induce the consecutive removal and addition of sugar protecting groups in order to isolate the reactivity of a specific chemical moiety on the monomer and effect its stepwise addition to the growing oligonucleotide chain. Assembly of the protected oligonucleotide chain can be carried out in chemical steps, for example, without limitation, deblocking, activation/coupling, oxidation, and capping. Cleavage and deprotection then reveal the single-stranded nucleic acid.

[00241] Nucleic acid synthesis methods disclosed herein can comprise, for example, oligonucleotide synthesis, column-based oligonucleotide synthesis, microarray-based oligonucleotide synthesis, gene synthesis from oligonucleotides, gene synthesis from array-derived oligonucleotide pools, and any of various error correction and sequence validation steps, or any combination thereof.

[00242] RNA chemical synthesis may be similar to that used for DNA. In some embodiments, RNA chemical synthesis methods may comprise an additional protecting group at the 2′ hydroxyl of ribose. The 2′ hydroxyl of ribose position may be protected with tert-butyldimethyl silyl groups, which can be stable throughout the synthesis, and can be removed at the final deprotection step by addition of a basic fluoride ion such as tetrabutylammonium fluoride (TBAF). The remaining positions on both the sugar and the bases can be protected in the same fashion as for DNA. By adjusting several parameters in the DNA synthesis protocol such as, but not limited to the coupling times, monomer delivery rate, frequency of washing steps, and types of capping reagents, stepwise coupling efficiencies of up to 99% can be obtained.

[00243] Viral nucleic acid synthesis may be catalyzed by both viral and host enzymes, the relative contribution of which can be determined by the type of virus and the specific molecule. Viruses with RNA genomes, except for the retroviruses, synthesize mRNA and replicate their genomes using virus-encoded RNA-dependent RNA polymerases. In contrast, retroviruses synthesize a double-stranded complementary DNA (cDNA) copy of their single-stranded RNA genome using a virion-encoded RNA-dependent DNA polymerase (reverse transcriptase). In subsequent steps, the retroviral cDNA may be integrated into the host chromosome and transcribed by host-encoded DNA-dependent RNA polymerase II (pol II) to yield viral messages and genomic RNA. DNA viruses, except for poxviruses, also use host-encoded pol II to transcribe their messages. Poxviruses, because they replicate in the cytoplasm and do not have access to pol II, assemble a novel transcriptase composed of multiple poxvirus-specific (and possibly one or more host-derived) subunits. Most DNA virus families (e.g. *Poxviridae*, *Iridoviridae*, *Herpesviridae*, *Adenoviridae*) synthesize a virus-encoded DNA-dependent, DNA polymerase. However, two families (i.e., *Parvoviridae* and *Papovaviridae*) utilize host DNA polymerase, and the *Hepadnaviridae* replicate viral DNA through an RNA intermediate using a virus-encoded reverse transcriptase.

[00244] Due to the degeneracy of the genetic code, nucleic acid molecules of different nucleotide sequence can encode the same amino acid sequence. For expression in various hosts, the polynucleotides may be codon-optimized.

[00245] In a further aspect, the disclosure provides a vector comprising a nucleic acid sequence according to the third aspect of the disclosure. The vector may include, in addition to a nucleic acid sequence encoding only a peptide of the disclosure, one or more additional nucleic acid sequences encoding one or more additional peptides. Such additional peptides may, once expressed, be fused to the N-terminus or the C-terminus of the peptide of the disclosure. Examples of such additional peptides are detailed in the sections above. In one embodiment, the vector includes a nucleic acid sequence encoding a peptide or protein tag such as, for example, a biotinylation site, a FLAG-tag, a MYC-tag, an HA-tag, a GST-tag, a Strep-tag or a poly-histidine tag.

[00246] The vector utilized in the context of the present disclosure desirably comprises sequences appropriate for introduction into cells. For instance, the vector may

be an expression vector, a vector in which the coding sequence of the polypeptide is under the control of its own cis-acting regulatory elements, a vector designed to facilitate gene integration or gene replacement in host cells, and the like.

[00247] In the context of the present disclosure, the term "vector" encompasses a DNA molecule, such as a plasmid, bacteriophage, phagemid, virus or other vehicle, which contains one or more heterologous or recombinant nucleotide sequences (e.g., an above-described nucleic acid molecule of the disclosure, under the control of a functional promoter and, possibly, also an enhancer) and is capable of functioning as a vector in the sense understood by those of ordinary skill in the art.

[00248] The following vectors are provided by way of example: bacteriophages such as lambda (X) bacteriophage, EMBL bacteriophage; bacterial vectors such as pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a; pTrc99A, pKK223-3, pKK233-3, pDR540, and pRIT5; eukaryotic vectors such as pWLneo, pSV2cat, pOG44, PXR1, pSG, pSVK3, pBPV, pMSG and pSVL; and transposons such as Sleeping Beauty transposon and PiggyBac transposon.

[00249] In some embodiments, the vector is a viral vector. Viral vectors can be derived from naturally occurring virus genomes, which typically are modified to be replication incompetent, e.g. non-replicating. Non-replicating viruses require the provision of proteins in trans for replication. Typically, those proteins are stably or transiently expressed in a viral producer cell line, thereby allowing replication of the virus. The viral vectors are, thus, typically infectious and non-replicating. Viral vectors may be adenovirus vectors, adeno-associated virus (AAV) vectors (e.g., AAV type 5 and type 2), alphavirus vectors (e.g., Venezuelan equine encephalitis virus (VEE), Sindbis virus (SIN), Semliki forest virus (SFV), and VEE-SIN chimeras), herpes virus vectors (e.g., vectors derived from cytomegaloviruses, like rhesus cytomegalovirus (RhCMV)), arena virus vectors (e.g. lymphocytic choriomeningitis virus (LCMV) vectors), measles virus vectors, pox virus vectors (e.g., vaccinia virus, modified vaccinia virus Ankara (MVA), NYVAC (derived from the Copenhagen strain of vaccinia), and avipox vectors (canarypox (ALVAC) and fowlpox (FPV) vectors), vesicular stomatitis virus (VSV) vectors, retrovirus vectors, lentivirus vectors, simian virus 40 (SV40), bovine papilloma viruses, Epstein-Barr viruses, Moloney murine leukemia viruses, Harvey murine sarcoma viruses, murine mammary

tumor viruses, Rous sarcoma viruses, poxvirus viral like particles, baculoviral vectors and bacterial spores.

[00250] As further examples, adenovirus vectors may be derived from human adenovirus (Ad) but also from adenoviruses that infect other species, such as bovine adenovirus (e.g. bovine adenovirus 3, BAdV3), a canine adenovirus (e.g. CAdV2), a porcine adenovirus (e.g. PAdV3 or 5), or great apes, such as Chimpanzee (Pan), Gorilla (Gorilla), Orangutan (Pongo), Bonobo (Pan paniscus) and common chimpanzee (Pan troglodytes). Poxvirus (Poxviridae) vectors may be derived from smallpox virus (variola), vaccinia virus, cowpox virus or monkeypox virus. Exemplary vaccinia viruses are the Copenhagen vaccinia virus (W), New York Attenuated Vaccinia Virus (NYVAC), ALVAC, TROVAC and Modified Vaccinia Ankara (MVA).

[00251] Many expression systems are known in the art, including bacteria (for example E. *coli* and *Bacillus subtilis*), yeasts (for example *Saccharomyces cerevisiae*), filamentous fungi (for example *Aspergillus spec.*), plant cells, animal cells (e.g., mammalian cells), and insect cells.

[00252] In yet another aspect, the disclosure provides a host cell comprising the vector of the disclosure. The host cell can be either prokaryotic or eukaryotic. Bacterial cells may be preferred prokaryotic host cells in some circumstances and typically are a strain of E. coli such as, for example, the *E. coli* strains DH5 and RR1. Non-limiting examples of eukaryotic host cells include yeast, insect, and mammalian cells (e.g., from a mouse, rat, monkey, or human cell lines). Non-limiting examples of yeast host cells include, e.g., YPH499, YPH500, and YPH501. Non-limiting examples of mammalian host cells include Chinese hamster ovary (CHO) cells, NIH Swiss mouse embryo cells NIH/3T3, monkey kidney-derived COS-1 cells, and 293 cells which are human embryonic kidney cells. Examples of insect cells include Sf9 cells, which can be transfected with baculovirus expression vectors.

[00253] Transformation of appropriate cell hosts with a DNA construct of the present disclosure is accomplished by well-known methods that typically depend on the type of vector used. Successfully transformed cells, i.e. cells that contain a DNA construct of the present disclosure, can be identified by, for example, PCR. Alternatively, the presence of the protein in the supernatant can be detected using antibodies.

[00254] It will be appreciated that certain host cells of the disclosure are useful in the preparation of the peptides or peptide-based molecules of the disclosure, for example bacterial, yeast, and insect cells. However, other host cells may be useful in certain therapeutic methods. For example, antigen-presenting cells (APCs), such as dendritic cells or B cells, may be used to express the peptides of the disclosure such that the peptides may be loaded into appropriate MHC molecules.

[00255] A further aspect of the disclosure provides a method of producing peptides or peptide-based molecules of the disclosure, the method comprising culturing a host cell and isolating the peptide or peptide-based molecule from the host cell or its culture medium.

Peptide and pMHC Binding Moieties

[00256] Peptides, pMHC complex, or other peptide-based molecules (such as a complex, fusion protein, or conjugate comprising a peptide disclosed herein) of the present disclosure can be used to identify and/or isolate binding moieties that bind specifically to a peptide, pMHC complex, or other peptide-based molecule of the disclosure. Such binding moieties may be used as immunotherapeutic reagents and may include, e.g., antibodies (or antigen-binding fragments thereof), alternative scaffolds, TCRs, and CARs.

[00257] In one aspect, the disclosure provides a peptide binding moiety that binds a peptide of the disclosure. Preferably the peptide binding moiety binds a peptide when the peptide is in complex with MHC. In the latter instance, the peptide binding moiety may bind partially to the MHC, provided that the peptide binding moiety also binds to the peptide. The peptide binding moiety may bind only the peptide, and that binding may be specific. The peptide binding moiety may bind only the pMHC complex and that binding may be specific.

[00258] The disclosure also provides a method of identifying a peptide binding moiety that binds a pMHC complex of the disclosure, the method comprising contacting a candidate peptide binding moiety with the pMHC complex and determining whether the candidate peptide binding moiety binds the complex. Methods to determine binding to pMHC complexes include, for example, surface plasmon resonance, or any other biosensor technique, ELISA, flow cytometry, chromatography, microscopy. Alternatively, or in

addition, binding may be determined by functional assays in which a biological response is detected upon binding, for example, cytokine release or cell apoptosis.

[00259] The candidate peptide binding moiety may be a peptide binding moiety of the type already described, such as an antibody or a TCR.

[00260] For example, antibodies and TCRs may be obtained from display libraries in which the pMHC complex of the disclosure is used to pan the library. TCRs can be displayed on the surface of phage particles and yeast particles, for example, and such libraries have been used for the isolation of high affinity variants of TCR derived from T cell clones. TCR phage libraries can be used to isolate TCRs with novel antigen specificity. Such libraries can be constructed with α- and β- chain sequences corresponding to those found in a natural repertoire. However, the random combination of these α- and β- chain sequences, which occurs during library creation, can produce a repertoire of TCRs that may not be naturally occurring.

[00261] In some embodiments, the pMHC complex of the disclosure may be used to screen a library of diverse TCRs displayed on the surface of phage particles. The TCRs displayed by said library may not correspond to those contained in a natural repertoire, for example, they may contain α- and β- chain pairing that would not be present *in vivo*, and/or the TCRs may contain non-natural mutations and/or the TCRs may be in soluble form. Screening may involve panning the phage library with pMHC complexes of the disclosure and subsequently isolating bound phage particles. For this purpose, pMHC complexes may be attached to a solid support, such as a magnet bead, or column matrix and phage bound pMHC complexes isolated, with a magnet, or by chromatography, respectively. The panning steps may be repeated several times. Isolated phage may be further expanded in *E. coli* cells. Isolated phage particles may be tested for specific binding to pMHC complexes of the disclosure. Binding can be detected using techniques including, but not limited to, ELISA, or SPR for example using a BiaCore instrument. The DNA sequence of the T cell receptor displayed by pMHC binding phage can be further identified by PCR methods.

[00262] Alternatively, antigen binding T cells and TCRs can be isolated from fresh blood obtained from patients or healthy donors. Such a method involves stimulating T cells using autologous dendritic cells (DCs), followed by autologous B cells, and then pulsed

with a peptide disclosed herein. Several rounds of stimulation may be carried out, for example three or four rounds. Activated T cells may then be tested for specificity by measuring cytokine release in the presence of T2 cells pulsed with the peptide of the disclosure (for example using an IFNγ ELISpot assay). Activated cells may then be sorted by fluorescence-activated cell sorting (FACS) using labelled antibodies to detect intracellular cytokine production (e.g. IFNγ), or expression of a cell surface marker (such as CD137). Sorted cells may be expanded and further validated, for example, by ELISpot assay and/or cytotoxicity against target cells and/or staining by peptide-MHC tetramer. The TCR chains from validated clones may then be amplified by rapid amplification of cDNA ends (RACE) and sequenced.

[00263] A peptide binding moiety disclosed herein can include, for example, without limitation, an antibody, a TCR, or a CAR.

[00264] In some embodiments, the peptide binding moiety of the disclosure may be an antibody or antigen-binding fragment thereof. Antibodies or antigen-binding fragments thereof encompass derivatives, functional equivalents, and homologues of antibodies, humanized antibodies, including any polypeptide comprising an immunoglobulin binding domain, whether natural or wholly or partially synthetic and any polypeptide or protein having a binding domain which is, or is homologous to, an antibody binding domain. Chimeric molecules comprising an immunoglobulin binding domain, or equivalent, fused to another polypeptide are therefore included. A humanized antibody may be a modified antibody having the variable regions of a non-human, e.g. murine, antibody, and the constant region of a human antibody. Examples of antibodies are the immunoglobulin isotypes (e.g., IgG, IgE, IgM, IgD and IgA) and their isotypic subclasses; or fragments that comprise an antigen binding domain such as Fab, scFv, Fv, dAb, Fd; and diabodies. Antibodies may be polyclonal or monoclonal. A monoclonal antibody may be referred to herein as "mAb".

[00265] In some embodiments, the antibody is a multispecific antibody. In some embodiments, the antibody is a bispecific antibody. The bispecific antibody may comprise a second targeting moiety that targets to the desired cell or tissue or to another desired antigen associated with the same or similar disease or disorder.

[00266] It is possible to take an antibody, for example a monoclonal antibody, and use recombinant DNA technology to produce other antibodies or chimeric molecules which retain the specificity of the original antibody. Such techniques may involve introducing DNA encoding the immunoglobulin variable region, or the complementary determining regions (CDRs), of an antibody to the constant regions, or constant regions plus framework regions, of a different immunoglobulin. A hybridoma (or other cell that produces antibodies) may be subject to genetic mutation or other changes, which may or may not alter the binding specificity of antibodies produced.

[00267] It has been shown that fragments of a whole antibody can perform the function of binding antigens. Examples of binding fragments are (i) the Fab fragment consisting of VL, VH, CL and CH1 domains; (ii) the Fd fragment consisting of the VH and CH1 domains; (iii) the Fv fragment consisting of the VL and VH domains of a single antibody; (iv) the dAb fragment which consists of a VH domain; (v) isolated CDR regions; (vi) F(ab′)2 fragments, a bivalent fragment comprising two linked Fab fragments; (vii) single chain Fv molecules (scFv), wherein a VH domain and a VL domain are linked by a peptide linker which allows the two domains to associate to form an antigen binding site; (viii) bispecific single chain Fv dimers, (ix) “diabodies”, multivalent or multispecific fragments constructed by gene fusion, and (x) VHH or VNAR antibodies, also known as single-domain antibodies or nanobodies (Nb), which may be derived from heavy-chain antibodies from e.g., dromedaries, camels, llamas, alpacas, or sharks.

[00268] Diabodies are multimers of polypeptides, each polypeptide comprising a first domain comprising a binding region of an immunoglobulin light chain and a second domain comprising a binding region of an immunoglobulin heavy chain, the two domains being linked (e.g. by a peptide linker) but unable to associate with each other to form an antigen binding site: antigen binding sites are formed by the association of the first domain of one polypeptide within the multimer with the second domain of another polypeptide within the multimer (WO94/13804). Where bispecific antibodies are to be used, these may be conventional bispecific antibodies, which can be manufactured in a variety of ways, e.g. prepared chemically or from hybrid hybridomas, or may be any of the bispecific antibody fragments mentioned above. It may be preferable to use scFv dimers or diabodies rather than whole antibodies. Diabodies and scFv can be constructed without an Fc region, using

only variable domains, potentially reducing the effects of anti-idiotypic reaction. Other forms of bispecific antibodies include the single chain "Janusins". Bispecific diabodies, as opposed to bispecific whole antibodies, may also be useful because they can be readily constructed and expressed in *E. coli*. Diabodies (and many other polypeptides such as antibody fragments) of appropriate binding specificities can be readily selected using phage display from libraries. If one arm of the diabody is to be kept constant, for instance, with a specificity directed against an antigen of interest, then a library can be made where the other arm is varied, and an antibody of appropriate specificity selected. An "antigen binding domain" is the part of an antibody which comprises the area which specifically binds to and is complementary to part or all of an antigen. Where an antigen is large, an antibody may only bind to a particular part of the antigen, which part is termed an epitope. An antigen binding domain may be provided by one or more antibody variable domains. An antigen binding domain may comprise an antibody light chain variable region (VL) and an antibody heavy chain variable region (VH).

[00269] In some embodiments, the peptide binding moiety may be an antibody-like molecule that has been designed to specifically bind a peptide or peptide-MHC complex of the disclosure. In some embodiments the peptide binding moiety may comprise a TCR-mimic antibody. In some embodiments, such TCR-mimic antibodies can comprise high-affinity soluble antibody molecules endowed with a TCR-like specificity towards tumor or viral epitopes that can target tumor and/or virus-infected cells and mediate their specific killing.

[00270] Also encompassed within the present disclosure are binding moieties based on engineered protein scaffolds or "alternative scaffolds". Alternative scaffolds are derived from stable, soluble, natural protein structures which have been modified to provide a binding site for a target molecule of interest. Examples of alternative scaffolds include, but are not limited to, affibodies, which are based on the Z-domain of staphylococcal protein A that provides a binding interface on two of its α-helices; anticalins, derived from lipocalins, that incorporate binding sites for small ligands at the open end of a β-barrel fold; monobodies, designed to incorporate the fibronectin type III domain (Fn3) of fibronectin or tenascin as a protein scaffold or synthetic FN3 domains (e.g., tencon); nanobodies, and DARPins. Additional alternative scaffolds include Adnectin™, iMab, EETI-II/AGRP,

Kunitz domain, thioredoxin peptide aptamer, Affilin, Tetranectin, Fynomer, and Avimer. Alternative scaffolds are typically targeted to bind the same antigenic proteins as antibodies, and are potential therapeutic agents. They may act as inhibitors or antagonists, or as delivery vehicles to target molecules, such as toxins, to a specific tissue *in vivo*. Short peptides may also be used to bind a target protein. Phylomers are natural structured peptides derived from bacterial genomes. Such peptides represent a diverse array of protein structural folds and can be used to inhibit/disrupt protein-protein interactions *in vivo*.

[00271] Alternative scaffolds are typically single chain polypeptidic frameworks that contain a highly structured core associated with variable domains of high conformational tolerance allowing insertions, deletions, or other substitutions within the variable domains. Libraries introducing diversity to one or more variable domains, and in some cases to the structured core, may be generated using known protocols and the resulting libraries may be screened for binding to the peptide and/or the pMHC complex of the disclosure, and the identified binders may be further characterized for their specificity using known methods. Alternative scaffolds may be derived from Protein A, in particular, the Z-domain thereof (affibodies), ImmE7 (immunity proteins), BPTI/APPI (Kunitz domains), CTLA-4, charybdotoxin (Scorpion toxin), Min-23 (knottins), lipocalins (anticalins), Ras-binding protein AF-6 (PDZ-domains), neokarzinostatin, a fibronectin domain, an ankyrin consensus repeat domain, or thioredoxin.

[00272] In some embodiments, the antibodies or alternative scaffolds described herein can be immobilized on viral vectors. Such modified recombinant viral vectors can be useful for the targeted introduction of genetic materials encoded by the viral vectors into cells and/or tissues. Various means can be used to mobilize the antibodies or alternative scaffolds to the viral vectors, for example, by using an affinity binding pair, such as c-Myc/anti-Myc antibody, streptavidin/biotin, or via spy-tag/spy-catcher system. Exemplary vectors that may be modified with the antibodies or alternative scaffolds described herein include, but are not limited to, adeno-associated virus (AAV) vectors (e.g., AAV1, AAV2, AAV6, AAV9, or AAV9.PHP), retroviral vectors, lentiviral vectors, and targeted oncolytic viruses (e.g., herpes simplex virus (HSV)).

[00273] In some embodiments, the peptide binding moiety may be a TCR. TCRs are described using the International Immunogenetics (IMGT) TCR nomenclature, and the IMGT public database of TCR sequences.

[00274] The TCRs of the present disclosure may be in any format. For example, the TCRs may be αβ heterodimers, or αα or ββ homodimers.

[00275] α/β heterodimeric TCRs have an α-chain and a β-chain. Broadly, each chain comprises variable, joining and constant region, and the β-chain also usually contains a short diversity region between the variable and joining regions, but this diversity region is often considered as part of the joining region. Each variable region comprises three hypervariable CDRs (Complementarity Determining Regions) embedded in a framework sequence; CDR3 is believed to be the main mediator of antigen recognition. There are several types of α- chain variable (Vα) regions and several types of β-chain variable (Vβ) regions distinguished by their framework, CDR1 and CDR2 sequences, and by a partly defined CDR3 sequence.

[00276] The TCRs of the disclosure may not correspond to TCRs as they exist in nature. For example, they may comprise α- and β- chain combinations that are not present in a natural repertoire. Alternatively or additionally, a TCR described herein may be soluble, and/or the α- and/or β- chain constant domain may be truncated relative to the native/naturally occurring TRAC/TRBC sequences such that, for example, the C-terminal transmembrane domain and intracellular regions are not present. Such truncation may result in removal of the cysteine residues from TRAC/TRBC that form the native interchain disulfide bond.

[00277] In addition, the TRAC/TRBC domains may contain modifications. For example, the α-chain extracellular sequence may include a modification relative to the native/naturally occurring TRAC whereby amino acid T48 of TRAC, with reference to IMGT numbering, is replaced with C48. Likewise, the β-chain extracellular sequence may include a modification relative to the native/naturally occurring TRBC1 or TRBC2 whereby S57 of TRBC1 or TRBC2, with reference to IMGT numbering, is replaced with C57. These cysteine substitutions relative to the native α- and β- chain extracellular sequences enable the formation of a non-native interchain disulfide bond which stabilizes the refolded soluble TCR, i.e. the TCR formed by refolding extracellular α- and β- chains.

This non-native disulfide bond facilitates the display of correctly folded TCRs on phage. In addition, the use of the stable disulfide linked soluble TCR enables more convenient assessment of binding affinity and binding half-life. Alternative positions for the formation of a non-native disulfide include, for example, Thr 45 of exon 1 of TRAC*01 and Ser 77 of exon 1 of TRBC1*01 or TRBC2*01; Tyr 10 of exon 1 of TRAC*01 and Ser 17 of exon 1 of TRBC1*01 or TRBC2*01; Thr 45 of exon 1 of TRAC*01 and Asp 59 of exon 1 of TRBC1*01 or TRBC2*01; and Ser 15 of exon 1 of TRAC*01 and Glu 15 of exon 1 of TRBC1*01 or TRBC2*01. TCRs with a non-native disulfide bond may be full length or may be truncated.

[00278] TCRs of the disclosure may be in single chain format. Single chain TCRs include αβ TCR polypeptides of the type: Vα-L-Vβ, Vβ-L-Vα, Vα-Cα-L-Vβ, Vα-L-Vβ-Cβ or Vα-Cα-L-Vβ-Cβ, optionally in the reverse orientation, wherein Vα and Vβ are TCR α and β variable regions respectively, Cα and Cβ are TCR α and β constant regions respectively, and L is a linker sequence. Single chain TCRs may contain a non-native disulfide bond. The TCR may be in a soluble form (i.e. having no transmembrane or cytoplasmic domains) or may contain full length α- and β- chains. The TCR may be provided on the surface of a cell, such as a T cell.

[00279] TCRs of the disclosure may be engineered to include mutations. Methods for producing mutated high affinity TCR variants such as phage display and site directed mutagenesis. Preferably, mutations to improve affinity are made within the variable regions of α- and/or β- chains. More preferably mutations to improve affinity are made within the CDRs. There may be between 1 and 15 mutations in the α- and or β- chain variable regions.

[00280] TCRs of the disclosure may also be labeled with an imaging compound, for example a label that is suitable for diagnostic purposes. Such labelled high affinity TCRs are useful in a method for detecting a TCR ligand selected from CD1-antigen complexes, bacterial superantigens, and MHC-peptide/superantigen complexes, which method comprises contacting the TCR ligand with a high affinity TCR (or a multimeric high affinity TCR complex) which is specific for the TCR ligand; and detecting binding to the TCR ligand. In multimeric high affinity TCR complexes (formed, for example, using biotinylated heterodimers) fluorescent streptavidin can be used to provide a detectable label. A fluorescently-labelled multimer is suitable for use in FACS analysis, for example

to detect antigen presenting cells carrying the peptide for which the high affinity TCR is specific.

[00281] A TCR of the present disclosure (or multivalent complex thereof) may alternatively or additionally be associated with (e.g. covalently or otherwise linked to) a therapeutic agent which may be, for example, a toxic moiety for use in cell killing, or an immunostimulating agent such as an interleukin or a cytokine. A multivalent high affinity TCR complex of the present disclosure may have enhanced binding capability for a TCR ligand compared to a non-multimeric wild-type or high affinity T cell receptor heterodimer. Thus, the multivalent high affinity TCR complexes according to the disclosure are particularly useful for tracking or targeting cells presenting particular antigens *in vitro* or *in vivo*, and are also useful as intermediates for the production of further multivalent high affinity TCR complexes having such uses. The high affinity TCR or multivalent high affinity TCR complex may therefore be provided in a pharmaceutically acceptable formulation for use *in vivo*.

[00282] High affinity TCRs of the disclosure may be used in the production of soluble bi-specific reagents. A preferred embodiment is a reagent which comprises a soluble TCR, fused via a linker to an anti-CD3 specific antibody fragment.

[00283] In a further aspect, the disclosure provides nucleic acid encoding the TCR of the disclosure, a TCR expression vector comprising nucleic acid encoding a TCR of the disclosure, as well as a cell harboring such a vector. The TCR may be encoded either in a single open reading frame or two distinct open reading frames. Also included in the scope of the disclosure is a cell harboring a first expression vector which comprises nucleic acid encoding an α- chain of a TCR of the disclosure, and a second expression vector which comprises nucleic acid encoding a β-chain of a TCR of the disclosure. Alternatively, one vector may encode both an α- and a β- chain of a TCR of the disclosure.

[00284] A further aspect of the present disclosure provides a cell displaying on its surface a TCR of the disclosure. The cell may be a T cell, or other immune cell. The T cell may be modified such that it does not correspond to a T cell as it exists in nature. For example, the cell may be transfected with a vector encoding a TCR of the disclosure such that the T cell expresses a further TCR in addition to the native TCR. Additionally or alternatively, the T cell may be modified such that it is not able to present the native TCR.

There are a number of methods suitable for the transfection of T cells with DNA or RNA encoding the TCRs of the disclosure. As a non-limiting example, the transfection method may comprise a rapid RNA-based transfection system. T cells expressing the TCRs of the disclosure are suitable for use in adoptive therapy-based treatment of diseases such as cancers. There are a number of suitable methods by which adoptive therapy can be carried out. For example, adoptive cell therapy (ACT) may comprise use of autologous tumor-infiltrating lymphocytes, and may include a lymphodepletion preparative regimen prior to ACT. In some embodiments, viruses, e.g., retroviruses, that encode TCRs may be used for genetic modification of lymphocytes to convert normal lymphocytes into lymphocytes with anti-cancer activity. The adoptive transfer of lymphocytes with anti-cancer activity into patients requiring treatment of, e.g., metastatic melanoma, can mediate tumor regression. In some embodiments, ACT may comprise treatment of patients with cancers expressing viral or alloantigens, treatment of patients with cancers expressing viral antigens, and/or ACT using gene-modified lymphocytes. In some embodiments, ACT methods may include, for example, genetic modification of lymphocytes to introduce new recognition specificities using, e.g., αβTCR(s) and/or chimeric TCR(s); genetic modification of lymphocytes to alter function of T cells using, e.g., co-stimulatory molecules (e.g., CD28, 41BB), cytokines (e.g., IL2, IL15), homing molecules (e.g., CD62L, CCR7), and/or molecules capable of preventing apoptosis (BCL2); modification of host lymphodepletion using, e.g., selective depletion of CD4+ cells or T regulatory cells; blocking of inhibitory signals on reactive lymphocytes using, e.g., antibodies to CTLA4 and/or PD-1; administration of vaccines to stimulate transferred cells using, e.g., recombinant virus encoding antigen(s); administration of alternative cytokines to support cell growth using, e.g., IL15 and/or IL21; stimulation of APCs using, e.g., toll-like receptor agonists; generation of less differentiated lymphocytes using, e.g., alternate culture conditions and growth promoting cytokines *in vitro*; and, overcoming antigen escape variants using, e.g., natural killer cells.

[00285] The TCRs of the disclosure intended for use in adoptive therapy are generally glycosylated when expressed by the transfected T cells. The glycosylation pattern of transfected TCRs may be modified by mutations of the transfected gene.

[00286] In some embodiments, the peptide binding moiety may be a chimeric antigen receptor (CAR). CARs are genetically engineered receptors. CARs may be generated that bind the peptides or pMHC complexes of the present disclosure by incorporating an antigen binding domain that specifically binds the peptide or pMHC complex to the extracellular domain of the CAR. CARs may be introduced into and expressed by immune cells, such as T cells, NK cells, or macrophages. CARs can be programmed to both recognize a specific antigen and, when bound to that antigen, activate the immune cell to attack and destroy the cell presenting that antigen. When these antigens exist on tumor cells, an immune cell that expresses the CAR can target and kill the tumor cell.

[00287] The general structure of a CAR typically comprises an extracellular domain that binds the antigen (e.g. the peptides or pMHC complexes of the present disclosure), a hinge, a transmembrane domain, and an intercellular domain comprising a signaling domain and optionally one or more co-stimulatory domains.

[00288] Extracellular domains of the CAR may contain any polypeptide that specifically binds the desired antigen (e.g. the peptides or pMHC complexes of the present disclosure). For example, the extracellular domain may comprise an antibody fragment such as scFv or VHH. The CARs may also be engineered to bind two or more desired antigens that may be arranged in tandem and separated by linker sequences. For example, one or more domain antibodies, scFvs, llama VHH antibodies or other VH only antibody fragments may be organized in tandem via a linker to provide bispecificity or multispecificity to the CAR.

[00289] A hinge domain may be present between the extracellular domain and the transmembrane domain of the CAR, e.g., to provide flexibility to allow effective binding of the extracellular domain to its intended target. The hinge domain may be a polypeptide of about 2 to 100 amino acids in length. The hinge may include or be composed of flexible residues such as Gly and Ser so that the adjacent protein domains are free to move relative to one another. Longer hinges may be used when it is desirable to ensure that two adjacent domains do not sterically interfere with one another. The hinge may be derived from a hinge region or portion of the hinge region of any immunoglobulin. Non-limiting examples

of linkers include a part of human CD8α chain, extracellular domain of CD28, an Ig hinge from IgG, IgM, IgA, IgD, or IgE, FcyRllla receptor, or a functional fragment thereof.

[00290] Transmembrane domains of the CAR may be derived transmembrane proteins, such as an alpha, beta or zeta chain of a T-cell receptor, CD28, CD3 epsilon, CD2, CD4, CD5, CD8, CD9, CD16, CD18, CD19, CD22, CD27, CD29, CD33, CD37, CD40, CD45, CD49a, CD64, CD80, CD84, CD86, CD96 (Tactile), CD100 (SEMA4D), CD103, CD134, CD154, CD160 (BY55), KIRDS2, OX40, LFA-1 (CD11a, CD18), CD11b, CD11c, CD11d, ICOS (CD278), 4-1 BB (CD137), 4-1 BBL, GITR, BAFFR, HVEM (LIGHTR), SLAMF7, NKp80 (KLRFI), IL2R beta, IL2R gamma, IL7Ra, ITGA1, VLA1, ITGA4, IA4, CD49D, ITGA6, VLA-6, CD49f, ITGAD, ITGAE, ITGAL, LFA-1, ITGAM, ITGAX, ITGB1, ITGB2, LFA-1, ITGB7, TNFR2, DNAM1 (CD226), SLAMF4 (CD244, 2B4), CEACAM1, CRT AM, Ly9 (CD229), PSGL1, SLAMF6 (NTB-A, Ly108), SLAM (SLAMF1, CD150, IPO-3), BLAME (SLAMF8), SELPLG (CD162), LTBR, PAG/Cbp, NKp30, NKp44, NKp46, NKG2D, and NKG2C, or functional fragment thereof.

[00291] The intracellular signaling domain of a CAR participates in transducing the signal of effective CAR binding to a target antigen into the interior of the immune effector cell to elicit an effector cell function, e.g., activation, cytokine production, proliferation, and cytotoxic activity, including the release of cytotoxic factors to the CAR-bound target cell, or other cellular responses elicited following antigen binding to the extracellular CAR domain. Non-limiting examples of intracellular signaling domains of the CAR include those derived from CD3ζ, CD3ε, CD3δ, CD3γ, CD5, CD22, CD39, CD79A, CD79B, CD66d, CD226, DAP10, DAP12, Fc epsilon receptor I gamma chain (FCER1G), or FcR β.

[00292] Intracellular co-stimulatory domains of the CAR can provide a second signal required for efficient activation and function of T lymphocytes upon binding to antigen. Such co-stimulatory domains may be derived from one or more co-stimulatory molecules, such as, but not limited to, 4-1BB, CD2, CD7, CD27, CD28, CD30, CD40, CD54 (ICAM), CD83, CD134 (OX40), CD150 (SLAMF1), CD152 (CTLA4), CD223 (LAG3), CD270 (HVEM), CD278 (ICOS), DAP10, LAT, NKD2C SLP76, TRIM, BTLA, GITR, CD226, HVEM, and ZAP70.

[00293] The CARs can be generated by standard molecular biology techniques. The extracellular domain that binds the desired antigen may be derived from the antibodies or their antigen binding fragments described herein.

[00294] In another aspect, the disclosure further provides cells that comprise peptide a binding moiety (e.g., TCRs and CARs) of the present disclosure. In some embodiments, the host cell is an immune cell. In some embodiments, the immune cell is T cell, NK cell, or a macrophage. The host cell may be autologous or allogeneic with respective to the subject receiving the cell (as treatment).

[00295] In some embodiments, TCRs of the present disclosure are provided as TCR-T cells. In some embodiments, CARs of the present disclosure are provided as CAR-T cells. Any methods known in the art for modifying T cells to express a TCR or CAR can be employed to generate the TCR-T or CAR-T cells of the present disclosure.

[00296] The cells expressing a peptide binding moiety (e.g., TCRs and CARs) of the present disclosure may also contain one or more additional genes. The additional genes can be used to increase the effector function of the cells expressing the peptide binding moiety (e.g., TCRs and CARs). Non-limiting examples of classes of additional genes include (a) a second targeting moiety, such as antibodies, including fragments thereof and bispecific antibodies (e.g., bispecific T cell engagers (BiTEs)), (b) secretable cytokines (e.g., GM-CSF, IL-7, IL-12, IL-15, IL-18), (c) membrane bound cytokines (e.g., IL-15), (d) chimeric cytokine receptors (e.g., IL-2/IL-7, IL-4/IL-7), (e) constitutive active cytokine receptors (e.g., C7R), (f) dominant negative receptors (DNR; e.g., TGFRII DNR), (g) ligands of co-stimulatory molecules (e.g., CD80, 4-1BBL), (h) nuclear factor of activated T cells (NFATs) (e.g., NFATc1, NFATc2, NFATc3, NFATc4, and NFAT5), or (j) suicide genes (e.g., CD20, truncated EGFR or HER2, inducible caspase 9 molecules). In some embodiments, the cells expressing a peptide binding moiety (e.g., TCRs and CARs) of the present disclosure may express a second targeting moiety that targets to a specific tissue and/or cell type or to a known cancer antigen.

Pharmaceutical Compositions, Dosage Forms, and Administration

[00297] In a further aspect, the disclosure provides a pharmaceutical composition comprising a peptide, a peptide-based molecule (such as a complex (e.g., peptide-MHC

(pMHC) complex), fusion protein, or conjugate comprising the peptide), a nucleic acid molecule, a vector, a cell, or a peptide binding moiety of the disclosure together with a pharmaceutically acceptable carrier and/or excipient. The pharmaceutical compositions of the disclosure may be in any suitable form (depending upon the desired method of administering to a patient). Suitable compositions and methods of administration are known to those skilled in the art, for example see, Johnson et al., Blood. 2009; 114(3):535-46.

[00298] The pharmaceutical compositions may comprise the peptides or peptide-based molecules of the disclosure either in the free form or in the form of a pharmaceutically acceptable salt. The term "pharmaceutically acceptable salt" as used herein refers to a derivative of the disclosed peptides wherein the peptide is modified by making acid or base salts of the agent. For example, acid salts are prepared from the free base (typically wherein the neutral form of the drug has a neutral —NH_2 group) involving reaction with a suitable acid. Suitable acids for preparing acid salts include both organic acids, e.g., acetic acid, benzoic acid, citric acid, propionic acid, glycolic acid, trifluoroacetic acid, pyruvic acid, oxalic acid, malic acid, malonic acid, maleic acid, succinic acid, fumaric acid, tartaric acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, p-toluenesulfonic acid, salicylic acid, and the like, as well as inorganic acids, e.g., hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid phosphoric acid and the like. Conversely, preparation of basic salts of acid moieties which may be present on a peptide are prepared using a pharmaceutically acceptable base such as sodium hydroxide, potassium hydroxide, ammonium hydroxide, calcium hydroxide, trimethylamine or the like.

[00299] Compositions of the disclosure may comprise multiple peptides, e.g., 2 to 50, 2 to 40, 2 to 30, 5 to 25, 5 to 20, or 10 to 15 peptides as described herein (e.g., SEQ ID NO: 1-89 and 143-146). In some embodiment, the compositions of the disclosure may comprise 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, or 93 amino acid

sequences selected from SEQ ID NO: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or fragment or derivative thereof.

[00300] In some embodiments, the peptides or peptide-based molecules may be present in a solution at a concentration of about 1 μg/mL to 50 mg/mL, for example, about 0.1 mg/mL to 10 mg/mL, about 0.2 mg/mL to 5 mg/mL, about 0.5 mg/mL to 8 mg/mL, about 0.8 mg/mL to 12 mg/mL, about 1 mg/mL to 15 mg/mL, about 2 mg/mL to 20 mg/mL, or about 5 mg/mL to 25 mg/mL, or about 0.1 mg/mL, 0.2 mg/mL, 0.3 mg/mL, 0.4 mg/mL, 0.5 mg/mL, 0.6 mg/mL, 0.7 mg/mL, 0.8 mg/mL, 0.9 mg/mL, 1 mg/mL, 1.25 mg/mL, 1.5 mg/mL, 1.75 mg/mL, 2 mg/mL, 2.25 mg/mL, 2. 5 mg/mL, 2.75 mg/mL, 3 mg/mL, 3.25 mg/mL, 3. 5 mg/mL, 3.75 mg/mL, 4 mg/mL, 5 mg/mL, 6 mg/mL, 7 mg/mL, 8 mg/mL, 9 mg/mL, 10 mg/mL, 11 mg/mL, 12 mg/mL, 13 mg/mL, 14 mg/mL, 15 mg/mL or 20 mg/mL.

[00301] The pharmaceutical composition may be adapted for administration by any appropriate route such as, e.g., parenteral (including subcutaneous, intramuscular, or intravenous), enteral (including oral or rectal), inhalation, or intranasal routes.

[00302] Such compositions may be prepared by any method known in the art of pharmacy, for example, by mixing the active ingredient with the carrier(s) or excipient(s) under sterile conditions.

[00303] In addition, disclosed herein are pharmaceutical dosage forms comprising the peptides, peptide-based molecules (such as complexes (e.g., peptide-MHC (pMHC) complexes), fusion proteins, or conjugates comprising the peptide(s)), nucleic acid molecules, vectors, cells, or binding moieties of the disclosure.

[00304] Pharmaceutical compositions based on the peptides, peptide-based molecules (such as complexes (e.g., peptide-MHC (pMHC) complexes), fusion proteins, or conjugates comprising the peptide(s)), nucleic acid molecules, vectors, cells, or binding moieties disclosed herein can be formulated in any conventional manner using one or more physiologically acceptable carriers and/or excipients. The peptides, peptide-based molecules (such as complexes (e.g., peptide-MHC (pMHC) complexes), fusion proteins, or conjugates comprising the peptide(s)), nucleic acid molecules, vectors, cells, or binding moieties may be formulated for administration by, for example, injection, inhalation, or insulation (either through the mouth or the nose) or by oral, buccal, parenteral or rectal administration, or by administration directly to an organ or tissue.

[00305] The pharmaceutical compositions can be formulated for a variety of modes of administration, including systemic, topical, or localized administration. Techniques and formulations can be found in, for example, Remington's Pharmaceutical Sciences, Meade Publishing Co., Easton, Pa. For systemic administration, injection is preferred, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For the purposes of injection, the pharmaceutical compositions can be formulated in liquid solutions, preferably in physiologically compatible buffers, such as Hank's solution or Ringer's solution. In addition, the pharmaceutical compositions may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms of the pharmaceutical composition are also suitable.

[00306] In some embodiments, the pharmaceutical compositions of the present disclosure may be lyophilized. As a non-limiting example, the obtained lyophilizate can be reconstituted into a hydrous composition by adding a hydrous solvent. In some embodiments, the hydrous composition may be able to be directly administered parenterally to a patient. Therefore, a further embodiment of the present disclosure is a hydrous pharmaceutical composition, obtainable via reconstitution of the lyophilizate with a hydrous solvent.

[00307] In some embodiments, the pharmaceutical composition disclosed herein may comprise a lyophilized formulation. As a non-limiting example, the lyophilization formulation may comprise peptides of the disclosure, mannitol, and/or TWEEN 80®. As another non-limiting example, the lyophilization formulation may comprise the peptides disclosed herein, mannitol and poloxamer 188. In some embodiments, the pharmaceutical composition may comprise a lyophilization formulation comprising a reconstituted-liquid composition.

[00308] In some embodiments, pharmaceutical compositions of the present disclosure may provide a formulation with an enhanced solubility and/or moistening of the lyophilizate over previously known compositions. As a non-limiting example, enhanced solubility and/or moistening of the lyophilizate may be achieved using an appropriate composition of excipients. In this way, pharmaceutical compositions of the present disclosure comprising peptides of SEQ ID NO: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof, may be developed to show a

desired shelf stability at (e.g., at −20° C, +5° C, or +25° C) and can be easily resolubilized such that the lyophilizate can be completely dissolved through the use of a buffer or other excipients from seconds up to two or more minutes, with or without the use of an of ultrasonic homogenizer. Furthermore, the composition can be easily provided to a patient in need of treatment via any appropriate delivery route disclosed herein, e.g., parenteral (including subcutaneous, intramuscular, or intravenous), enteral (including oral or rectal), inhalation, or intranasal routes. As a non-limiting example, the pH-value of the resulting solution may be between pH 2.7 and pH 9.

[00309] For oral administration, the pharmaceutical compositions may take the form of, for example, tablets or capsules prepared by conventional means with pharmaceutically acceptable excipients such as binding agents (e.g. pregelatinized maize starch, polyvinylpyrrolidone or hydroxypropyl methylcellulose); fillers (e.g. lactose, microcrystalline cellulose or calcium hydrogen phosphate); lubricants (e.g. magnesium stearate, talc or silica); disintegrants (e.g. potato starch or sodium starch glycolate); or wetting agents (e.g. sodium lauryl sulfate). The tablets can also be coated by methods well known in the art. Liquid preparations for oral administration may take the form of, for example, solutions, syrups or suspensions, or they may be presented as a dry product for constitution with water or other suitable vehicle before use. Such liquid preparations may be prepared by conventional means with pharmaceutically acceptable additives such as suspending agents (e.g. sorbitol syrup, cellulose derivatives or hydrogenated edible fats); emulsifying agents (e.g. lecithin or acacia); non-aqueous vehicles (e.g. ationd oil, oily esters, ethyl alcohol or fractionated vegetable oils); and preservatives (e.g. methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations can also contain buffer salts, flavoring, coloring and sweetening agents as appropriate.

[00310] The pharmaceutical compositions can be formulated for parenteral administration by injection, e.g. by bolus injection or continuous infusion. Formulations for injection can be presented in a unit dosage form, e.g. in ampoules or in multi-dose containers, with an optionally added preservative. The pharmaceutical compositions can further be formulated as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain other agents including suspending, stabilizing and/or dispersing agents.

[00311] Additionally, the pharmaceutical compositions can also be formulated as a depot preparation. These long acting formulations can be administered by implantation (e.g. subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (e.g. as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt. Other suitable delivery systems include microspheres, which offer the possibility of local noninvasive delivery of drugs over an extended period of time. This technology can include microspheres having a precapillary size, which can be injected via a coronary catheter into any selected part of an organ without causing inflammation or ischemia. The administered therapeutic is men slowly released from the microspheres and absorbed by the surrounding cells present in the selected tissue.

[00312] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, bile salts, and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration can occur using nasal sprays or suppositories. For topical administration, the vector particles described herein can be formulated into ointments, salves, gels, or creams as generally known in the art. A wash solution can also be used locally to treat an injury or inflammation in order to accelerate healing.

[00313] Pharmaceutical forms suitable for injectable use can include sterile aqueous solutions or dispersions; formulations including sesame oil, peanut oil or aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In all cases, the form must be sterile and must be fluid. It must be stable under the conditions of manufacture and certain storage parameters (e.g. refrigeration and freezing) and must be preserved against the contaminating action of microorganisms, such as bacteria and fungi.

[00314] If formulations disclosed herein are used as a therapeutic to boost an immune response in a subject, a therapeutic agent can be formulated into a composition in a neutral or salt form. Pharmaceutically acceptable salts, include the acid addition salts

(formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like.

[00315] A carrier can also be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be brought about by various antibacterial and antifungal agents known in the art. In many cases, it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

[00316] Sterile injectable solutions can be prepared by incorporating the active compounds or constructs in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization.

[00317] Upon formulation, solutions can be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms, such as the type of injectable solutions described above, but slow release capsules or microparticles and microspheres and the like can also be employed.

[00318] For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intratumorally, intramuscular, subcutaneous and intraperitoneal administration. In this context, sterile aqueous media that can be employed will be known to those of skill in the art in light of the present disclosure. For example, one dosage could

be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion.

[00319] The person responsible for administration will, in any event, determine the appropriate dose for the individual subject. For example, a subject may be administered peptides, peptide-based molecules (such as complexes (e.g., peptide-MHC (pMHC) complexes), fusion proteins, or conjugates comprising the peptide(s)), nucleic acid molecules, vectors, cells, or binding moieties described herein on a daily or weekly basis for a time period or on a monthly, bi-yearly or yearly basis depending on need or exposure to a pathogenic organism (e.g., HTLV-1) or to a condition in the subject (e.g. cancer).

[00320] In addition to the compounds formulated for parenteral administration, such as intravenous, intratumorally, intradermal or intramuscular injection, other pharmaceutically acceptable forms include, e.g., tablets or other solids for oral administration; liposomal formulations; time release capsules; biodegradable and any other form currently used.

[00321] One may also use intranasal or inhalable solutions or sprays, aerosols or inhalants. Nasal solutions can be aqueous solutions designed to be administered to the nasal passages in drops or sprays. Nasal solutions can be prepared so that they are similar in many respects to nasal secretions. Thus, the aqueous nasal solutions usually are isotonic and slightly buffered to maintain a pH of 5.5 to 7.5. In addition, antimicrobial preservatives, similar to those used in ophthalmic preparations, and appropriate drug stabilizers, if required, may be included in the formulation. Various commercial nasal preparations are known and can include, for example, antibiotics and antihistamines and are used for asthma prophylaxis.

[00322] Oral formulations can include excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders. In certain defined embodiments, oral pharmaceutical compositions will include an inert diluent or assimilable edible carrier, or they may be enclosed in hard or soft-shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compounds may be incorporated

with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like.

[00323] The tablets, troches, pills, capsules and the like may also contain the following: a binder, as gum tragacanth, acacia, cornstarch, or gelatin; excipients, such as dicalcium phosphate; a disintegrating agent, such as corn starch, potato starch, alginic acid and the like; a lubricant, such as magnesium stearate; and a sweetening agent, such as sucrose, lactose or saccharin may be added or a flavoring agent, such as peppermint, oil of wintergreen, or cherry flavoring. When the dosage unit form is a capsule, it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills, or capsules may be coated with shellac, sugar, or both. A syrup of elixir may contain the active compounds sucrose as a sweetening agent methyl and propylparabens as preservatives, a dye and flavoring, such as cherry or orange flavor.

[00324] Further embodiments disclosed herein can concern kits for use with methods and compositions. Kits can also include a suitable container, for example, vials, tubes, mini- or microfuge tubes, test tube, flask, bottle, syringe or other container. Where an additional component or agent is provided, the kit can contain one or more additional containers into which this agent or component may be placed. Kits herein will also typically include a means for containing the peptides, peptide-based molecules (such as complexes (e.g., peptide-MHC (pMHC) complexes), fusion proteins, or conjugates comprising the peptide(s)), nucleic acid molecules, vectors, cells, or binding moieties and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained. Optionally, one or more additional active agents may be needed for compositions described.

[00325] Dose ranges and frequency of administration can vary depending on the nature of the composition and the medical condition as well as parameters of a specific patient and the route of administration used. A dose can also depend on the subject in which it is being administered. For example, a lower dose may be required if the subject is juvenile, and a higher dose may be required if the subject is an adult human subject. In certain embodiments, a more accurate dose can depend on the weight of the subject. A

suitable, non-limiting example of a dosage of a pharmaceutical composition containing the same disclosed herein may vary depending upon the age and the size of a subject to be administered, target disease, the purpose of the treatment, conditions, route of administration, and the like. Non-limiting examples of suitable dosages include, e.g., 0.01 to about 20 mg/kg body weight, more preferably about 0.02 to about 7, about 0.03 to about 5, or about 0.05 to about 3 mg/kg body weight. Depending on the severity of the condition, the frequency and the duration of the treatment can be adjusted. In certain embodiments, the initial dose may be followed by administration of a second or a plurality of subsequent doses in an amount that can be approximately the same or less than that of the initial dose, wherein the subsequent doses are separated by at least 1 day to 3 days; at least one week, at least 2 weeks; at least 3 weeks; at least 4 weeks; at least 5 weeks; at least 6 weeks; at least 7 weeks; at least 8 weeks; at least 9 weeks; at least 10 weeks; at least 12 weeks; or at least 14 weeks.

[00326] Compositions may include administration to a subject intravenously, intratumorally, intradermally, intraarterially, intraperitoneally, intralesionally, intracranially, intraarticularly, intraprostaticaly, intrapleurally, intratracheally, intranasally, intravitreally, intravaginally, intrarectally, topically, intratumorally, intramuscularly, intrathecally, subcutaneously, subconjunctival, intravesicularlly, mucosally, intrapericardially, intraumbilically, intraocularly, orally, locally, by inhalation, by injection, by infusion, by continuous infusion, by localized perfusion, via a catheter, via a lavage, in a cream, or in a lipid composition.

[00327] Certain additional agents used in the combination therapies can be formulated and administered by any means known in the art.

[00328] Compositions as disclosed herein can also include adjuvants such as aluminum salts and other mineral adjuvants, tensoactive agents, bacterial derivatives, vehicles and cytokines. Adjuvants can also have antagonizing immunomodulating properties. For example, adjuvants can stimulate Th1 or Th2 immunity. Compositions and methods as disclosed herein can also include adjuvant therapy.

[00329] The peptides or peptides-based molecules of the disclosure may be provided in the form of a vaccine composition. The vaccine composition may be useful for the treatment, prevention, or reduction of the likelihood of HTLV-1 infection and/or HTLV-

1-induced diseases or disorders. As will be appreciated, vaccines may take several forms (see, e.g., Schlom, J Natl Cancer Inst. 2012; 104(8):599-613; Salgaller, Cancer Res. 1996; 56(20):4749-57 and Marchand, Int J Cancer. 1999; 80(2):219-30). The vaccine composition may include additional peptides or peptides-based molecules such that the peptide or peptides-based molecule of the disclosure is one of a mixture of peptides or peptides-based molecules. Adjuvants may be added to the vaccine composition to augment the immune response. In particular for peptide-containing vaccines compositions of the disclosure, pharmaceutically acceptable adjuvants include, but are not limited to, aluminum salts, Amplivax, AS 15, Aquila's QS21 stimulon, AsA404 (DMXAA), beta-glucan, BCG, CP-870,893, CpG7909, CyaA, dSLIM, GM-CSF, IC30, IC31, Imiquimod, ImuFact EV1P321, IS Patch, ISS, 1018 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, poly-ICLC, PepTel®, Pam3Cys, PLGA microparticles, resiquimod, SRL172, Virosomes and other Virus-like particles, YF-17D, VEGF trap, R848, and/or vadimezan.

[00330] Alternatively, the vaccine composition may take the form of an APC displaying the peptide of the disclosure in complex with MHC. Preferably the APC is an immune cell, more preferably a dendritic cell or a B cell. The peptide may be pulsed onto the surface of the cell (Thurner, J Exp Med. 1999; 190(11):1669-78), or nucleic acid encoding for the peptide of the disclosure may be introduced into dendritic cells or B cells (e.g., by electroporation. Van Tendeloo, Blood. 2001; 98(1):49-56).

[00331] The pharmaceutical compositions of the disclosure may be administered directly into the patient, into the affected organ or systemically i.d., i.m., s.c., i.p. and i.v., or applied *ex vivo* to cells derived from the patient or a human cell line which are subsequently administered to the patient, or used *in vitro* to select a subpopulation of immune cells derived from the patient, which are then re-administered to the patient. If the nucleic acid is administered to cells *in vitro*, it may be useful for the cells to be transfected so as to co-express immune-stimulating cytokines, such as interleukin-2. The peptide or peptide-based molecule may be substantially pure, or combined with an immune-stimulating adjuvant or used in combination with immune-stimulatory cytokines, or be administered with a suitable delivery system, e.g., liposomes, viral particles, VLPs. The

peptide or peptide-based molecule may also be conjugated to a suitable carrier such as keyhole limpet haemocyanin (KLH) or mannan (see, e.g., WO 95/18145 and Longenecker et al., 1993).

[00332] In some embodiments, the peptide-containing compositions described herein further comprise an accessory molecule which can modulate a survival or an activity of TCR-expressing cells.

[00333] Non-limiting examples of useful accessory molecules include, e.g., an anti-CD28 antibody, an anti-CD80 (B7.1) antibody, an anti-CD86 (B7.2) antibody, an anti-anti-CD3 antibody, an anti-CD2 antibody, an anti-CD4 antibody, an anti-CD8 antibody, an anti-CD47 antibody, and functional derivatives, mutants and fragments thereof.

[00334] Accessory molecules used in the peptide-containing compositions described herein include molecules that provide a signal which, in addition to the primary signal provided by, for instance, binding of a TCR/CD3 complex with a pMHC complex, mediates a T cell response, including, but not limited to, proliferation, activation, differentiation, and the like.

[00335] The accessory molecule can be, for example, an inhibitory or stimulatory antibody, a peptide ligand, a costimulatory peptide, a cytokine, etc. Non-limiting examples of accessory molecules that can be used in the peptide-containing compositions described herein include, e.g., CD7, B7.1 (CD80), B7.2 (CD86), PD-L1 , PD-L2, 4-1BBL, OX40L, Fas ligand (FasL), inducible co stimulatory ligand (ICOS-L), intercellular adhesion molecule (ICAM), CD30L, CD40, CD70, CD83, HLA-G, MICA, MICB, FIVEM, lymphotoxin β receptor, 3/TR6, ILT3, ILT4, HVEM, an agonist or antibody that binds Toll ligand receptor and a ligand that specifically binds to B7-H3 as well as antibodies that specifically bind to CD27, CD28, B7.1 (CD80), B7.2 (CD86), 4-1BB, OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD3, CD7, LIGHT, NKG2C, B7-H3, and a ligand that specifically binds to CD83.

[00336] Additional non-limiting examples of accessory molecules include, e.g., TNF/TNF family members (e.g., OX40L, ICOSL, FASL, LTA, LTB TRAIL, CD153, TNFSF9, RANKL, TWEAK, TNFSF13, TNFSF13b, TNFSF14, TNFSF15, TNFSF18, CD40LG, CD70); members of the Immunoglobulin superfamily (e.g., VISTA, PD1, PD-L1 , PD-L2, B71 , B72, CTLA4, CD28, TIM3, CD4, CD8, CD19, T cell receptor chains,

ICOS, ICOS ligand, HHLA2, butyrophilms, BTLA, B7-H3, B7-H4, CD3, CD79a, CD79b, IgSF CAMS (including CD2, CD58, CD48, CD150, CD229, CD244, ICAM-1), Leukocyte immunoglobulin like receptors (LILR), killer cell immunoglobulin like receptors (KIR)), lectin superfamily members, selectins, cytokines/chemokine and cytokine/chemokine receptors, growth factors and growth factor receptors), adhesion molecules (integrins, fibronectins, cadherins), or ecto-domains of multi-span integral membrane proteins, or antibodies directed to any of these molecules.

[00337] In some embodiments, the peptide-containing compositions described herein further comprise a cytotoxic agent. In one specific embodiment, the cytotoxic agent is a toxin or a radioactive isotope (e.g., a radioconjugate) or a suicide gene. Non-limiting examples of toxins which can be used in the peptide-containing compositions described herein include, e.g., enzymatically active toxins of bacterial, fungal, plant, or animal origin, or fragments, mutants or derivatives thereof. Enzymatically active toxins and fragments thereof that can be used include, for example, diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, α-sarcin, *Aleurites fordii* proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. Non-limiting examples of suicide genes include, e.g., thymidine kinase, cytosine deaminase, purine nucleoside phosphorylase, nitroreductase, β-galactosidase, hepatic cytochrome P450-2B1, linamarase, horseradish peroxidase, and carboxypeptidase.

[00338] Methods for introducing polypeptide or polynucleotides of the present disclosure into a cell or subject can include, for example, vector delivery, particle-mediated delivery, exosome-mediated delivery, lipid-nanoparticle-mediated delivery, cell-penetrating-peptide-mediated delivery, or implantable-device-mediated delivery. In some embodiments, a nucleic acid or protein can be introduced into a cell or subject in a carrier such as a poly(lactic acid) (PLA) microsphere, a poly(D,L-lactic-coglycolic-acid) (PLGA) microsphere, a liposome, a micelle, an inverse micelle, a lipid cochleate, or a lipid microtubule.

[00339] The use of nanoparticles to deliver the polypeptide or polynucleotides compositions of the disclosure is contemplated herein. Exemplary nanoparticles include, but are not limited to, polymeric nanoparticles, inorganic nanoparticles, liposomes, lipid nanoparticles (LNP), an immune stimulating complex (ISCOM), a virus-like particle (VLP), or a self-assembling protein. The nanoparticles may be calcium phosphate nanoparticles, silicon nanoparticles or gold nanoparticles. For examples, the polymeric nanoparticles may comprise one or more synthetic polymers, such as poly(d,l-lactide-co-glycolide) (PLG), poly(d,l-lactic-coglycolic acid) (PLGA), poly(g-glutamic acid) (g-PGA), poly(ethylene glycol) (PEG), or polystyrene or one or more natural polymers such as a polysaccharide, for example pullulan, alginate, inulin, and chitosan. The use of a polymeric nanoparticles may be advantageous due to the properties of the polymers that may be include in the nanoparticle. For instance, the natural and synthetic polymers recited above may have good biocompatibility and biodegradability, a non-toxic nature and/or the ability to be manipulated into desired shapes and sizes. The polymeric nanoparticle may also form hydrogel nanoparticles, hydrophilic three-dimensional polymer networks with favorable properties including flexible mesh size, large surface area for multivalent conjugation, high water content, and high loading capacity for antigens. Polymers such as Poly(L-lactic acid) (PLA), PLGA, PEG, and polysaccharides are suitable for forming hydrogel nanoparticles. Inorganic nanoparticles typically have a rigid structure and comprise a shell in which an antigen is encapsulated or a core to which the antigen may be covalently attached. The core may comprise one or more atoms such as gold (Au), silver (Ag), copper (Cu) atoms, Au/Ag, Au/Cu, Au/Ag/Cu, Au/Pt, Au/Pd or Au/Ag/Cu/Pd or calcium phosphate (CaP).

[00340] Other molecules suitable for complexing with the polypeptide or polynucleotides of the disclosure include cationic molecules, such as, polyamidoamine , dendritic polylysine, polyethylene irinine or polypropylene imine, polylysine, chitosan, DNA-gelatin coarcervates, DEAE dextran, dendrimers, or polyethylenimine (PEI).

[00341] In some embodiments, antibodies of the present disclosure can be conjugated to nanoparticles. Nanoparticles that may be used for conjugation with antibodies of the present disclosure include but not are limited to PEGylated liposomes, poly(d,l-lactide-co-glycolide)/montmorillonite nanoparticles (PLGA/MMT NPs),

poly(lactide-co-glycolide) (PLGA) nanoparticles, poly-(malic acid)-based nanoparticles, chitosan-shelled nanoparticles, carbon nanotubes, and other inorganic nanoparticles (such as nanoparticles made of magnesium–aluminium layered double hydroxides with disuccinimidyl carbonate (DSC), and TiO_2 nanoparticles). Nanoparticles can be developed and conjugated to an antibody contained in a pharmaceutical composition for targeting virus-infected cells.

Treatment Methods

[00342] Compositions of the present disclosure, including the peptides, peptide-based molecules (such as complexes (e.g., peptide-MHC (pMHC) complexes), fusion proteins, or conjugates comprising the peptide(s)), nucleic acid molecules, vectors, cells, or binding moieties of the disclosure, may be used in the prophylaxis and/or treatment of a viral infection (e.g., HTLV-1 infection) and/or diseases or disorders caused by the viral infection (e.g., HTLV-1 infection).

[00343] In one aspect, disclosed herein is a method for modulating an activity, proliferation or survival of a cell comprising a TCR, comprising contacting the cell with a composition (e.g., peptide, complex (e.g., pMHC complex), fusion protein, or conjugate) of the disclosure.

[00344] In some embodiments, the cell is a lymphocyte such as, e.g., a T-cell (e.g., a CD4+ T-cell or a CD8+ T-cell). In some embodiments, the target T cell is a CD4+ T cell such as, e.g., a helper T cell (e.g., a Th1, Th2, or Th17 cell) or a CD4+/CD25+/FOXP3+ regulatory T (Treg) cell. In some cases, the target T cell is a CD8+ T cell such as, e.g., a cytotoxic T cell. In some cases, the target T cell is a memory T cell, which can be a CD4+ T cell or a CD8+ T cell, where memory T cells are generally CD45RO+. In some cases, the target T cell is an NK-T cell.

[00345] In some embodiments, the contacting is *ex vivo*. In some embodiments, the contacting is *in vivo* in a subject (e.g., human).

[00346] In some embodiments, the cell is a mammalian cell (e.g., a human cell).

[00347] In some embodiments, e.g., where the target T cell is a CD8+ T cell, the peptide is presented by a class I MHC polypeptide. In some embodiments, e.g., where the target T cell is a CD4+ T cell, the peptide is presented by class II MHC polypeptides.

[00348] The interaction of a T cell with the peptides described herein can result in, e.g., activation, induction of anergy, or death of a T cell that occurs when the TCR of the T cell is bound by a TCR-binding molecule (e.g., pMHC complex). "Activation of a T cell" refers to induction of signal transduction pathways in the T cell resulting in production of cellular products (e.g., interleukin-2) by that T cell. "Anergy" refers to the diminished reactivity by a T cell to an antigen. Activation and anergy can be measured by, for example, measuring the amount of IL-2 produced by a T cell after a pMHC complex has bound to the TCR. Anergic cells will have decreased IL-2 production when compared with stimulated T cells. Another method for measuring the diminished activity of anergic T cells includes measuring intracellular and/or extracellular calcium mobilization by a T cell upon engagement of its TCR's. "T cell death" refers to the permanent cessation of substantially all functions of the T cell.

[00349] In another aspect, provided herein is a method of inducing an immune response against HTLV infection (e.g., a HTLV-1, HTLV-2, HTLV-3, and/or HTLV-4 infection) in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of a composition (e.g., one or more peptides, complexes (e.g., pMHC complex), fusion proteins, conjugates, nucleic acid molecules, vectors, cells, or binding moieties) of the present disclosure.

[00350] In certain embodiments, generating an immune response comprises an increase in target antigen-specific cytotoxic T lymphocytes (CTL) activity of about 1.5-fold to 20-fold, or more fold in a subject administered a composition of the disclosure as compared to a control. In certain embodiments, generating an immune response comprises an increase in target-specific CTL activity of about 1.5-fold to 20-fold, or more fold in a subject administered the composition of the disclosure as compared to a control. In a further embodiment, generating an immune response that comprises an increase in target antigen-specific cell-mediated immunity activity as measured by ELISpot assays measuring cytokine secretion, such as interferon-gamma (IFN-γ), interleukin-2 (IL-2), tumor necrosis

factor-alpha (TNF-α), or other cytokines, of about 1.5-fold to 20-fold, or more fold as compared to a control.

[00351] In a further embodiment, generating an immune response comprises an increase in target- specific antibody production of between 1.5-fold and 5-fold in a subject administered the composition of the disclosure as compared to an appropriate control. In another embodiment, generating an immune response comprises an increase in target-specific antibody production of about 1.5-fold to 20-fold, or more fold in a subject administered the composition of the disclosure as compared to a control.

[00352] T cell activation may be determined, e.g., by measuring changes in the level of expression of cytokines and/or T cell activation markers, and/or the induction of antigen-specific proliferating cells. Techniques known to those of skill in the art, including, but not limited to, immunoprecipitation followed by western blot analysis, ELISAs, flow cytometry, northern blot analysis, and RT-PCR can be used to measure the expression cytokines and T cell activation markers. Cytokine release may be measured by measuring secretion of cytokines including but not limited to Interleukin-2 (IL-2), Interleukin-4 (IL-4), Interleukin-6 (IL-6), Interleukin-12 (IL-12), Interleukin-16 (IL-16), PDGF, TGF-α, TGF-β, TNF-α, TNF-β, GCSF, GM-CSF, MCSF, IFN-α, IFN-β, IFN-γ, TFN-γ, IGF-I, and IGF-II.

[00353] T cell modulation may also be evaluated by measuring, e.g., proliferation by, e.g., ^{3}H-thymidine incorporation, trypan blue cell counts, and fluorescence activated cell sorting (FACS).

[00354] The anti-tumor responses of T cells may be determined in xenograft tumor models. Tumors may be established using any human cancer cell line expressing the relevant tumor associated antigen. To establish xenograft tumor models, about 5×10^{6} viable cells, may be injected, e.g., subcutaneously into nude athymic mice using for example Matrigel (Becton Dickinson). The endpoint of the xenograft tumor models can be determined based on the size of the tumors, weight of animals, survival time and histochemical and histopathological examination of the cancer, using methods known to one skilled in the art.

[00355] In a related aspect, disclosed herein is a method of treating an HTLV-1 infection in a subject in need thereof, the method comprising administering to the subject

an effective amount of a composition (e.g., one or more peptides, complexes (e.g., pMHC complex), fusion proteins, conjugates, nucleic acid molecules, vectors, cells, binding moieties) of the present disclosure.

[00356] In a related aspect, disclosed herein is a method of preventing or reducing the likelihood of an HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of a composition (e.g., one or more peptides, complexes (e.g., pMHC complex), fusion proteins, conjugates, nucleic acid molecules, vectors, a cells, binding moieties) of the present disclosure.

[00357] The HTLV-1-induced disease or disorder can be, e.g., adult T-cell leukemia/lymphoma (ATL) or HTLV-1-myelopathy/tropical spastic paraparesis (HAM/TSP). In some embodiments, the HTLV-1-induced disease or disorder can be, e.g., a hypersensitivity reaction such as, without limitation, arthritis, uveitis, or HTLV-1-associated Infective Dermatitis (IDH).

[00358] Non-limiting examples of ATL are acute ATL, lymphoma ATL, chronic ATL, and smoldering ATL. The most benign but detectable form of ATL is an asymptomatic pre-leukemic phase which can be diagnosed, e.g., incidentally upon examination of a peripheral blood smear which reveals abnormal lymphocytes with characteristics of lobulated nuclei Smoldering ATL is the most benign symptomatic form of ATL and can be characterized by a normal peripheral blood leukocyte count, a few circulating leukemic cells, and cutaneous lesions but not visceral lesions. The development of chronic ATL is associated with visceral involvement and is evidenced by hepatosplenomegaly, lymphadenopathy, and peripheral blood leukocytosis. Patients with acute ATL can have elevated levels of lactate dehydrogenase and bilirubin, peripheral blood leukocytosis, hypercalcemia, and cutaneous and visceral involvement. ATL can be accompanied by severe immunosuppression and many ATL patients can be highly susceptible to various infections such, but not limited to cryptococcal meningitis, pneumocystis carinii pneumonia, disseminated cytomegalovirus infection, and candida esophagitis. Acute ATL is typically unresponsive to conventional chemotherapy and the median life expectancy of patients with acute ATL is approximately 11 months.

[00359] HTLV-1 is thought to cause approximately 80 percent of the cases of HAM/TSP such as by weakening the immune system. A majority of patients with

HAM/TSP have few or no symptoms over the lifespan. In addition to neurological symptoms of weakness and muscle spasms and/or stiffness, in rare instances patients with HAM/TSP may also have, e.g., arthritis (inflammation of one or more joints); infectious dermatitis (inflammation of the skin); keratoconjunctivitis sicca (persistent dryness of the cornea and/or conjunctiva); polymyositis (an inflammatory muscle disease); pulmonary lymphocytic alveolitis (inflammation of the lung); and/or uveitis (inflammation of the uveal tract of the eye).

[00360] When the disease being treated is a cancer, the cancer may specifically be of the following histological type, though it is not limited to these: neoplasm, malignant; carcinoma; carcinoma, undifferentiated; giant and spindle cell carcinoma; small cell carcinoma; papillary carcinoma; squamous cell carcinoma; lymphoepithelial carcinoma; basal cell carcinoma; pilomatrix carcinoma; transitional cell carcinoma; papillary transitional cell carcinoma; adenocarcinoma; gastrinoma, malignant; cholangiocarcinoma; hepatocellular carcinoma; combined hepatocellular carcinoma and cholangiocarcinoma; trabecular adenocarcinoma; adenoid cystic carcinoma; adenocarcinoma in adenomatous polyp; adenocarcinoma, familial polyposis coli; solid carcinoma; carcinoid tumor, malignant; branchiolo-alveolar adenocarcinoma; papillary adenocarcinoma; chromophobe carcinoma; acidophil carcinoma; oxyphilic adenocarcinoma; basophil carcinoma; clear cell adenocarcinoma; granular cell carcinoma; follicular adenocarcinoma; papillary and follicular adenocarcinoma; nonencapsulating sclerosing carcinoma; adrenal cortical carcinoma; endometroid carcinoma; skin appendage carcinoma; apocrine adenocarcinoma; sebaceous adenocarcinoma; ceruminous adenocarcinoma; mucoepidermoid carcinoma; cystadenocarcinoma; papillary cystadenocarcinoma; papillary serous cystadenocarcinoma; mucinous cystadenocarcinoma; mucinous adenocarcinoma; signet ring cell carcinoma; infiltrating duct carcinoma; medullary carcinoma; lobular carcinoma; inflammatory carcinoma; paget's disease, mammary; acinar cell carcinoma; adenosquamous carcinoma; adenocarcinoma w/squamous metaplasia; thymoma, malignant; ovarian stromal tumor, malignant; thecoma, malignant; granulosa cell tumor, malignant; and roblastoma, malignant; sertoli cell carcinoma; leydig cell tumor, malignant; lipid cell tumor, malignant; paraganglioma, malignant; extra-mammary paraganglioma, malignant; pheochromocytoma; glomangiosarcoma; malignant melanoma; amelanotic melanoma;

superficial spreading melanoma; malig melanoma in giant pigmented nevus; epithelioid cell melanoma; blue nevus, malignant; sarcoma; fibrosarcoma; fibrous histiocytoma, malignant; myxosarcoma; liposarcoma; leiomyosarcoma; rhabdomyosarcoma; embryonal rhabdomyosarcoma; alveolar rhabdomyosarcoma; stromal sarcoma; mixed tumor, malignant; mullerian mixed tumor; nephroblastoma; hepatoblastoma; carcinosarcoma; mesenchymoma, malignant; brenner tumor, malignant; phyllodes tumor, malignant; synovial sarcoma; mesothelioma, malignant; dysgerminoma; embryonal carcinoma; teratoma, malignant; struma ovarii, malignant; choriocarcinoma; mesonephroma, malignant; hemangiosarcoma; hemangioendothelioma, malignant; kaposi's sarcoma; hemangiopericytoma, malignant; lymphangiosarcoma; osteosarcoma; juxtacortical osteosarcoma; chondrosarcoma; chondroblastoma, malignant; mesenchymal chondrosarcoma; giant cell tumor of bone; ewing's sarcoma; odontogenic tumor, malignant; ameloblastic odontosarcoma; ameloblastoma, malignant; ameloblastic fibrosarcoma; pinealoma, malignant; chordoma; glioma, malignant; ependymoma; astrocytoma; protoplasmic astrocytoma; fibrillary astrocytoma; astroblastoma; glioblastoma; oligodendroglioma; oligodendroblastoma; primitive neuroectodermal; cerebellar sarcoma; ganglioneuroblastoma; neuroblastoma; retinoblastoma; olfactory neurogenic tumor; meningioma, malignant; neurofibrosarcoma; neurilemmoma, malignant; granular cell tumor, malignant; malignant lymphoma; Hodgkin's disease; Hodgkin's lymphoma; paragranuloma; malignant lymphoma, small lymphocytic; malignant lymphoma, large cell, diffuse; malignant lymphoma, follicular; mycosis fungoides; other specified non-Hodgkin's lymphomas; malignant histiocytosis; multiple myeloma; mast cell sarcoma; immunoproliferative small intestinal disease; leukemia; lymphoid leukemia; plasma cell leukemia; erythroleukemia; lymphosarcoma cell leukemia; myeloid leukemia; basophilic leukemia; eosinophilic leukemia; monocytic leukemia; mast cell leukemia; megakaryoblastic leukemia; myeloid sarcoma; and hairy cell leukemia.

[00361] It is contemplated that when used to treat various diseases, the compositions and methods can be combined with other therapeutic agents suitable for the same or similar diseases. Also, two or more embodiments described herein may be also co-administered to generate additive or synergistic effects. When co-administered with a second therapeutic

agent, the embodiment described herein and the second therapeutic agent may be simultaneously or sequentially (in any order). Suitable therapeutically effective dosages for each agent may be lowered due to the additive action or synergy.

[00362] In some embodiments, the compositions and methods disclosed herein are useful to enhance the efficacy of vaccines directed to HTLV-1 infection or HTLV-1-induced diseases. Thus, the compositions and methods described herein can be administered to a subject either simultaneously with or before (e.g., 1-30 days before) a reagent (including but not limited to small molecules, antibodies, or cellular reagents) that acts to elicit an immune response (e.g., to treat HTLV-1, HTLV-2, HTLV-3, and/or HTLV-4 infection or cancer) is administered to the subject.

[00363] The compositions and methods described herein can be also administered in combination with an anti-tumor antibody or an antibody directed at a pathogenic antigen (e.g., HTLV-1) or allergen.

[00364] The compositions and methods described herein can be combined with other immunomodulatory treatments such as, e.g., therapeutic vaccines (including but not limited to GVAX, DC-based vaccines, etc.), checkpoint inhibitors (including but not limited to agents that block CTLA4, PD1, LAG3, TIM3, etc.) or activators (including but not limited to agents that enhance 41BB, OX40, etc.). The inhibitory treatments described herein can be also combined with other treatments that possess the ability to modulate NKT function or stability, including but not limited to CD1d, CD1d-fusion proteins, CD1d dimers or larger polymers of CD1d either unloaded or loaded with antigens, CD1d-chimeric antigen receptors (CD1d-CAR), or any other of the five known CD1 isomers existing in humans (CD1a, CD1b, CD1c, CD1e), in any of the aforementioned forms or formulations, alone or in combination with each other or other agents.

[00365] Therapeutic methods described herein can be combined with additional immunotherapies and therapies. For example, when used for treating cancer, NKT cells described herein can be used in combination with cancer therapies, such as, e.g., surgery, radiotherapy, chemotherapy or combinations thereof, depending on type of the tumor, patient condition, other health issues, and a variety of factors. In certain aspects, other therapeutic agents useful for combination cancer therapy with the inhibitors described herein include anti-angiogenic agents. Many anti-angiogenic agents have been identified

and are known in the art, including, e.g., TNP-470, platelet factor 4, thrombospondin-1, tissue inhibitors of metalloproteases (TIMP1 and TIMP2), prolactin (16-Kd fragment), angiostatin (38-Kd fragment of plasminogen), endostatin, bFGF soluble receptor, transforming growth factor β, interferon-α, soluble KDR and FLT-1 receptors, placental proliferin-related protein, as well as those listed by Carmeliet and Jain (2000). In some embodiments, the inhibitors described herein can be used in combination with a VEGF antagonist or a VEGF receptor antagonist such as anti-VEGF antibodies, VEGF variants, soluble VEGF receptor fragments, aptamers capable of blocking VEGF or VEGFR, neutralizing anti-VEGFR antibodies, inhibitors of VEGFR tyrosine kinases and any combinations thereof (e.g., anti-hVEGF antibody A4.6.1, bevacizumab or ranibizumab).

[00366] Non-limiting examples of chemotherapeutic compounds which can be used in combination treatments include, for example, aminoglutethimide, amsacrine, anastrozole, asparaginase, bcg, bicalutamide, bleomycin, buserelin, busulfan, campothecin, capecitabine, carboplatin, carmustine, chlorambucil, cisplatin, cladribine, clodronate, colchicine, cyclophosphamide, cyproterone, cytarabine, dacarbazine, dactinomycin, daunorubicin, dienestrol, diethylstilbestrol, docetaxel, doxorubicin, epirubicin, estradiol, estramnustine, etoposide, exemestane, filgrastim, fludarabine, fludrocortisone, fluorouracil, fluoxymesterone, flutamide, gemcitabine, genistein, goserelin, hydroxyurea, idarubicin, ifosfamide, imatinib, interferon, irinotecan, ironotecan, letrozole, leucovorin, leuprolide, levamisole, lomustine, mechlorethamine, medroxyprogesterone, megestrol, melphalan, mercaptopurine, mesna, methotrexate, mitomycin, mitotane, mitoxantrone, nilutamide, nocodazole, octreotide, oxaliplatin, paclitaxel, pamidronate, pentostatin, plicamycin, porfimer, procarbazine, raltitrexed, rituximab, streptozocin, suramin, tamoxifen, temozolomide, teniposide, testosterone, thioguanine, thiotepa, titanocene dichloride, topotecan, trastuzumab, tretinoin, vinblastine, vincristine, vindesine, and vinorelbine.

[00367] These chemotherapeutic compounds may be categorized by their mechanism of action into, for example, following groups: anti-metabolites/anti-cancer agents, such as pyrimidine analogs (5-fluorouracil, floxuridine, capecitabine, gemcitabine and cytarabine) and purine analogs, folate antagonists and related inhibitors (mercaptopurine, thioguanine, pentostatin and 2-chlorodeoxyadenosine (cladribine));

antiproliferative/antimitotic agents including natural products such as vinca alkaloids (vinblastine, vincristine, and vinorelbine), microtubule disruptors such as taxane (paclitaxel, docetaxel), vincristin, vinblastin, nocodazole, epothilones and navelbine, epidipodophyllotoxins (etoposide, teniposide), DNA damaging agents (actinomycin, amsacrine, anthracyclines, bleomycin, busulfan, camptothecin, carboplatin, chlorambucil, cisplatin, cyclophosphamide, cytoxan, dactinomycin, daunorubicin, doxorubicin, epirubicin, hexamethyhnelamineoxaliplatin, iphosphamide, melphalan, merchlorehtamine, mitomycin, mitoxantrone, nitrosourea, plicamycin, procarbazine, taxol, taxotere, teniposide, triethylenethiophosphoramide and etoposide (VP16)); antibiotics such as dactinomycin (actinomycin D), daunorubicin, doxorubicin (adriamycin), idarubicin, anthracyclines, mitoxantrone, bleomycins, plicamycin (mithramycin) and mitomycin; enzymes (L-asparaginase which systemically metabolizes L-asparagine and deprives cells which do not have the capacity to synthesize their own asparagine); antiplatelet agents; antiproliferative/antimitotic alkylating agents such as nitrogen mustards (mechlorethamine, cyclophosphamide and analogs, melphalan, chlorambucil), ethylenimines and methylmelamines (hexamethylmelamine and thiotepa), alkyl sulfonates-busulfan, nitrosoureas (carmustine (BCNU) and analogs, streptozocin), trazenes-dacarbazinine (DTIC); antiproliferative/antimitotic antimetabolites such as folic acid analogs (methotrexate); platinum coordination complexes (cisplatin, carboplatin), procarbazine, hydroxyurea, mitotane, aminoglutethimide; hormones, hormone analogs (estrogen, tamoxifen, goserelin, bicalutamide, nilutamide) and aromatase inhibitors (letrozole, anastrozole); anticoagulants (heparin, synthetic heparin salts and other inhibitors of thrombin); fibrinolytic agents (such as tissue plasminogen activator, streptokinase and urokinase), aspirin, dipyridamole, ticlopidine, clopidogrel, abciximab; antimigratory agents; antisecretory agents (breveldin); immunosuppressives (cyclosporine, tacrolimus (FK-506), sirolimus (rapamycin), azathioprine, mycophenolate mofetil); anti-angiogenic compounds (e.g., TNP-470, genistein, bevacizumab) and growth factor inhibitors (e.g., fibroblast growth factor (FGF) inhibitors); angiotensin receptor blocker; nitric oxide donors; anti-sense oligonucleotides; antibodies (trastuzumab); cell cycle inhibitors and differentiation inducers (tretinoin); mTOR inhibitors, topoisomerase inhibitors (doxorubicin (adriamycin), amsacrine, camptothecin, daunorubicin,

dactinomycin, eniposide, epirubicin, etoposide, idarubicin and mitoxantrone, topotecan, irinotecan), corticosteroids (cortisone, dexamethasone, hydrocortisone, methylpednisolone, prednisone, and prenisolone); growth factor signal transduction kinase inhibitors; mitochondrial dysfunction inducers and caspase activators; and chromatin disruptors.

[00368] For treatment of viral infections, combined therapy described herein can encompass co-administering compositions and methods described herein with an anti-viral drug. Non-limiting examples of useful anti-viral drugs include, adefovir and entecavir, telbivudine, immune system modulators such as interferon-α, -β or -γ, didanosine, lamivudine, zanamavir, lopanivir, nelfinavir, efavirenz, indinavir, valacyclovir, zidovudine, amantadine, rimantidine, ribavirin, ganciclovir, foscarnet, and acyclovir or any salts or variants thereof. See also Physician's Desk Reference, 59.sup.th edition, (2005), Thomson P D R, Montvale N.J.; Gennaro et al., Eds. Remington's The Science and Practice of Pharmacy 20.sup.th edition, (2000), Lippincott Williams and Wilkins, Baltimore Md.; Braunwald et al., Eds. Harrison's Principles of Internal Medicine, 15.sup.th edition, (2001), McGraw Hill, NY; Berkow et al., Eds. The Merck Manual of Diagnosis and Therapy, (1992), Merck Research Laboratories, Rahway N.J.

Kits

[00369] The present disclosure further comprises a kit which may comprise any of various compositions of the present disclosure, including the peptides, peptide-based molecules (such as complexes (e.g., peptide-MHC (pMHC) complexes), fusion proteins, or conjugates comprising the peptide(s)), nucleic acid molecules, vectors, cells, or binding moieties of the disclosure.

[00370] In one aspect, the present disclosure may include a kit comprising, for example: (a) a container that contains a pharmaceutical composition disclosed herein, for example, a pharmaceutical composition in solution or in lyophilized form; (b) optionally, a second container containing a diluent or reconstituting solution for the lyophilized formulation; and/or (c) optionally, instructions for (i) use of the solution or (ii) reconstitution and/or use of the lyophilized formulation.

[00371] In some embodiments, the kit may further comprise, for example, without limitation, one or more of (i) a buffer, (ii) a diluent, (iii) a filter, (iv) a needle, and/or (v) a syringe. As a non-limiting example, the container may be a bottle, a vial, a syringe or test tube. In some embodiments, the container may be a multi-use container. In some the pharmaceutical composition may be lyophilized.

[00372] Kits of the present disclosure may comprise a lyophilized formulation of the present disclosure in a suitable container and instructions for its reconstitution and/or use. Suitable containers include, for example, bottles, vials (e.g. dual chamber vials), syringes (such as dual chamber syringes) and test tubes. The container may be formed from a variety of materials such as glass or plastic. The kit and/or container may contain instructions on or associated with the container that indicate directions for reconstitution of the lyophilized formulation and/or use of the kit. For example, the label may indicate that the lyophilized formulation is to be reconstituted to an appropriate peptide concentration. The label may indicate that the formulation is useful or intended for any route of administration disclosed herein, e.g., parenteral administration routes disclosed herein.

[00373] The container holding the formulation may be a multi-use vial, which may allow for repeat administrations (e.g., from 2-6 administrations) of the reconstituted formulation. The kit may further comprise a second container comprising a suitable diluent (e.g., sodium bicarbonate solution).

[00374] Upon mixing of the diluent and the lyophilized formulation, the final peptide concentration in the reconstituted formulation is reached. The kit may further include other materials desirable from a commercial and/or user standpoint, including, for example, without limitation, other buffers, diluents, filters, needles, syringes, and/or package inserts which may comprise, e.g., instructions for use.

[00375] Kits of the present disclosure may have a single container that contains the formulation of the pharmaceutical compositions according to the present disclosure with or without other components (e.g., other compounds or pharmaceutical compositions of these other compounds) or may have a distinct container for each component.

[00376] In some embodiments, kits of the disclosure may include a formulation of the disclosure packaged for use in combination with the coadministration of a second compound (such as adjuvants (e.g., GM-CSF, a chemotherapeutic agent, a natural product,

a hormone or antagonist, an anti-angiogenesis agent or inhibitor, an apoptosis-inducing agent or a chelator) or a pharmaceutical composition thereof. The components of the kit may be pre-complexed or each component may be in a separate distinct container prior to administration to a patient. The components of the kit may be provided in one or more liquid solutions. A liquid solutions described herein may be an aqueous solution, for example, a sterile aqueous solution. The components of the kit may also be provided as solids, which may be converted into liquids such as by addition of suitable solvents, which may be provided in another distinct container.

[00377] The container of a therapeutic kit may be a vial, test tube, flask, bottle, syringe, or any other means of enclosing a solid or liquid. When there is more than one component, the kit may contain a second vial or other container, which may allow for separate dosing. The kit may also contain another container for a pharmaceutically acceptable liquid. In some embodiment, a kit may contain an apparatus (e.g., one or more needles, syringes, eye droppers, pipettes, etc.), which may allow for administration of the agents of the disclosure that are components of the present kit.

Method and System for Identifying an Immunogenic Virus-Derived Peptide

[00378] The present disclosure further comprises a method and system for identifying an immunogenic virus-derived peptide. Some or all the steps of the method can be executed by a computational device. In some aspects, some or all of the steps of the method can be stored as computer-readable instructions within a memory such that the instructions can be executed by one or more processors to perform functions associated with the method.

[00379] In one aspect, the present disclosure may include a method for identifying an immunogenic virus-derived peptide that includes the following steps: (a) obtaining a plurality of RNA contig sequences derived from an infected subject infected with a virus, wherein the plurality of RNA contig sequences comprise a plurality of virus-derived RNA contig sequences and a plurality of infected-subject endogenous RNA contig sequences; (b) identifying the plurality of virus-derived RNA contig sequences from within the plurality of RNA contig sequences; (c) assembling a viral RNA sequence based on the plurality of virus-derived RNA contig sequences; (d) identifying a protein sequence based

on the viral RNA sequence; and (e) identifying the immunogenic virus-derived peptide based at least in part on the identified protein sequence.

[00380] In one aspect, a system configured to perform functions associated with the foregoing method can include a non-transitory computer-readable medium configured to communicate with one or more processor(s) of a computational device. The non-transitory computer-readable medium may include instructions thereon, that when executed by the processor(s), cause the computational device to: (a) receive, as an input, a plurality of RNA contig sequences derived from an infected subject infected with a virus such that the plurality of RNA contig sequences comprise a plurality of virus-derived RNA contig sequences and a plurality of infected-subject endogenous RNA contig sequences, and wherein the infected subject is infected with a virus; (b) identify the plurality of virus-derived RNA contig sequences from within the plurality of RNA contig sequences; (c) assemble a viral RNA sequence based on the plurality of virus-derived RNA contig sequences; (d) identify a protein sequence based on the viral RNA sequence; (e) identify an immunogenic virus-derived peptide based at least in part on the protein sequence; and (f) provide, as an output, the immunogenic virus-derived peptide.

[00381] In some embodiments, the plurality of RNA contig sequences may be derived from one infected subject.

[00382] In some embodiments, the infected subject may be a human.

[00383] In some embodiments, the plurality of virus-derived RNA contig sequences are derived from the virus infecting the infected subject.

[00384] In some embodiments, the plurality of infected-subject endogenous RNA contig sequences are derived from RNA endogenous to the infected subject.

[00385] In some embodiments, identifying the plurality of virus-derived RNA contig sequences from within the plurality of RNA contig sequences (step b) can further include: comparing at least a portion of contig sequences of the plurality of RNA contig sequences to a reference viral sequence; and identifying the plurality of virus-derived RNA contig sequences such that each contig sequence of the plurality of virus-derived RNA contig sequences comprises at least a portion that corresponds to the reference viral sequence.

[00386] In some embodiments, each contig sequence of the plurality of virus-derived RNA contig may be distinct from the plurality of infected-subject endogenous RNA contig sequences.

[00387] In some embodiments, each contig sequence of the plurality of virus-derived RNA contig sequences may lack infected-subject endogenous RNA contig sequences.

[00388] In some embodiments, the reference viral sequence may include a reference genome.

[00389] In some embodiments, the reference genome may include a HTLV-1 virus genome.

[00390] In some embodiments, assembling the viral RNA sequence based on the plurality of virus-derived RNA contig sequences (step c) may include: overlapping common sequence portions at ends of at least a portion of the plurality of virus-derived RNA contig sequences such that the at least a portion of the plurality of virus-derived RNA contig sequences overlap linearly to assemble the viral RNA sequence.

[00391] In some embodiments, identifying a protein sequence based on the viral RNA sequence such that the protein sequence includes a translation of the viral RNA sequence (step d) may include: identifying the protein sequence without requiring a comparison to a database of viral proteins.

[00392] In some embodiments, identifying the protein sequence based on the viral RNA sequence such that the protein sequence includes a translation of the viral RNA sequence (step d) may include: identifying a plurality of protein sequences each based on the viral RNA sequence such that each of the plurality of protein sequences respectively include a translation of the viral RNA sequence, and identifying the protein sequence as a frequently occurring protein sequence within the plurality of protein sequences.

[00393] In some embodiments, the protein sequence may be identified based on the viral RNA sequence associated with a single infected subject.

[00394] In some embodiments, identifying the immunogenic virus-derived peptide based at least in part on the protein sequence (step 3) may include: identifying a MHC molecule associated with the single infected subject; identifying one or more peptides based at least in part on the protein sequence such that the one or more peptides each form

a respective MHC-peptide complex with the MHC molecule; and identifying the immunogenic virus-derived peptide based on the one or more peptides.

Method and System for Identifying an Integration Site of a Viral Gene within a Subject Gene

[00395] The present disclosure further comprises a method and system for identifying an integration site of a viral gene within a subject gene. Some or all the steps of the method can be executed by a computational device. In some aspects, some or all of the steps of the method can be stored as computer-readable instructions within a memory such that the instructions can be executed by one or more processors to perform functions associated with the method.

[00396] In one aspect, the present disclosure may include a method for identifying an integration site of a viral gene within a subject gene that includes the following steps: (a) obtaining a plurality of RNA contig sequences derived from an infected subject infected with a virus such that the plurality of RNA contig sequences comprise a plurality of virus-derived RNA contig sequences, a plurality of infected-subject endogenous RNA contig sequences, and a plurality of hybrid RNA contig sequences comprising viral and infected-subject endogenous portions; (b) identifying the plurality of hybrid RNA contig sequences from within the plurality of RNA contig sequences; (c) comparing, for at least a portion of the plurality of hybrid RNA contig sequences, infected-subject endogenous portions to a subject reference genome; and (d) identifying, based at least in part on the comparison of infected-subject endogenous portions to the subject reference genome, an integration site comprising the subject gene.

[00397] In one aspect, a system configured to perform functions associated with the foregoing method can include a non-transitory computer-readable medium configured to communicate with one or more processor(s) of a computational device. The non-transitory computer-readable medium may include instructions thereon, that when executed by the processor(s), cause the computational device to: (a) receive, as an input, a plurality of RNA contig sequences derived from an infected subject infected with a virus such that the plurality of RNA contig sequences comprise a plurality of virus-derived RNA contig sequences, a plurality of infected-subject endogenous RNA contig sequences, and plurality

of hybrid RNA contig sequences comprising viral and infected-subject endogenous portions; (b) identify the plurality of hybrid RNA contig sequences from within the plurality of RNA contig sequences; (c) compare, for at least a portion of the plurality of hybrid RNA contig sequences, infected-subject endogenous portions to a subject reference genome; (d) identify, based at least in part on the comparison of infected-subject endogenous portions to the subject reference genome, an integration site comprising a subject gene; and (e) provide, as an output, the integration site.

[00398] In some embodiments, the plurality of RNA contig sequences may be derived from one infected subject.

[00399] In some embodiments, the infected subject may be a human.

[00400] In some embodiments, the plurality of virus-derived RNA contig sequences may be derived from the virus infecting the infected subject.

[00401] In some embodiments, the virus may include a HTLV-1 virus.

[00402] In some embodiments, the plurality of infected-subject endogenous RNA contig sequences may be derived from RNA endogenous to the infected subject.

[00403] In some embodiments, the subject reference genome can include the human genome.

[00404] Certain embodiments and implementations of the disclosed technology are described above with reference to systems and methods and/or computer program products according to example embodiments or implementations of the disclosed technology. It will be understood that method steps can be implemented by computer-executable program instructions. Likewise, some method steps need not be performed in the order presented, may be repeated, or may not necessarily need to be performed at all, according to some embodiments or implementations of the disclosed technology.

[00405] These computer-executable program instructions may be loaded onto a general-purpose computer, a special-purpose computer, a processor, or other programmable data processing apparatus to produce a particular machine, such that the instructions that execute on the computer, processor, or other programmable data processing apparatus create means for implementing one or more functions associated with method steps. These computer program instructions may also be stored in a computer-readable memory that can direct a computer or other programmable data processing

apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory produce an article of manufacture including instruction means that implement one or more functions associated with method steps.

[00406] Non-transitory computer-readable media may include, but is not limited to, random access memory (RAM), read-only memory (ROM), electronically erasable programmable ROM (EEPROM), flash memory or other memory technology, compact disc ROM (CD-ROM), digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other tangible, physical medium which can be used to store computer readable information.

[00407] Embodiments or implementations of the disclosed technology may provide for a computer program product, including a computer-usable medium having a computer-readable program code or program instructions embodied therein, said computer-readable program code adapted to be executed to implement one or more functions related to example methods presented herein. Likewise, the computer program instructions may be loaded onto a computer or other programmable data processing apparatus to cause a series of operational elements or steps to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions that execute on the computer or other programmable apparatus provide elements or steps for implementing the functions related to example methods presented herein.

[00408] Accordingly, illustrations and descriptions of methods support combinations of means for performing the specified functions, combinations of elements or steps for performing the specified functions, and program instruction means for performing the specified functions. It will also be understood that at least some method steps, and combinations of method steps, can be implemented by special-purpose, hardware-based computer systems that perform the specified functions, elements or steps, or combinations of special-purpose hardware and computer instructions.

[00409] Certain implementations of the disclosed technology can be utilized with customer devices that may include mobile computing devices. Those skilled in the art will recognize that there are several categories of mobile devices, generally known as portable computing devices that can run on batteries but are not usually classified as laptops. For

example, mobile devices can include, but are not limited to portable computers, tablet PCs, internet tablets, PDAs, ultra-mobile PCs (UMPCs), wearable devices, and smart phones.

[00410] Certain implementations of the disclosed technology can be utilized with medical equipment, medical devices, and/or associated peripherals.

Non-limiting Examples of Methods and Systems

[00411] In various embodiments, the present disclosure provides a method for identifying an immunogenic virus-derived peptide, the method comprising:

a) obtaining a plurality of RNA contig sequences derived from an infected subject infected with a virus, wherein the plurality of RNA contig sequences comprises a plurality of virus-derived RNA contig sequences and a plurality of infected-subject endogenous RNA contig sequences;

b) identifying the plurality of virus-derived RNA contig sequences from within the plurality of RNA contig sequences;

c) assembling a viral RNA sequence based on the plurality of virus-derived RNA contig sequences;

d) identifying a protein sequence based on the viral RNA sequence; and

e) identifying the immunogenic virus-derived peptide based at least in part on the identified protein sequence.

[00412] In some embodiments, the plurality of RNA contig sequences are derived from one infected subject.

[00413] In some embodiments, the infected subject is a human.

[00414] In some embodiments, the plurality of virus-derived RNA contig sequences are derived from the virus infecting the infected subject.

[00415] In some embodiments, the plurality of infected-subject endogenous RNA contig sequences are derived from RNA endogenous to the infected subject.

[00416] In some embodiments, the identifying the plurality of virus-derived RNA contig sequences from within the plurality of RNA contig sequences comprises:

comparing at least a portion of contig sequences of the plurality of RNA contig sequences to a reference viral sequence; and

identifying the plurality of virus-derived RNA contig sequences such that each contig sequence of the plurality of virus-derived RNA contig sequences comprises at least a portion that corresponds to the reference viral sequence.

[00417] In some embodiments, each contig sequence of the plurality of virus-derived RNA contig is distinct from the plurality of infected-subject endogenous RNA contig sequences.

[00418] In some embodiments, each contig sequence of the plurality of virus-derived RNA contig sequences lacks infected-subject endogenous RNA contig sequences.

[00419] In some embodiments, the reference viral sequence comprises a reference genome.

[00420] In some embodiments, the reference genome comprises a HTLV-1 virus genome.

[00421] In some embodiments, the assembling the viral RNA sequence based on the plurality of virus-derived RNA contig sequences comprises:

overlapping common sequence portions at ends of at least a portion of the plurality of virus-derived RNA contig sequences such that the at least a portion of the plurality of virus-derived RNA contig sequences overlap linearly to assemble the viral RNA sequence.

[00422] In some embodiments, the identifying a protein sequence based on the viral RNA sequence such that the identified protein sequence includes a translation of the viral RNA sequence comprises:

identifying the protein sequence without requiring a comparison to a database of viral proteins.

[00423] In some embodiments, the identifying the protein sequence based on the viral RNA sequence such that the identified protein sequence includes a translation of the viral RNA sequence further comprises:

identifying a plurality of protein sequences each based on the viral RNA sequence such that each of the plurality of protein sequences respectively include a translation of the viral RNA sequence, and

identifying the protein sequence as a frequently occurring protein sequence within the plurality of protein sequences.

[00424] In some embodiments, the protein sequence identified based on the viral RNA sequence is associated with a single infected subject.

[00425] In some embodiments, the identifying the immunogenic virus-derived peptide based at least in part on the protein sequence comprises:

identifying an MHC molecule associated with the single infected subject;

identifying one or more peptides based at least in part on the protein sequence such that the one or more peptides each form a respective MHC-peptide complex with the MHC molecule; and

identifying the immunogenic virus-derived peptide based on the one or more peptides.

[00426] In various embodiments, the present disclosure provides a non-transitory computer-readable medium configured to communicate with one or more processor(s) of a computational device, the non-transitory computer-readable medium including instructions thereon, that when executed by the processor(s), cause the computational device to:

a) receive, as an input, a plurality of RNA contig sequences derived from an infected subject infected with a virus such that the plurality of RNA contig sequences comprise a plurality of virus-derived RNA contig sequences and a plurality of infected-subject endogenous RNA contig sequences, and wherein the infected subject is infected with a virus;

b) identify the plurality of virus-derived RNA contig sequences from within the plurality of RNA contig sequences;

c) assemble a viral RNA sequence based on the plurality of virus-derived RNA contig sequences;

d) identify a protein sequence based on the viral RNA sequence;

e) identify an immunogenic virus-derived peptide based at least in part on the protein sequence; and

f) provide, as an output, the immunogenic virus-derived peptide.

[00427] In some embodiments, the plurality of RNA contig sequences are derived from only one infected subject.

[00428] In some embodiments, the infected subject comprises a human.

[00429] In some embodiments, the plurality of virus-derived RNA contig sequences are derived from the virus infecting the infected subject.

[00430] In some embodiments, the plurality of infected-subject endogenous RNA contig sequences are derived from RNA endogenous to the infected subject.

[00431] In some embodiments, the instructions which cause the computational device to identify the plurality of virus-derived RNA contig sequences from within the plurality of RNA contig sequences further comprise instructions, that when executed by the processor(s), cause the computational device to:

compare at least a portion of contig sequences of the plurality of RNA contig sequences to a reference viral sequence; and

identify the plurality of virus-derived RNA contig sequences such that each contig sequence of the plurality of virus-derived RNA contig sequences comprises at least a portion that corresponds to the reference viral sequence.

[00432] In some embodiments, each contig sequence of the plurality of virus-derived RNA contig is distinct from the plurality of infected-subject endogenous RNA contig sequences.

[00433] In some embodiments, each contig sequence of the plurality of virus-derived RNA contig sequences lacks portions of infected-subject endogenous RNA contig sequences.

[00434] In some embodiments, the reference viral sequences comprises a reference genome.

[00435] In some embodiments, the reference genome comprises a HTLV-1 genome.

[00436] In some embodiments, the instructions which cause the computational device to assemble the viral RNA sequence based on the plurality of virus-derived RNA contig sequences further comprise instructions, that when executed by the processor(s), cause the computational device to:

overlap common sequence portions at ends of at least a portion of the plurality of virus-derived RNA contig sequences such that the at least a portion of the plurality of virus-derived RNA contig sequences overlap linearly to assemble the viral RNA sequence.

[00437] In some embodiments, the instructions which cause the computational device to identify a protein sequence based on the viral RNA sequence such that the protein

sequence includes a translation of the viral RNA sequence further comprise instructions, that when executed by the processor, cause the computational device to:

identify the protein sequence without requiring a comparison to a database of viral proteins.

[00438] In some embodiments, the protein sequence identified based on the viral RNA sequence is a novel protein.

[00439] In some embodiments, the protein sequence identified based on the viral RNA sequence is associated with a single infected subject.

[00440] In some embodiments, the instructions which cause the computational device to identify a protein sequence based on the viral RNA sequence such that the protein sequence includes a translation of the viral RNA sequence further comprise instructions, that when executed by the processor(s), cause the computational device to:

identify a plurality of protein sequences each based on the viral RNA sequence such that each of the plurality of protein sequences respectively include a translation of the viral RNA sequence; and

identify the protein sequence as a frequently occurring protein sequence within the plurality of protein sequences.

[00441] In some embodiments, the instructions which cause the computational device to identify the immunogenic virus-derived peptide based at least in part on the protein sequence further comprise instructions, that when executed by the processor(s), cause the computational device to:

identify a major histocompatibility complex (MHC) molecule associated with the single infected subject;

identify one or more peptides based at least in part on the protein sequence such that the one or more peptides are each capable of forming a respective MHC-peptide complex with the MHC molecule; and

identify the immunogenic virus-derived peptide based on the one or more peptides.

[00442] In some embodiments, the instructions, when executed by the processor(s), further cause the computational device to:

store the protein sequence to a database such that the protein sequence is associated with the infected subject within the database.

[00443] In various embodiments, the present disclosure provides a method for identifying an integration site of a viral gene within a subject gene, the method comprising:

a) obtaining a plurality of RNA contig sequences derived from an infected subject infected with a virus such that the plurality of RNA contig sequences comprise a plurality of virus-derived RNA contig sequences, a plurality of infected-subject endogenous RNA contig sequences, and a plurality of hybrid RNA contig sequences comprising viral and infected-subject endogenous portions;

b) identifying the plurality of hybrid RNA contig sequences from within the plurality of RNA contig sequences;

c) comparing, for at least a portion of the plurality of hybrid RNA contig sequences, infected-subject endogenous portions to a subject reference genome; and

d) identifying, based at least in part on the comparison of infected-subject endogenous portions to the subject reference genome, an integration site comprising the subject gene.

[00444] In some embodiments, the plurality of RNA contig sequences are derived from one infected subject.

[00445] In some embodiments, the infected subject is a human.

[00446] In some embodiments, the plurality of virus-derived RNA contig sequences are derived from the virus infecting the infected subject.

[00447] In some embodiments, the virus is a HTLV-1.

[00448] In some embodiments, the plurality of infected-subject endogenous RNA contig sequences are derived from RNA endogenous to the infected subject.

[00449] In some embodiments, the subject reference genome comprises the human genome.

[00450] In various embodiments, the present disclosure provides a non-transitory computer-readable medium configured to communicate with one or more processor(s) of a computational device, the non-transitory computer-readable medium including instructions thereon, that when executed by the processor(s), cause the computational device to:

a) receive, as an input, a plurality of RNA contig sequences derived from an infected subject infected with a virus such that the plurality of RNA contig sequences

comprise a plurality of virus-derived RNA contig sequences, a plurality of infected-subject endogenous RNA contig sequences, and plurality of hybrid RNA contig sequences comprising viral and infected-subject endogenous portions;

b) identify the plurality of hybrid RNA contig sequences from within the plurality of RNA contig sequences;

c) compare, for at least a portion of the plurality of hybrid RNA contig sequences, infected-subject endogenous portions to a subject reference genome;

d) identify, based at least in part on the comparison of infected-subject endogenous portions to the subject reference genome, an integration site comprising a subject gene; and

e) provide, as an output, the integration site.

[00451] In some embodiments, the plurality of RNA contig sequences are derived from only one infected subject.

[00452] In some embodiments, the infected subject comprises a human.

[00453] In some embodiments, the plurality of virus-derived RNA contig sequences are derived from a virus infecting the infected subject.

[00454] In some embodiments, the virus comprises a HTLV-1.

[00455] In some embodiments, the plurality of infected-subject endogenous RNA contig sequences are derived from RNA endogenous to the infected subject.

[00456] In some embodiments, the subject reference genome comprises the human genome.

EXAMPLES

[00457] The present disclosure is also described and demonstrated by way of the following examples. However, the use of these and other examples anywhere in the specification is illustrative only and in no way limits the scope and meaning of the disclosure or of any exemplified term. Likewise, the disclosure is not limited to any particular preferred embodiments described here. Indeed, many modifications and variations of the disclosure may be apparent to those skilled in the art upon reading this specification, and such variations can be made without departing from the disclosure in spirit or in scope. The disclosure is therefore to be limited only by the terms of the appended claims along with the full scope of equivalents to which those claims are entitled.

[00458] Human T-lymphotropic Virus, type 1 (HTLV-1) is the etiological agent for Adult T-cell Leukemia/Lymphoma (ATL), HTLV-1 associated myelopathy (HAM), HTLV-1-associated infective dermatitis (IDH) and other inflammatory disorders like uveitis and arthritis. While HTLV-1 is the most well-studied of the four HTLV genotypes (HTLV-1, HTLV-2, HTLV-3 and HTLV-4) the immunopeptidome of HTLV-1-infected T-cells and cell lines have not yet been characterized. The surface presentation of HTLV-1-specific peptides by Human Leukocyte Antigens (HLA) could be leveraged for developing immunotherapies targeting infected lymphocytes in ATL. In the Examples described below, HLA-I immunopeptidomics were performed on four model HTLV-transformed cell lines: MT-4, C8166, MT-2 and C5/MJ; primary cells and ATL patient samples.

Example 1. Immunopeptidomics for HTLV-transformed cells.

[00459] HLA-I complexes were purified from the 4 HTLV-transformed cell lines MT-4, C8166, MT-2 and C5/MJ, using anti-HLA-I W6/32 affinity purification (**Figure 1A**). Briefly, 100 million cells for each cell line were lysed in 1% NP-40 lysis buffer under mild conditions, and lysates were incubated with the W6/32-conjugated sepharose beads in column format. The beads were washed followed by an HLA elution using an acidic glycine solution (pH 2.7). The HLA-associated peptides were then separated from HLA molecules by 30% Acetonitrile/H_2O selective elution from a C18-based sep-pak. The peptides were separated on a 25 cm column, and analyzed by liquid chromatography with tandem mass spectrometry (LC-MS/MS) using a Thermo Orbitrap Fusion Lumos connected online to nano-LC 1200 (Thermo). The mass spectrometry raw files were searched against a combined FASTA database containing both human (Uniprot) and HTLV-1 (Uniprot) protein sequences. More than 13,000 peptides were detected from the MT-4 cell line (**Figure 1D**) and between 8,000 to 10,000 peptides were detected from the C8166, MT-2, and C5/MJ cell lines (**Figures 1B-1C** and **Figure 1E**), with nonamers being the most abundant peptide length.

[00460] Furthermore, a number of unique HTLV peptides were also identified. For example, 19 HTLV-1 peptides were detected from MT-4 cells, and 16 of the peptides overlapped between the three biological replicates (**Figure 4** and **Table 5**). Similarly, 11 and 32 HTLV-1 specific peptides from C8166 and MT-2 cells were detected, respectively,

with high reproducibility between replicates (**Figure 2** and **Table 3**; **Figure 3** and **Table 4**). For the C5/ MJ cells, the HTLV-1 peptides presented by HLA, with and without Interferon-γ stimulation, were compared. No significant increase in number of HTLV peptides was observed after Interferon (IFN)-γ stimulation of C5/ MJ cells (**Figure 5** and **Table 6**). To increase confidence in the disclosed peptide identifications, HLA genotyping of these samples was performed, and the panNetMHC 4.0 algorithm was used to predict the binding affinity of the identified HTLV-1 peptides with the HLA alleles of the respective cell lines. Most of the HTLV-1 peptides identified from the four cell lines were strong (panNetMHC 4.0 Rank below 0.5) or weak (panNetMHC 4.0 Rank between 0.5 – 2.0) predicted binders to the HLA of described cell lines (**Tables 3-6**).

Table 3. HTLV-1 Peptides Detected from C8166 Cells

			HLA-A01:01		HLA-B08:01		HLA-B44:02		HLA-C05:01		HLA-C07:01	
Peptide	**Len**	**Protein**	**nM**	**Rank***	**nM**	**Rank***	**nM**	**Rank***	**nM**	**Rank***	**nM**	**Rank***
EELLYKISL (SEQ ID NO: 51)	9	TAX_HTL1A	38320	55.43	3473	3.90	547	**0.48**	38893	38.92	25227	18.72
LENRVLTGW (SEQ ID NO: 24)	9	ENV_HTL1M	33983	35.91	30892	53.25	19	**0.01**	34275	26.60	27688	22.69
LSFPDPGLRP (SEQ ID NO: 29)	10	TAX_HTL1A	30070	23.85	41497	91.10	37928	56.27	35221	28.71	34298	37.80
SARLHRHAL (SEQ ID NO: 63)	9	TAX_HTL1A	31354	27.27	9	**0.01**	29128	22.94	14399	4.86	2220	*0.75*
VPYKRIEEL (SEQ ID NO: 70)	9	TAX_HTL1A	35071	40.20	165	**0.29**	33692	36.10	16224	5.77	6279	2.35
NVPYKRIEEL (SEQ ID NO: 31)	10	TAX_HTL1A	36556	46.65	513	*0.78*	36175	46.81	32503	22.98	14327	7.07

*Strong binders (rank < 0.5) shown in **bold**; Weak binders (rank 0.5-2) shown in *italics*; Len = length

Table 4. HTLV-1 Peptides Detected from MT-2 Cells

	HLA-A24:02	HLA-B40:02	HLA-B51:01	HLA-C03:03	HLA-C14:02

Peptide	Len	Protein	nM	Rank	nM	Rank	nM	Rank	nM	Rank	nM	Rank
GYPGRVNEI (SEQ ID NO: 52)	9	POL_HTL1A	519	*0.68*	35247	50.32	38130	33.35	25508	19.67	246	*0.72*
GYPGRVNEIL (SEQ ID NO: 17)	10	POL_HTL1A	1957	*1.67*	33643	44.90	28084	12.94	29747	25.01	1922	3.80
DPPDSDPQI (SEQ ID NO: 11)	9	POL_HTL1A	44062	68.58	44332	90.49	8357	*1.60*	28896	23.81	26881	42.26
DPASILPVI (SEQ ID NO: 9)	9	POL_HTL1A	32076	25.07	29380	33.02	300	**0.05**	10191	7.27	12157	17.23
LPMDNALSI (SEQ ID NO: 67)	9	POL_HTL1A	22730	13.51	14188	11.05	42	**0.01**	70	**0.28**	443	*1.22*
ITNSHVSIL (SEQ ID NO: 20)	9	ENV_HTL1M	14306	7.59	26588	26.90	15469	3.89	75	**0.30**	378	*1.06*
SYSSYHATY (SEQ ID NO: 44)	9	ENV_HTL1M	509	*0.67*	33422	44.21	32809	19.96	11418	8.02	13	**0.03**
EELLYKISL (SEQ ID NO: 51)	9	TAX_HTL1A	37669	38.16	54	**0.23**	32889	20.11	32713	29.68	16712	23.87
RQWKGPQEAL (SEQ ID NO: 4)	10	POL_HTL1A	10032	5.36	339	*0.97*	37923	32.68	9913	7.09	1600	3.32
TEASNILRSC (SEQ ID NO: 45)	10	POL_HTL1A	42210	57.16	2047	2.75	38471	34.51	32410	29.16	35487	64.71
HEVDKDISQL (SEQ ID NO: 60)	10	ENV_HTL1M	39913	46.22	292	*0.88*	38819	35.76	17793	12.44	17429	25.02
QEVSRLNINL (SEQ ID NO: 62)	10	ENV_HTL1M	34910	30.79	111	**0.43**	39137	36.93	30665	26.37	24530	37.34
YQLSPPITW (SEQ ID NO: 57)	9	TAX_HTL1A	793	*0.92*	2358	2.97	16877	4.50	1267	*1.73*	815	*1.99*
LYQLSPPITW (SEQ ID NO: 54)	10	TAX_HTL1A	70	**0.15**	16256	12.89	27404	12.15	19331	13.68	1727	3.52
QFLIPRLPSF (SEQ ID NO: 35)	10	TAX_HTL1A	491	*0.65*	21909	19.22	28816	13.85	6860	5.26	475	*1.30*
WQLHHSPRL (SEQ ID NO: 55)	9	POL_HTL1A	10913	5.79	346	*0.98*	23142	8.21	734	*1.26*	435	*1.20*

VSSPCHNSL (SEQ ID NO: 46)	9	ENV_HTL1M	13254	7.01	24991	24.01	22174	7.48	81	**0.32**	530	*1.42*
KECQKLLQA (SEQ ID NO: 21)	9	POL_HTL1A	40910	50.63	601	*1.38*	44240	64.51	43402	61.43	37260	70.40
WAIIGRDAL (SEQ ID NO: 5)	9	POL_HTL1A	30147	21.99	6119	5.41	14943	3.70	14	**0.06**	363	*1.02*
AAHHWLNFL (SEQ ID NO: 49)	9	POL_HTL1A	17152	9.31	14968	11.73	16783	4.46	186	*0.54*	436	*1.20*

*Strong binders (rank < 0.5) shown in **bold**; Weak binders (rank 0.5-2) shown in *italics*; Len = length

Table 5. HTLV-1 Peptides Detected from MT-4 Cells

			HLA-A02:01		HLA-A11:01		HLA-A24:02		HLA-B07:02		HLA-B15:01	
Peptide	**Len**	**Protein**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**
ATMASLISH (SEQ ID NO: 1)	9	POL_HTL1A	23292	29.62	45	**0.27**	38680	41.49	16525	10.16	1399	3.31
VVCMYLYQL (SEQ ID NO: 65)	9	TAX_HTL1A	810	3.65	13484	11.45	7476	4.19	29122	26.05	9426	13.25
KVLTPPITH (SEQ ID NO: 66)	9	TAX_HTL1A	35565	55.59	1211	2.79	43344	63.83	28264	24.44	9835	13.75
VPYKRIEEL (SEQ ID NO: 70)	9	TAX_HTL1A	21206	26.74	30695	34.56	21672	12.63	26	**0.10**	24036	36.34
NVPYKRIEEL (SEQ ID NO: 31)	10	TAX_HTL1A	13518	17.90	33224	41.63	30214	22.09	776	*1.05*	19206	27.14
IQYSSFHSL (SEQ ID NO: 68)	9	TAX_HTL1A	188	*1.58*	13522	11.48	1313	*1.30*	7166	4.47	32	**0.15**
IPISLLFNE (SEQ ID NO: 18)	9	TAX_HTL1A	31874	45.55	28048	28.71	39840	45.92	15931	9.72	36047	69.04
YLFPHWIKK (SEQ ID NO: 56)	9	ENV_HTL1M	1866	5.62	16	**0.07**	17085	9.27	27351	22.79	11960	16.49
SLYLFPHWIKK (SEQ ID NO: 42)	11	ENV_HTL1M	7893	12.40	48	**0.29**	23786	14.45	40781	68.69	28930	47.91
SKCGFPFSL (SEQ ID NO: 41)	9	ENV_HTL1M	13258	17.65	23760	21.65	22454	13.27	29112	26.03	13457	18.50
SPCHNSLIL (SEQ ID NO: 64)	9	ENV_HTL1M	29362	39.90	30697	34.56	31568	24.22	201	**0.47**	24214	36.72
GKFLATLIL (SEQ ID NO: 59)	9	ENV_HTL1M	20595	25.96	35398	49.01	33590	27.95	28920	25.67	7580	11.00
YPVYVFGDCV (SEQ ID NO: 47)	10	TAX_HTL1A	713	3.41	26043	25.08	32241	25.38	1649	*1.64*	13525	18.59
LPEDCLPTTL (SEQ ID NO: 27)	10	TAX_HTL1A	22362	28.27	38024	59.88	35372	31.87	714	*1.00*	23605	35.45

LPAPHLTLP (SEQ ID NO: 53)	9	ENV_HTL1M	34016	51.13	38214	60.74	41338	52.69	3154	2.47	32077	56.66
LPAPHLTLPF (SEQ ID NO: 26)	10	ENV_HTL1M	21334	26.90	21407	18.66	3796	2.56	24	**0.09**	432	*1.55*
LPFRTTPIVL (SEQ ID NO: 2)	10	POL_HTL1A	22416	28.34	33551	42.64	13539	7.17	18	**0.06**	8535	12.17
GPSSYDFHQLK (SEQ ID NO: 15)	11	POL_HTL1A	31035	43.56	60	0.37	39648	45.16	22578	15.92	33133	59.80

*Strong binders (rank < 0.5) shown in **bold**; Weak binders (rank 0.5-2) shown in *italics*; Len = length

Table 6. HTLV-1 Peptides Detected from C5/MJ Cells ± IFN-γ Treatment

			HLA-A32:01		HLA-B35:03		HLA-B51:01		HLA-C04:01		HLA-C15:02	
Peptide	**Len**	**Protein**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**
ITNSHVSIL (SEQ ID NO: 20)	9	ENV_HTL1M	1017	*1.32*	16511	2.77	15469	3.89	12776	*1.29*	72	**0.06**
VPSSSSTPL (SEQ ID NO: 58)	9	ENV_HTL1M	28292	26.65	118	**0.03**	6490	*1.17*	17255	2.71	10231	8.51
SPCHNSLIL (SEQ ID NO: 64)	9	ENV_HTL1M	25872	23.06	1817	**0.26**	12163	2.69	24354	7.45	13736	11.72
RLHVWVDTF (SEQ ID NO: 36)	9	POL_HTL1A	10	**0.01**	32304	11.11	30892	16.73	16367	2.36	7366	6.10
RVNEILHIL (SEQ ID NO: 69)	9	POL_HTL1A	30	**0.06**	19571	3.62	26636	11.31	10693	*0.87*	50	**0.03**
DPASILPVI (SEQ ID NO: 9)	9	POL_HTL1A	28086	26.33	1151	**0.17**	300	**0.05**	31874	19.87	18084	16.48
RVLTGWGLNW (SEQ ID NO: 38)	10	ENV_HTL1M	48	**0.09**	42728	36.85	32647	19.65	31776	19.63	12857	10.86
ATCPEHQITW (SEQ ID NO: 8)	10	TAX_HTL1A	1099	*1.39*	43255	39.86	36476	28.32	35749	31.97	18531	17.00
FPDPGLRPQNL (SEQ ID NO: 14)	11	TAX_HTL1A	42371	65.22	1033	**0.15**	12452	2.79	15239	*1.96*	13126	11.12

EALQTGITL (SEQ ID NO: 50)	9	ENV_HTL1M	1553	12.08	703	**0.11**	3947	*0.69*	2568	8.87	4484	3.81
RSRRAVPVAVW (SEQ ID NO: 37)	11	ENV_HTL1M	1588	*1.83*	4515	54.11	3688	29.49	3282	22.43	1220	10.27

*Strong binders (rank < 0.5) shown in **bold**; Weak binders (rank 0.5-2) shown in *italics*; Len = length

Example 2. Immunopeptidomics for ATL primary cells.

[00461] Next, the HLA-I immunopeptidome of patient-derived ATL primary sample ED41214(-) was analyzed using the same method as discussed in **Example 1**. A slight increase in total peptide count with interferon-γ treatment (**Figure 6A** and **Figure 6C**) was observed. More importantly, 16 and 21 unique HTLV-1 peptides from untreated and interferon-γ treated ED41214(-) cells, respectively, were identified (**Figure 6B**, **Figure 6D** and **Figure 6E**). Only a few HTLV peptides were predicted binders to the common HLA alleles in the patient population (see, e.g., **Table 7**).

Table 7. HTLV-1 Peptides Detected from C5/MJ Cells ± IFN-γ Treatment

				HLA-A02:01		HLA-A11:01		HLA-A24:02		HLA-B07:02		HLA-B15:01		HLA-B54	
Peptide	**Protein**	**AA**	**Len**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**	**nM**	**Rank**
AETRGITGY (SEQ ID NO: 7)	Gag polyprotein	212-220	9	38897	66.79	25083	23.55	43055	62.05	36851	49.01	2363	4.72	41857	65.64
QTQAQIPSR (SEQ ID NO: 3)	Gag polyprotein	89-97	9	36382	58.16	587	*1.92*	38529	40.96	35023	41.99	33036	59.51	40584	58.32
MGQIFSRSA (SEQ ID NO: 30)	Gag polyprotein	1-9	9	21573	27.21	30659	34.47	34285	29.41	13225	7.85	12116	16.7	4419	2.01
GYPGRVNEI (SEQ ID NO: 52)	Gag polyprotein	75-83	9	27663	36.65	35250	48.48	519	*0.68*	34319	39.62	35313	66.67	43494	76.25
PRPPRGLAA (SEQ ID NO: 33)	Gag polyprotein	13-21	9	44884	90.62	46010	95.58	43968	67.95	37063	49.98	42808	91.13	43897	78.94
LPTTLFQPA (SEQ ID NO: 61)	Protein Tax-1	213-221	9	22311	28.19	32331	38.93	39658	45.2	247	*0.54*	23256	34.74	8	**0.01**

SASPIPRPPR (SEQ ID NO: 39)	Gag polyprotein	8-17	10	33761	50.39	706	2.11	41402	53	28680	25.21	27677	44.76	32092	26.95
IPRPPRGLAA (SEQ ID NO: 19)	Gag polyprotein	12-21	10	33950	50.94	37380	57.15	41429	53.13	10	**0.03**	24704	37.75	152	**0.20**
YFKLPGLNSR (SEQ ID NO: 6)	Gag-Pro-Pol polyprotein	1396-1405	10	32021	45.9	12301	10.59	28936	20.27	35455	43.54	31447	54.78	35850	37.41
LPMDNALSIA (SEQ ID NO: 28)	Gag-Pro-Pol polyprotein	1343-1352	10	1614	5.21	31263	35.99	37443	37.48	237	*0.53*	12437	17.13	9	**0.01**
SEAETRGITGY (SEQ ID NO: 40)	Gag polyprotein	210-220	11	39884	70.49	30498	34.05	42997	61.69	35812	44.85	3620	6.24	41634	64.29
KTRWQLHHSPR (SEQ ID NO: 23)	Gag-Pro-Pol polyprotein	1367-1377	11	32282	46.54	1879	3.48	35328	31.77	26374	21.08	26627	42.2	39089	50.68
DPIDGRVIGSA (SEQ ID NO: 10)	Protein Tax-1	57-67	11	40071	71.23	43295	85.75	46707	88.48	14902	8.98	41307	86.62	4398	*2.00*
SPIPRPPRGLAA (SEQ ID NO: 43)	Gag polyprotein	10-21	12	35965	56.85	36810	54.71	44220	69.65	204	**0.48**	35201	66.3	1853	*1.06*
PMDNALSIAL (SEQ ID NO: 32)	Gag-Pro-Pol polyprotein	1344-1353	10	4232	8.61	34757	46.7	26732	17.45	25759	20.19	7345	10.74	30177	22.92
LNFLQAAYR (SEQ ID NO: 25)	Gag polyprotein	25-33	9	28961	39.11	1450	3.05	34260	29.36	41888	75.18	28966	48	31613	25.88
QAQIPSRPA (SEQ ID NO: 34)	Gag polyprotein	91-99	9	29990	41.22	31208	35.85	44276	70.03	3476	2.65	16570	22.94	5085	2.26
EAISSLLQA (SEQ ID NO: 13)	Gag-Pro-Pol polyprotein	1266-1274	9	10389	14.81	22940	20.56	40629	49.31	17589	10.98	16755	23.23	794	*0.59*
DTFSGAISA (SEQ ID NO: 12)	Gag-Pro-Pol polyprotein	1248-1256	9	10656	15.06	18906	15.98	41444	53.2	20832	13.98	21867	31.99	1839	*1.06*
GYNPLAGPL (SEQ ID NO: 16)	Gag polyprotein	219-227	9	27446	36.28	31906	37.72	47983		18525	11.82	16924	23.5	39963	54.96

KNTSVLGAGGQ (SEQ ID NO: 22)	Gag-Pro-Pol polyprotein	501-511	11	43138	83.66	34975	47.47	46807	89.23	44713	91.16	36685	71.15	47010	95.64
LPVMHPHGA (SEQ ID NO: 48)	Gag polyprotein	130-138	9	30584	42.53	37864	59.2	44633	72.64	916	*1.16*	30621	52.52	33	**0.05**

*Strong binders (rank < 0.5) shown in **bold**; Weak binders (rank 0.5-2) shown in *italics* Len = length

[00462] Peptide distributions were obtained for all HTLV primary samples tested; however, HTLV peptides were detected only for ED-41214(-) (combined search against human UniProt + HTLV sequence) (**Figure 6A** and **Figures 7A-7D**).

Example 3. Validation of HTLV peptides with heavy analogues in cell lines and calculation of copy number.

[00463] A few HTLV peptides are also selected for validation using the synthetic heavy analogues. Briefly, synthetic peptides containing heavy Leucine ($^{13}C(6)^{15}N(1)$) (Thermo) are mixed with the HLA-eluted peptides from HTLV-transformed cell lines and are analyzed by LC-MS/MS. The co-elution and similarity in fragmentation profile of endogenous and heavy peptides are used to confirm the endogenous peptide identity.

Example 4. Identification of HTLV-specific peptides from HLA-I immunopeptidomics of ATL patient samples.

[00464] ATL (adult T-cell leukemia) patient samples were procured. Peripheral blood mononuclear cells (PBMCs) were suspended in lysis buffer containing 1% NP-40 and incubated at 4°C for 1 hour. HLA class I complexes were then purified using a pan class I anti-HLA W6/32 antibody covalently conjugated to sepharose beads, followed by subsequent washing of the beads and elution of the HLA complexes in acidic solution (pH 2.7). A C18-based solid-phase extraction was utilized to separate the HLA proteins from the associated peptides, which were then further analyzed by liquid chromatography-mass spectrometry (LC-MS). Out of 23 patient samples, 5 samples with cell counts in the range of 50-100 million cells were analyzed by a Thermo Orbitrap mass spectrometer, while the remaining 18 samples (10-20 million cells) were analyzed by a Bruker timsTOF SCP mass spectrometer. PEAKS online (Bioinformatics Solutions Inc.) was used to identify peptide sequences from the mass spectrometer raw files, by searching the raw spectrum against human Uniprot protein sequences. For identification of HLA-I-restricted HTLV-1

peptides, FASTA-containing sequences derived from HTLV RNA reads from the patient samples were used in combination with human Uniprot. Patient DNAseq and RNAseq data were also used to identify HLA genotyping of individual patient samples.

[00465] Analysis of all ATL patient samples revealed an abundance of nonamers representing peptides presented by HLA class I (**Figures 8A-8D**). While 800-2,300 nonamers were isolated from ATL1-5 with the Thermo Orbitrap mass spectrometer (**Figures 8A-8B**), much higher 9-mer peptide counts were recovered from the HLA-I peptidome analysis of ATL6-23 using the Bruker timsTOF SCP mass spectrometer (**Figures 8C-8D**). HLA genotyping of ATL patient samples highlighted the diversity of HLA-I in the patient population with a high proportion of patients with HLA-A02 (12 patient samples), HLA-A11 (6 patient samples) and HLA-A24 (9 patient samples) alleles (**Table 8**).

Table 8. HLA Genotyping of ATL Patient Samples

Lab ID	HLA-A		HLA-B		HLA-C	
ATL1C	*24:02	*26:03	*51:01	*54:01	*01:02	*14:02
ATL2C	*02:01	*31:01	*35:01	*51:01	*03:04	*03:04
ATL3C	*02:01	*02:01	*38:02	*44:03	*07:02	*14:03
ATL4C	*24:02	*33:03	*44:03	*52:01	*12:02	*14:02
ATL5C	*02:01	*24:02	*59:01	*59:01	*01:02	*01:02
ATL6C	*11:01	*68:17	*35:19	*54:01	*01:02	*08:01
ATL7C	*02:06	*33:03	*44:03	*52:01	*12:02	*14:03
ATL8C	*11:01	*31:01	*40:01	*51:01	*03:04	*14:02
ATL9C	*26:01	*26:01	*39:01	*52:01	*07:02	*12:02
ATL10C	*02:06	*26:02	*15:01	*48:01	*03:03	*08:22
ATL11C	*26:03	*31:01	*15:01	*35:01	*03:03	*03:03
ATL12C	*02:01	*24:02	*15:27	*51:01	*04:01	*15:02
ATL13C	*02:07	*31:01	*46:01	*54:01	*01:02	*01:02
ATL14C	*26:01	*26:01	*35:01	*51:01	*03:03	*15:02
ATL15C	*11:01	*24:02	*40:02	*54:01	*01:02	*03:04
ATL16C	*02:06	*24:02	*13:01	*55:02	*01:02	*03:04
ATL17C	*11:01	*24:02	*39:01	*54:01	*01:12	*07:50
ATL18C	*02:01	*11:01	*40:01	*40:02	*03:04	*15:02
ATL19C	*02:06	*26:01	*15:11	*48:01	*03:03	*08:03
ATL20C	*02:01	*31:01	*51:02	*51:02	*03:04	*03:04
ATL21C	*02:06	*26:01	*07:02	*40:02	*03:04	*07:02
ATL22C	*11:01	*24:02	*44:03	*54:01	*01:02	*14:03
ATL23C	*24:02	*26:01	*35:01	*54:01	*01:02	*03:03

[00466] Using analysis from the Thermo Orbitrap mass spectrometer, peptide LPAPHLTLP (SEQ ID NO: 53) originating from the HTLV envelope protein and a predicted HLA-B51 and HLA-B54 binder, was identified from the ATL1C sample, and an 8-mer originating from the HTLV polymerase protein, ALPELQAL (SEQ ID NO: 72), was identified from the ATL1C and ATL2C samples (**Table 9**). However, the two unique HTLV peptides detected at the highest frequencies in the ATL patient samples were QSSSFIFHK (SEQ ID NO: 143) and EYTNIPISLL (SEQ ID NO: 144), as determined using analysis from the Bruker timsTOF SCP mass spectrometer (**Table 10**). Both peptides originated from the HTLV Tax protein and were detected in 2/6 HLA-A11 (33%) and 3/6 HLA-A24 (50%) of patient samples respectively. The HTLV peptides identified from the individual patient samples were matched with their respective HLA alleles using the panNetMHC 4.0 binding prediction algorithm to further enhance the confidence in peptide identifications.

Table 9. HTLV Peptides Detected From ATL1-5 Samples Using Thermo Orbitrap Mass Spectrometer

Sample	Peptide (aa sequence)	Len	Ion Score	Uniprot ID	% Rank_EL (netMHC)
ATL1C	LPAPHLTLP (SEQ ID NO: 53)	9	18	ENV_HTL1M	0.73, 0.07 (HLA-B51:01, HLA-B54:01)
	ALPELQAL (SEQ ID NO: 72)	8	18	POL_HTL1A	0.07 (HLA-C01:02)
ATL2C	ALPELQAL (SEQ ID NO: 72)	8	19	POL_HTL1A	1.52 (HLA-A02:01)

Table 10. HTLV Peptides Detected From ATL6- 23 Samples Using Bruker timsTOF SCP Mass Spectrometer

Uniprot IDs	Peptide (aa sequence)	Len	Ion Scores	Samples	Sample Count	Frequency by Genotype	%Rank_EL (netMHC)
rex_exon1; tax_exon2	QSSSFIFHK (SEQ ID NO: 143)	9	18; 20; 21; 23	ATL16C; ATL17C; ATL18C; ATL21C; ATL23C	5	2/ 6 (HLA-A11)	0.02 (HLA-A11:01)
rex_exon1; tax_exon2	EYTNIPISLL (SEQ ID NO: 144)	10	16; 17; 21; 23	ATL16C; ATL20C; ATL22C; ATL23C	4	3/ 6 (HLA-A24)	0.41 (HLA-A24:02)
rex_exon1; tax_exon2	EYTNIPISL (SEQ ID NO: 145)	9	23	ATL16C	1	1/ 6 (HLA-A24)	0.11 (HLA-A24:02)
gag_pro_pol	LTAFAALPGS (SEQ ID NO: 146)	10	18	ATL22C	1	NA	NA
env	YLFPHWIKK (SEQ ID NO: 56)	9	24	ATL12C	1	1/ 9 (HLA-A02)	1.55 (HLA-A02:01)

Len = Length

Example 5. Generation of a Viral Sequence Database from HTLV-infected Patient Samples.

[00467] An RNAseq-based approach was employed to generate a database of patient-specific HTLV genome sequences directly reconstituted from transcribed virus RNA sequences amplified from patient samples (**Figure 9**). Total human RNA reads were first converted into large contigs using a *de novo* sequencing approach, and contigs with

loose homology to the HTLV reference genome (NCBI Accession No. NC_001436.1) were marked as HTLV-specific sequences. Fully-mapped contigs to the HTLV reference genome were re-arranged in a reference-free fashion to assemble patient-specific HTLV genomic sequences. Using this approach, seven patient-specific HTLV genome sequences were reconstructed. Genome reconstruction was partial, ranging from 13% to 42% of the HTLV reference genome. The patient-specific viral sequence was then used to extract and translate all HTLV coding sequences. Non-canonical protein sequences with start codons embedded within canonical protein sequences were also identified in each sequence. The RNA sequence read coverage of the patient-specific reconstructed genomes is depicted in **Figure 10**.

[00468] A summary of the RNAseq data of the present Example, including patient-specific reconstructed genome length and percentage (%) of the HTLV reference genome, is provided in **Table 11**.

Table 11. Summary of RNAseq data

Lab ID	Reconstructed Genome Length	Reconstructed Genome %
ATL7C	1141	13%
ATL8C	3571	42%
ATL9C	1684	20%
ATL12C	2057	24%
ATL20C	1859	22%
ATL22C	1331	16%
ATL23C	1960	23%

RNA Sequencing

[00469] Total RNA was extracted from human tissue using a MagMAX kit (ThermoFisher). Strand-specific RNA-seq libraries were prepared from 1 μg RNA using a KAPA stranded mRNA-Seq Kit (KAPA Biosystems). Twelve-cycle PCR was performed to amplify libraries. The amplified libraries were size-selected at 400-600 bp (base pairs) using PippinHT (Sage Science). Sequencing was performed on Illumina HiSeq®2500 (Illumina) by multiplexed paired-read run with 2X100 cycles.

Mapping of Patient RNA Sequences to HTLV Reference Genome

[00470] Bulk RNA-seq reads were aligned to the HTLV reference genome (NCBI Accession No. NC_001436.1) using minimap2 (v2.17) (see, e.g., Li 2018). Alignments were then sorted by coordinates and quality controlled by samtools flagstats (v1.9) (see, e.g., Li et al., 2009) and bedtools genomeCoverageBed (v2.17.0) (see, e.g., Quinlan et al., 2010) inspection. Duplicates were then marked and removed by the Picard toolkit (v2.18.2) (github.com/broadinstitute/picard).

Workflow to Reconstruct HTLV Genomes from Patient RNAseq Data

[00471] Paired-end Illumina RNA reads from each sample were *de novo* assembled into large contigs using megahit (see, e.g., Li et al., 2015) (options: --min-count 3 --k-min 27 --k-max 127 --prune-level 2) and mapped using BLAST to the HTLV reference genome (NCBI Accession No. NC_001436.1) to select HTLV-specific sequences. BLAST parameters for sequence comparisons included outfmt '7 std sgi stitle'; minimum E value = 0.001; cost to open a gap = 5; cost to extend a gap = 2; length of best perfect match = 11; reward for a nucleotide match = 2; reward for a nucleotide mis-match = -3. Contigs without BLAST matches were discarded, as well as BLAST results with E values greater than 0.001, percentage identity below 79%, or alignment length of less than 50 nucleotides. Overlapping contigs were merged using custom scripts and final sequences that covered the entire length or partial length of the reference genome sequence with the highest identity were selected. The final sequence was mapped against the reference genome to extract and translate all coding sequences. Non-redundant protein sequences were then added to customized databases for peptide identification by mass-spectrometry.

References

1. Li, H., Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 2018. 34(18): p. 3094-3100.

2. Heng Li, B.H., Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. Bioinformatics, 2009. 25(16): p. 2078-2079.

3. Aaron R Quinlan, I.M.H., BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 2010. 26(6): p. 841-842.

4. Dinghua Li, C.-M.L., Ruibang Luo, Kunihiko Sadakane, Tak-Wah Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 2015. 31(10): p. 1674-1676.

Example 6. Identification of HTLV-1 antigens or epitopes.

[00472] The present Example relates to nucleic acid assay methods and compositions for identifying HTLV-1 antigens or epitopes.

[00473] RNA or DNA is sequenced from a cancer patient's tumor cell(s) and/or healthy tissue(s)/cell(s) to identify sequences which may contain HTLV-1-associated insertions in genes expressed in the tumor cell.

[00474] Tumor material is assessed for HTLV-1-associated insertions by sequencing such as by RNA sequencing. Sequencing data are used to investigate HTLV-1-associated insertions expressed in genes. Peptide stretches comprising any of the identified HTLV-1-associated insertions are created *in silico* and are filtered through the application of prediction algorithms or used to identify MHC-associated epitopes via mass spectrometry data. Where appropriate, suitable MHC-binding HTLV-1 epitope sets are used to identify physiologically relevant HTLV-1 epitope-specific T cell responses via functional assays and/or MHC multimer-based screens within both or either of CD8+ and/or CD4+ T cell populations.

[00475] HTLV-1 epitopes are determined by sequencing the genome and/or exome of tumor tissue and/or healthy tissue from a cancer patient using next generation sequencing (NGS) approaches. Genes selected based on the presence of HTLV-1-associated insertions and ability to act as an antigen are sequenced using NGS technology. NGS applies to, without limitation, genome sequencing, genome resequencing, epigenome characterization, DNA-protein interactions (ChIP-sequencing), and transcriptome profiling (RNA-Seq). Similar to DNA-based assays using, e.g., NGS or massively parallel sequencing (MPS) approaches, RNA from tumors is analyzed by conversion to cDNA and generation of a library suitable for sequencing.

[00476] Assays are employed to identify HTLV-1 epitopes in biological samples. Suitable assays used to identify HTLV-1 epitopes in biological samples include, but are not limited to, proteomics, NGS, solution hybridization, array hybridization nucleic amplification, polymerase chain reaction (PCR), RT-PCR, quantitative PCR, branched

DNA (bDNA) assay, rolling circle amplification (RCA), in situ hybridization, Northern hybridization, hybridization protection assay (HPA), single molecule hybridization detection, Invader assay, and/or Oligo Ligation Assay (OLA), hybridization, and array analysis.

[00477] The sequencing data derived from determining the presence of HTLV-1 epitopes in a cancer patient is analyzed to predict personal HTLV-1 peptides that can bind to HLA molecules of the individual. The data are analyzed using a computer. The sequence data are analyzed, in particular, for the presence of HTLV-1 antigens.

[00478] HTLV-1 antigens are assessed by their affinity to MHC molecules.

[00479] Neural network-based learning approaches with validated binding and non-binding peptides are used in prediction algorithms for the major HLA-A and -B alleles. Algorithms are used for predicting missense mutations that create strong binding peptides to a cancer patient's cognate MHC molecules. A set of peptides representative of optimal HTLV-1 epitopes for each patient is identified and prioritized. HTLV-1 epitope prediction algorithms are used to predict binding of candidate peptides to MHC class I molecules or MHC class II molecules.

[00480] In some cases, a peptide binding tool can be one of the following: Antibody Epitope Prediction, ANTIGENIC, BepiPred, CTLPred, DiscoTope, EPIPREDICT, Epitope Cluster Analysis, Epitope Conservancy Analysis, EUiPro, HLA Peptide Binding Predictions, HLABinding, MAPPP, MHCBench, MHC-I processing predictions, Mosaic Vaccine Tool Suite, NetChop, NetCTL, NetMHC, NetMHCII, NetMHCpan, nHLAPred-I, OptiTope, PAProC, POPI, PREDEP, Prediction of Antigenic Determinants, ProPred, ProPred-1, RankPep, SMM, SVMHC, TAPPred, VaxiJen, or combinations thereof. Additional exemplary programs are used such BIMAS or SYFPEITHI, Rankpep.

[00481] In some cases, the Immune Epitope Database and Analysis Resource (IEDB) (Vita R, et al. Nucleic Acids Res. 2015; 43 (Database issue):D405-D412) is used to identify a suitable tumor HTLV-1 antigen. Such algorithms predict peptide binding to different MHC class I variants based on artificial neural networks (ANN), providing predicted IC_{50} as an output. NetMHC (Lundegaard C, et al. Nucleic Acids Res. 2008;36 (Web Server issue):W509-W512.) is used. Programs such as SMMPMBEC (Kim Y, et al. BMC Bio informatics. 2009; 10:394) and/or SMM (Peters B, et al. BMC Bio informatics.

2005; 6: 132) are used. These programs use position-weight matrices to describe statistical preferences from peptide-MHC I binding data. This approach suppresses noise caused, for example, by a limited number of data points present in the training set and/or experimental error.

[00482] In some instances, single nucleotide polymorphisms (SNPs) are removed from candidate HTLV-1 antigens or HTLV-1 epitopes. SNPs contain a range of molecular variation: (1) SNPs, (2) multinucleotide polymorphisms (MNPs), (3) short deletion and insertion polymorphisms (indels/DIPs), (4) micro satellite markers or short tandem repeats (STRs), (5) heterozygous sequences, and (6) named variants.

[00483] Proteomic-based methods for identifying tumor specific HTLV-1 antigens such as direct protein sequencing are used. Protein sequencing of enzymatic digests using multidimensional MS techniques including tandem mass spectrometry (MS/MS) is used to identify HTLV-1 antigens. High-throughput methods for de novo sequencing of unknown proteins is used, for example, to analyze the proteome of a cancer patient's tumor to identify expressed HTLV-1 antigens. In some instances, meta-shotgun protein sequencing is used to identify expressed HTLV-1 antigens.

[00484] Tumor specific HTLV-1 antigens are identified using MHC multimers to identify HTLV-1 antigen-specific T-cell responses. For example, high-throughput analysis of HTLV-1 antigen- specific T-cell responses in cancer patient samples may be performed using MHC tetramer-based screening approaches. Such tetramer-based screening approaches are used for the identification of tumor specific HTLV-1 antigens, or as a secondary screening protocol to assess HTLV-1 antigens to which a patient may have already been exposed, which may support the selection of candidate HTLV-1 antigens. Where appropriate, filters are applied to eliminate (1) epitopes with lower binding affinity than the corresponding wild-type sequences and/or (2) epitopes predicted to be poorly processed by the immunoproteasome. Candidate mutated peptides are synthesized and screened to identify T cell HTLV-1 antigens.

[00485] Pulsing antigen presenting cells (APCs) with relatively long synthetic peptides that encompass minimal T cell epitopes is used to identify HTLV-1 epitopes. Nonsynonymous mutated epitopes are identified in tumors by evaluating the response of CD4+ tumor infiltrating lymphocytes (TILs) to autologous B cells that are pulsed with

peptides encompassing individual mutations. Use of this approach results in the identification of mutated cell epitopes. A peptide screening assay is performed based on the combination of two peptide libraries: (1) overlapping long-peptides (2) peptides according to MHC-binding prediction. Screening leads to identification of mutated HTLV-1-reactive T cells.

[00486] A tandem minigene screening approach is used to identify HTLV-1 epitopes. A tandem minigene construct comprised, for example, without limitation 6 to 24 minigenes that encoded polypeptides comprising a mutated amino acid residue flanked on the N- and/or C-terminus by, e.g., 12 amino acids. Tandem minigene constructs are synthesized and used to transfect autologous APCs and/or cell lines co-expressing autologous HLA molecules. Using this approach, HTLV-1 epitopes are identified in cancer patients, e.g., patients with adult T-cell leukemia/lymphoma (ATL) or HTLV-1-myelopathy/tropical spastic paraparesis (HAM/TSP).

[00487] HTLV-1 epitopes are identified using an approach combining whole-exome/transcriptome sequencing analysis, MHC binding prediction, and mass spectrometric technique to detect peptides eluted from HLA molecules. Predicted high-binding peptides are confirmed by mass spectrometry.

Example 7. Determination of MHC binding capacity.

[00488] Candidate peptides according to the present disclosure are tested for their MHC binding capacity (affinity). The individual peptide-MHC (pMHC) complexes are produced by UV-ligand exchange. A UV-sensitive peptide is cleaved upon UV-irradiation, and exchanged with the peptide of interest. Peptide candidates that effectively bind and stabilize the peptide-receptive MHC molecules prevent dissociation of the MHC complexes. To determine the yield of the exchange reaction, an ELISA is performed based on the detection of the light chain (β2m) of stabilized MHC complexes. Briefly, 96-well plates are coated with streptavidin, washed, and blocked. Refolded HLA-A monomers serve as standards, covering a pre-determined concentration range. Peptide-MHC monomers of the UV-exchange reaction are diluted in blocking buffer. Samples are incubated, washed, incubated with HRP conjugated anti- β2m, washed again and detected with a chromogenic substrate solution that is stopped per the manufacturer's protocol. Absorption is measured, for example, at 450 nm. Candidate peptides that show a high

exchange yield are preferred for generation and production of antibodies or fragments thereof, and/or T cell receptors or fragments thereof. Candidate peptides demonstrate avidity to the MHC molecules and prevent dissociation of the MHC complexes.

Example 8. Preparation of peptide-MHC (pMHC) complexes.

[00489] This example relates to a method for the preparation of soluble recombinant HLA loaded with an HTLV-1-derived peptide.

[00490] Class I HLA molecules (HLA-heavy chain and HLA light-chain (β2m)) are expressed separately in *E. coli* as inclusion bodies using suitable expression vectors. HLA-heavy chain additionally comprises a C-terminal biotinylation tag which replaces, for example, the transmembrane and/or cytoplasmic domains. *E. coli* cells are lysed and inclusion bodies processed to approximately 80% purity.

[00491] Inclusion bodies of β2m and heavy chain are denatured separately in denaturation buffer. Refolding buffer is prepared. Synthetic peptides are dissolved to a final concentration and added to the refold buffer. Then β2m followed by heavy chain are added. Refolding is performed to completion.

[00492] The refold mixture is then dialyzed. The protein solution is subsequently filtered through a filter and loaded onto an exchange column (pre-equilibrated). Protein is eluted such as by way of a linear salt gradient using additional purifier. HLA-peptide complex is eluted, and peak fractions are collected. A cocktail of protease inhibitors is added and the fractions are chilled on ice.

[00493] Biotin-tagged pHLA molecules are buffer exchanged into a buffer using a fast desalting column equilibrated in the same buffer. Upon elution, the protein-containing fractions are chilled on ice and protease inhibitor cocktail is added. Biotinylation reagents are then added. The mixture is then allowed to incubate.

[00494] The biotinylated pHLA molecules are further purified, for example, by gel filtration chromatography using purifier with a column pre-equilibrated with filtered PBS. The biotinylated pHLA mixture is concentrated to a final volume, loaded onto the column and developed. Biotinylated pHLA molecules elute, e.g., as a single peak. Fractions containing protein are pooled, chilled on ice, and protease inhibitor cocktail is added. Protein concentration is determined and aliquots of biotinylated pHLA molecules are stored frozen.

[00495] Such peptide-MHC (pMHC) complexes are used in soluble form or immobilized through their C-terminal biotin moiety on to a solid support, to be used for the detection of T cells and T cell receptors which bind the peptide-MHC complex. For example, such complexes are used in panning phage libraries, performing ELISA assays and/or preparing sensor chips for measurements of affinity and binding kinetics.

Example 9. Identification of T cell receptors (TCRs) that bind to pMHC complexes.

[00496] Antigen binding T cell receptors (TCRs) are obtained using peptides disclosed herein to pan a TCR phage library. The library is constructed using α- and β-chain sequences obtained from a natural repertoire. The random combination of these α- and β- chain sequences occurs during library creation, thereby producing a non-natural repertoire of α/β chain combinations.

[00497] TCRs obtained from the library are assessed by enzyme-linked immunoassay (ELISA) to confirm specific antigen recognition. ELISA plates are coated with streptavidin and incubated with the biotinylated peptide-HLA complex. TCR-bearing phage clones are added to each well and detection is carried out using an HRP antibody conjugate. Bound antibody is detected using a peroxidase Substrate System. An absence of binding to alternative peptide-HLA complexes indicated that the TCR is not highly cross reactive.

[00498] Further confirmation that TCRs can bind a peptide-HLA complex of the disclosure is obtained by surface plasmon resonance (SPR) using isolated TCRs. In this case α- and β- chain sequences are expressed in *E. coli* as soluble TCRs. Binding of the soluble TCRs to the complexes is analyzed by surface plasmon resonance. Biotinylated peptide-HLA monomers are prepared and immobilized on a streptavidin-coupled sensor chip. To measure affinity, serial dilutions of the soluble TCRs are flowed over the immobilized peptide-HLAs and the response values at equilibrium are determined for each concentration. Data are analyzed, for example, by plotting the specific equilibrium binding against protein concentration followed by a least squares fit to the Langmuir binding equation, assuming a 1:1 interaction.

[00499] TCRs that specifically recognize peptide-HLA complexes of the disclosure are obtained from the library. Data generated according to the above-described experiments confirm that antigen specific TCRs can be isolated.

Example 10. Characterization of binding to MHC and stability of pMHC complex.

[00500] *T2 cell-based peptide binding assay*. T2 cells which do not express the transporter associated with antigen processing (TAP), and as such do not assemble stable MHC class I on the cell surface, are pulsed with different concentrations of peptides (controls or HTLV-1 peptide of interest [POI]), washed, detected with fluorescently-tagged antibody recognizing MHC class I (e.g., A2 allele), and run through a FACS Scan analyzer. The difference between the MFI (mean fluorescence intensity) corresponding to a given concentration of POI and the negative control (non-MHC binder) is a function of the number of stabilized pMHC complexes displayed on the cell surface. Therefore, at limiting concentrations of the peptide, it is largely a measurement of K_{on}, and at saturation levels of the peptide it is a measurement of both K_{on} and K_{off}. The binding is quantified by two related factors: relative affinity (1/RA) and half maximal binding (the peptide concentration responsible for 50% of the signal corresponding to saturation). Relative affinity, RA, is binding normalized to a reference (e.g., a wild-type peptide in instances where a mutant POI is being tested), e.g., the ratio between half maximal binding of control relative to POI. The higher the 1/RA index and the lower the half maximal binding, the higher the K_{on} of the interaction between the POI and the MHC.

[00501] *Characterization of binding and stability by ELISA*. Avidin-coated microtiter plates containing class I monomer loaded with a placeholder peptide are used to evaluate peptide binding, affinity, and off-rate. The monomer-coated plates are supplied as part of a kit, e.g., the iTopia Epitope Discovery System Kit. Assay buffers, anti-MHC-FITC mAb and β2-microglobulin and control peptides are also supplied with the kits.

[00502] Binding assay: POIs are first evaluated for their ability to bind each MHC molecule by binding assay. This assay measures the ability of individual peptides to bind HLA molecules under optimal standardized binding conditions. Monomer-coated plates are first stripped, releasing the placeholder peptide and leaving only the MHC heavy chain bound to the plate. Test peptides are then introduced under optimal folding conditions, along with the anti-MHC-FITC mAb. Plates were incubated. The anti-MHC-FITC mAb binds preferentially to a refolded MHC complex. Thus, the fluorescence intensity resultant from each peptide is related to the peptide's capacity to complex with MHC molecule. Each

peptide's binding is evaluated relative to a positive control peptide, and the results are expressed, for example, as percent (%) binding.

[00503] Affinity assay: For the affinity assay, after the initial stripping of the placeholder peptide, increasing concentrations of POI are added to a series of wells and incubated under the conditions described previously. Plates are read on the fluorometer. Dose response curves are generated. The amount of peptide required to achieve 50% of the maximum is recorded as ED50 value.

[00504] Off-Rate assay: Plates are washed after incubation under conditions to remove excess peptide. The plates are then incubated on allele-specific monomer plates. The plates are measured at multiple time points (e.g., 0, 0.5, 1, 1.5, 2, 4, 6 and 8 hrs) for relative fluorescence intensity. The time required for 50% of the peptide to dissociate from the MHC monomer is defined as the T½ value (hrs).

[00505] iScore calculation: an iScore is a multi-parameter calculation provided within the iTopia software. Its value is calculated based on the binding, affinity, and stability data.

Example 11. Determination of responses against tumor cells.

[00506] A suitable number of groups of mice are immunized with a plasmid expressing HTLV-1 peptides disclosed herein by direct inoculation. By way of a non-limiting example, mice are inoculated into lymph nodes, e.g., inguinal lymph nodes, with plasmids at an appropriate concentration at day 0, and at subsequent days over the time course of the experiment, e.g., at days 3, 14, and 17. In certain cases, this is followed by one or more additional peptide boost(s) on following days, e.g., days 28 and 31, using a negative control peptide and POI. Splenocytes are stimulated *ex vivo* with POI and tested against Chromium-51 (^{51}Cr)-labeled tumor cells at various E:T ratios.

Example 12. *In vivo* assessment of enhanced immunity against HTLV-1 peptides.

[00507] A suitable number of groups of mice are immunized with a plasmid expressing HTLV-1 peptides disclosed herein by direct inoculation. By way of a non-limiting example, mice are inoculated into lymph nodes, e.g., inguinal lymph nodes, with plasmids at an appropriate concentration at day 0, and at subsequent days over the time course of the experiment, e.g., at days 3, 14, and 17. In certain cases, this is followed by

one or more additional peptide boost(s) on following days, e.g., days 28 and 31, using a negative control peptide and POI.

[00508] To evaluate the *in vivo* response against HTLV-1 peptides, splenocytes were isolated from littermate control mice and incubated with one or more appropriate concentrations of POI for a pre-determined period of time. These cells were then stained with CFSEhi fluorescence and intravenously co-injected into immunized mice with an equal number of control splenocytes stained with CFSElo fluorescence. After a pre-determined period of time, the specific elimination of target cells was measured by removing spleen and PBMCs from challenged animals and measuring CFSE fluorescence by flow cytometry. The relative depletion of the populations corresponding to peptide loaded splenocytes was calculated relative to the control (unloaded) population and expressed as percent (%) specific lysis.

Example 13. Testing of POI's increased immunogenicity and ability to overcome tolerization

[00509] POIs are used in *in vitro* for immunization of blood to generate cytotoxic T lymphocytes (CTLs).

[00510] PBMCs from normal donors are purified from buffy coats by centrifugation in standard sterile medium designed for isolating lymphocytes. Cultures are carried out using autologous plasma (AP). For *in vitro* generation of peptide-specific CTL, autologous dendritic cells (DCs) are used as antigen presenting cells (APCs). DCs are generated and CTLs are induced with DCs and peptides from PBMCs. Monocyte-enriched cell fractions are cultured to induce maturation. Specific numbers of CD8+-enriched T lymphocytes and peptide-pulsed DCs are co-cultured. Cultures are restimulated on various days with autologous irradiated peptide-pulsed DCs. Immunogenicity is assayed using *in vitro* cytotoxicity and cytokine production assays.

Example 14. CD8+ T cell responses against peptides.

[00511] Based on the predicted or experimentally verified HLA-binding peptides, whether T cells can be generated to recognize the tumor-specific peptides is determined. Peptides with appropriate binding scores are synthesized. To generate T cells of desired specificity, T cells are stimulated with peptide-pulsed (individual peptide or peptide pool) autologous APCs such as dendritic cells and/or CD40L-expanded autologous B cells on

a predetermined schedule, for example, in the presence of IL-2 and IL-7. After a number of rounds of stimulation, the expanded CD8+ cells are tested on ELISpot for evidence of reactivity against the peptide based on IFNγ secretion.

Example 15. Cytokine production assay.

[00512] For cytokine (e.g., IL-2 and IFNγ) production assays, T cells are harvested after contact with the peptide-pulsed APC, centrifuged, and both cell pellets and supernatants are collected. Cytokine production is measured by ELISA from the supernatant.

Example 16. *In vivo* activation of viral peptide-specific T cells as a method to limit viral production.

[00513] For *in vivo* studies, the ability of the viral peptide in the context of MHC I molecules (e.g., HLA-A2, HLA-A24, HLA-A11) to enhance a CD8+ T cell response as a method to limit viral production following AAV-HTLV-1 challenge is tested in transgenic mice that express human HLA molecule(s):

[00514] (i) MHC class I molecule(s) displaying the HTLV-1 viral peptide of interest (POI) or (ii) a control MHC class I molecule displaying the $OVA_{257-264}$ peptide (OVA) are generated.

[00515] A cohort of transgenic mice that express human HLA molecule(s) is initially injected with the MHC I molecule/peptide or MHC I molecule/OVA complexes and the level of effector and memory CD8+ T cells is monitored at different time points post-injection (from spleen and/or blood). Mock injected mice will serve as a negative control group in this experiment. A second cohort of mice are then injected with the MHC I molecule/POI complex and subsequently challenged using AAV-HTLV-1, at post-injection time points that will be determined according to the measurements of effector and memory CD8 + T cells levels from the first experiment. Control groups are constituted with (i) mice not injected with the MHC I molecule/POI complex but challenged with AAV-HTLV-1, (ii) mice injected with MHC I molecule/OVA complexes, and (iii) mice injected with MHC I molecule/POI complex but not challenged with AAV-HTLV-1. The viral burden is analyzed at different time points post-injection to determine MHC I molecule/POI limitation of viral infection from CD8+ T cells activation.

Example 17. *In vivo* activation of antigen-specific T cells by the viral peptide as a method to decrease viral production.

[00516] For *in vivo* studies, the ability of the HTLV-1 viral peptide of interest (POI) in the context of MHC I molecules (e.g., HLA-A2, HLA-A24, HLA-A11) to enhance a T cell response against HTLV-1 virus following AAV-HTLV-1 challenge is tested in transgenic mice that express human HLA molecule(s). More particularly, the *in vivo* activation of peptide-specific CD8+ T cells after (i) POI immunization, (ii) acute challenge with AAV-HTLV-1, and/or (iii) chronic challenge with AAV-HTLV-1, is compared.

[00517] (i) MHC class I molecule(s) displaying the HTLV-1 viral POI or (ii) a control MHC class I molecule displaying the $OVA_{257-262}$ peptide (OVA) are generated.

[00518] Transgenic mice that express human HLA molecule(s) are first challenged with AAV-HTLV-1. The MHC class I molecule/peptide or MHC class I molecule/OVA complexes produced for this example are then injected 1 day and 7 days after infection with the acute challenge, and 1 day, 7 days, and 14 days after the chronic challenge. The viral burden is analyzed at different time points post-peptide injection by blood collection and post-mortem organs collection. At those time points, the levels of effector and memory CD8+ T cells are measured and compared to the control conditions to determine the capacity of MHC class I molecule/peptide to decrease viral production in response to an AAV-HTLV-1 challenge.

Example 18. Enhancement of a peptide-specific CD8+ T cell immunity in a peptide-presenting tumor model.

[00519] For *in vivo* studies, the ability of the HTLV-1 peptide of interest in the context of MHC I molecules to enhance a CD8+ T cell response against peptide-presenting tumors is tested in a general purpose strain mice, e.g., C57B1/6 mice, grafted with a peptide-expressing tumor cell line.

[00520] The peptide-expressing tumor cell line is first grown *in vitro*, then injected (e.g., subcutaneously) into the mice. After tumor cell injection, when tumors are of palpable size, the HTLV-1 peptide, for example, in a specific groove of the MHC I or control peptide (e.g., ovalbumin [OVA] peptide) in an alternate groove of the MHC I molecule, are injected intratumorally. Tumor cell volume is measured. Tumor, blood, and/or spleen is collected, and homogenized in single cell suspensions, when applicable.

The level of CD8+ T cells in the homogenized samples is determined at all suitable time points.

[00521] Below are additional example methods that may be used in accordance with the disclosure.

[00522] *Splenocytes and antigen presenting cell primary cultures.* Spleens of adult general purpose strain mice, e.g., C57Bl/6 mice, are excised and placed in cold buffer. Tissues are then homogenized to break apart the spleens. Dissociated cells are centrifuged. After centrifugation, the cell pellet is resuspended in a suitable volume of lysis buffer designed to remove red blood cells, e.g., ACK (Ammonium-Chloride-Potassium) lysis buffer, and incubated in the buffer. After incubation the cell suspension is added to buffer, and centrifuged. The cell pellet is then resuspended in buffer and the cell solution is filtered. The filtered cell suspension is centrifuged again, and the subsequent pellet is resuspended in a suitable volume of buffer. Total spleen cells are counted, and global antigen-presenting cells (APC) are sorted using anti-MHC class II Microbeads and appropriate sorting technology. After elution, the MHC class II-positive cell fraction is collected and resuspended at a given concentration in APC cell medium.

[00523] *Pulsing of APCs with peptides.* APC are incubated with HTLV-1 or control peptides. Peptide-pulsed APC are harvested and washed prior to addition to any of various suitable T cell lines (e.g., J.RT3-T3.5 derived cells), after infection with the viral particles. The T cells are co-cultured with the pulsed APC and subsequently used in proliferation, luciferase and/or cytokines production assays.

[00524] *T cell activation post transduction.* Alternatively, to co-culture with peptides-pulsed APCs, suitable T cell lines (e.g., J.RT3-T3.5-derived cell lines) are activated after transduction with either phytohemagglutinin (PHA,), phorbol 12-myristate 13-acetate (PMA), or a combination of PHA and PMA.

[00525] Another alternative method to activate T cells after transduction is to use soluble or immobilized antibodies such as, but not limited to anti-CD3 and/or anti-CD28 monoclonal antibodies. Soluble antibodies are added at an appropriate concentration. In some experiments, plates pre-coated with anti-CD3 antibodies, for example, are used in combination with soluble anti-CD28 antibody.

[00526] Another alternative method to activate T cell lines is activating beads coupled with anti-CD3/anti-CD28 antibodies. After transduction, infected cells are counted, and anti-CD3/anti-CD28 beads are added to the culture medium at a given ratio.

[00527] *Cells staining and FACS analysis.* Fluorescence-activated cell sorting (FACS) is performed after the transduction. Transduced cells are counted and seeded into a cell culture plate. Cells are spun, washed, and then spun again any number of times. Cells are incubated with a dye for determining viability of cells, e.g., Live/Dead™ Fixable Near-IR stain, washed, and incubated with Fc block in buffer, e.g., FACS stain buffer. After being washed any number of times again with FACS stain buffer, cells are subsequently incubated with antibodies such as, but not limited to, fluorescent-labeled antibodies targeting TCR and control. Cells are washed with FACS stain buffer, fixed, washed again, and resuspended in FACS stain buffer. Samples are run for analysis with a FACS analyzer.

[00528] *Proliferation and cytokine production assays.* To measure T cell proliferation, 3H-thymidine is added to assay cultures following contact with the peptide-pulsed APC. Following incubation, cultures are harvested onto filter bottom microplates. MicroScint 20 scintillation fluid or the like is added to each well, and plates are counted on a Scintillation Counter.

[00529] For transcription factor activity analysis and cytokines production assay, cells are harvested after contact with the peptide-pulsed APC, centrifuged, and cell pellets and supernatants are collected.

[00530] Cell pellets are processed for RNA extraction for transcriptomics analysis via qPCR. Cytokine production is measured by ELISA from the supernatant. Cell pellets are processed for luciferase detection assay to measure programmed cell death, e.g., AP1 activity.

[00531] *Splenocytes isolation and CD8+T cell culture.* Spleens of adult general purpose strain mice, e.g., C57Bl/6 mice, are excised and placed in cold buffer. Tissues are then homogenized to break apart the spleens. Dissociated cells are centrifuged. After centrifugation, the cell pellet is resuspended in a suitable volume of lysis buffer designed to remove red blood cells, e.g., ACK (Ammonium-Chloride-Potassium) lysis buffer, and incubated in the buffer. After incubation, the cell suspension is added to buffer, and centrifuged. The cell pellet is then resuspended in buffer and the cell solution is filtered.

The filtered cell suspension is centrifuged again and the subsequent pellet is resuspended in a suitable volume of buffer. CD8+ T cells are then isolated such as, for example, by using a CD8a+ T cell Isolation Kit.

[00532] *Peptide immunization of mice.* For immunization, a suitable amount of HTLV-1 peptide or control peptide is diluted in buffer and emulsified at a given ratio with an adjuvant such as, but not limited to, complete Freund's adjuvant (CFA) or incomplete Freund's adjuvant (IFA) using, for example, a double syringes system, to reach a final desired volume of peptide/adjuvant emulsion. A measured volumed (e.g., 200 µl) of the peptide/adjuvant emulsion is then injected by a given route of administration, for example, subcutaneously, into mice (e.g., C57Bl/6 mice) under appropriate anesthesia in appropriate location(s).

[00533] *Quantification of cytotoxicity activity by flow cytometry.*

[00534] Preparation of target cells: For preparation of target cells (e.g., autologous B cells), the target cells are isolated from splenocytes of OT-1 or P14 mice by using a Mouse B cell Kit (or the like, according to the manufacturer's instruction), and are subsequently stimulated for a predetermined duration in the presence of a set concentration of IFN-γ. Targets cells are then counted, divided into tubes, and washed in buffer. A proportion of the target cells are stained with a high concentration (e.g., 0.2 µM) of CFSE (CFSEHigh) and a separate proportion with a low concentration (e.g, 0.02 µM) of CFSE (CFSELow) in buffer. After the incubation, cells are pelleted and resuspended in an appropriate media to quench the labeling reaction. Target cells stained with the low concentration of CFSE are pulsed by adding the HTLV-1 peptides or control OVA peptides at a suitable final concentration under appropriate culture conditions. Both target cells stained with different concentration of CFSE are then washed, resuspended at a measured concentration in media and mixed with an appropriate ratio, such as a 1:1 ratio (CFSEHigh:CFSELow).

[00535] Effector cells: Total CD8+ T cells containing the effector cells are enriched from splenocytes of P14 and OT-1 mice using a Negative Selection Human CD8 T cell isolation Kit. CD8+ T cells are counted, resuspended in complete T cell medium, and serially diluted volume:volume in a given volume of complete T cell medium. From each dilution, a given volume of cells are seeded in duplicate to multi-well cell culture plate.

The mixed target cells are added to each dilution of effector cells. To measure basal apoptosis, a number of wells are seeded with target cells alone. Cell mixtures are incubated under standard cell culture conditions.

[00536] Flow cytometry staining and acquisition: Cells are transferred to multi-well cell culture plate, washed in FACS staining buffer and stained with, for example, iTag Tetramer/APC-H-2Kb OVA9® (MBL), iTag Tetramer/APC-H-2Db HTLV-1®Alexa Fluor® and Live/Dead® Fixable Near-IR stain (ThermoFisher) under appropriate conditions. Cells are then washed in FACS stain buffer before staining with, for example, Fluorescently-labeled monoclonal antibody that specifically binds to CD8 alpha such as, but not limited to, BV421 αCD8a. Cells are washed with FACS stain buffer and resuspended fixative. Acquisition is performed and all cells are acquired. Post-acquisition data analysis performed.

[00537] *Preparation of vectors encoding peptides.* A therapeutic DNA or RNA vaccine comprising polynucleotides or vectors encoding polynucleotides to be used is prepared by GMP manufacturing of the plasmid vaccine according to regulatory authorities' guidelines. The vaccine is appropriately formulated, for example, by dissolving in a saline solution, at a suitable concentration. The vaccine may be administered either intradermal or intramuscular with or without following electroporation or alternatively with a jet injector.

* * *

[00538] The claimed subject matter is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the claimed subject matter in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

[00539] All patents, applications, publications, test methods, literature, and other materials cited herein are hereby incorporated by reference in their entirety as if physically present in this specification.

CLAIMS

1. An isolated peptide comprising an amino acid sequence that is at least 90% identical to the amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof, wherein the isolated peptide is 5-20 amino acids in length.

2. The isolated peptide of claim 1, wherein the isolated peptide comprises an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146.

3. The isolated peptide of claim 1 or 2, wherein the isolated peptide consists essentially of an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146.

4. The isolated peptide of any one of claims 1-3, wherein the isolated peptide consists of an amino acid sequence of any one of SEQ ID NOs: 1-89 and 143-146.

5. An isolated peptide comprising two or more amino acid sequences selected from any one of SEQ ID NO: 1-89 and 143-146, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof.

6. The isolated peptide of any one of claims 1-5, wherein the isolated peptide comprises one or more reverse peptide bonds, one or more non-peptide bonds, one or more D-isomers of amino acids, one or more chemical modifications, or any combination thereof.

7. The isolated peptide of any one of claims 1-6, wherein the isolated peptide is produced by expression in a heterologous host cell.

8. The isolated peptide of any one of claims 1-6, wherein the isolated peptide is produced synthetically.

9. The isolated peptide of any one of claims 1-8, wherein the isolated peptide, or pharmaceutically acceptable salt thereof, or fragment or derivative thereof induces a

Human T-Lymphotropic Virus Type-1 (HTLV-1)-specific immune response in a subject when presented in a complex with a major histocompatibility complex (MHC) molecule on the surface of an antigen presenting cell (APC).

10. A fusion protein comprising one or more isolated peptides of any one of claims 1-9 fused to one or more heterologous molecules.

11. The fusion protein of claim 10, wherein the one or more heterologous molecules enhance a peptide-specific immune response in a subject.

12. The fusion protein of claim 10, wherein the one or more heterologous molecules mediate peptide delivery to a specific site within a subject.

13. The fusion protein of any one of claims 10-12, wherein the one or more heterologous molecules are a MHC molecule, or a fragment or derivative thereof.

14. A conjugate comprising one or more isolated peptides of any one of claims 1-9 conjugated to one or more heterologous molecules.

15. The conjugate of claim 14, wherein the one or more heterologous molecules enhance a peptide-specific immune response in a subject.

16. The conjugate of claim 14, wherein the one or more heterologous molecules mediate peptide delivery to a specific site within a subject.

17. The conjugate of any one of claims 14-16, wherein the one or more heterologous molecules are an MHC molecule, or a fragment or derivative thereof.

18. The conjugate of any one of claims 14-16, wherein the one or more peptides are conjugated to a particle.

19. An oligomeric complex comprising two or more isolated peptides of any one of claims 1-9.

20. A non-covalent complex comprising the isolated peptide of any one of claims 1-9 and an MHC molecule, or a fragment or derivative thereof.

21. The non-covalent complex of claim 20, wherein the MHC molecule, or the fragment thereof, is a class I MHC molecule.

22. The non-covalent complex of claim 21, wherein the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

23. The non-covalent complex of claim 20, wherein the MHC molecule, or the fragment thereof, is a class II MHC molecule.

24. The non-covalent complex of claim 23, wherein the class II MHC molecule is a class II HLA molecule.

25. A fusion protein comprising the isolated peptide of any one of claims 1-9 and an MHC molecule, or a fragment or derivative thereof.

26. The fusion protein of claim 25, wherein the MHC molecule, or the fragment thereof, is a class I MHC molecule.

27. The fusion protein of claim 26, wherein the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

28. The fusion protein of claim 25, wherein the MHC molecule, or the fragment thereof, is a class II MHC molecule.

29. The fusion protein of claim 28, wherein the class II MHC molecule is a class II HLA molecule.

30. A conjugate comprising the isolated peptide of any one of claims 1-9 and a MHC molecule, or a fragment or derivative thereof.

31. The conjugate of claim 30, wherein the MHC molecule, or the fragment thereof, is a class I MHC molecule.

32. The conjugate of claim 31, wherein the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

33. The conjugate of claim 30, wherein the MHC molecule, or the fragment thereof, is a class II MHC molecule.

34. The conjugate of claim 33, wherein the class II MHC molecule is a class II HLA molecule.

35. A pharmaceutical composition comprising (i) one or more isolated peptides of any one of claims 1-9, one or more fusion proteins of any one of claims 10-13 and 25-29, one or more conjugates of any one of claims 14-18 and 30-34, one or more oligomeric complexes of claim 19, or one or more non-covalent complexes of any one of claims 20-24, or any combination thereof; and (ii) a pharmaceutically acceptable carrier or excipient.

36. The pharmaceutical composition of claim 35, further comprising an adjuvant.

37. An isolated molecule that binds the isolated peptide of any one of claims 1-9, the fusion protein of any one of claims 10-13 and 25-29, the conjugate of any one of claims 14-18 and 30-34, the oligomeric complex of claim 19, or the non-covalent complex of any one of claims 20-24.

38. The isolated molecule of claim 37, wherein the molecule is an antibody or an antigen-binding fragment thereof.

39. The isolated molecule of claim 38, wherein the antibody is a bispecific antibody.

40. The isolated molecule of claim 37, wherein the molecule is an alternative scaffold.

41. The isolated molecule of claim 37, wherein the molecule is a chimeric antigen receptor (CAR).

42. The isolated molecule of claim 37, wherein the molecule is a T cell receptor (TCR).

43. An isolated cell comprising the CAR of claim 41.

44. The isolated cell of claim 43, wherein the isolated cell is an immune cell.

45. The isolated cell of claim 44, wherein the immune cell is a T cell, an NK cell, or a macrophage.

46. An isolated cell comprising the TCR of claim 42.

47. The isolated cell of claim 46, wherein the isolated cell is an immune cell.

48. The isolated cell of claim 47, wherein the immune cell is a T cell, an NK cell, or a macrophage.

49. A pharmaceutical composition comprising (i) the isolated molecule of any one of claims 38-43, or the isolated cell of any one of claims 43-48; and (ii) a pharmaceutically acceptable carrier or excipient.

50. An isolated polynucleotide comprising a nucleotide sequence encoding one or more isolated peptides of any one of claims 1-9 or the fusion protein of any one of claims 10-13 and 25-29.

51. The isolated polynucleotide of claim 50, wherein the nucleotide sequence is operably linked to a promoter.

52. The isolated polynucleotide of claim 50 or claim 51, wherein the isolated polynucleotide comprises DNA.

53. The isolated polynucleotide of claim 50 or claim 51, wherein the isolated polynucleotide comprises RNA.

54. The isolated polynucleotide of claim 53, wherein the RNA is mRNA.

55. The isolated polynucleotide of claim 53, wherein the RNA is self-replicating RNA.

56. A vector comprising the isolated polynucleotide of any one of claims 50-55.

57. The vector of claim 56, wherein the vector is an expression vector.

58. The vector of claim 56 or claim 57, wherein the vector is a viral vector.

59. A host cell comprising the isolated polynucleotide of any one of claims 50-55 or the vector of any one of claims 56-58.

60. The host cell of claim 59, wherein the host cell is a prokaryotic cell.

61. The host cell of claim 59, wherein the host cell is a eukaryotic cell.

62. The host cell of claim 61, wherein the host cell is an APC.

63. A pharmaceutical composition comprising (i) the isolated polynucleotide of any one of claims 50-55, or the vector of any one of claims 56-58; and (ii) a pharmaceutically acceptable carrier or excipient.

64. The pharmaceutical composition of claim 63, wherein the pharmaceutically acceptable carrier is a lipid nanoparticle carrier.

65. A method of inducing an immune response against a HTLV-1infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of:
a) one or more isolated peptides of any one of claims 1-9;
b) the fusion protein of any one of claims 10-13 and 25-29;
c) the conjugate of any one of claims 14-18 and 30-34;
d) the oligomeric complex of claim 19;
e) the non-covalent complex of any one of claims 20-24;
f) the pharmaceutical composition of any one of claims 35, 36, 49, 63, and 64;
g) the molecule of any one of claims 37-42;
h) the isolated cell of any one of claims 43-48 and 59-62;
i) the isolated polynucleotide of any one of claims 50-55; or
j) the vector of any one of claims 56-58.

66. A method of inducing an immune response against a HTLV-1 infection in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of one or more isolated peptides of any one of claims 1-9.

67. A method of inducing an immune response against a HTLV-1 infection in a subject in need thereof, the method comprising administering to the subject an activated T cell that is produced by contacting a T cell with an APC that presents the isolated peptide of any one of claims 1-9 in complex with an MHC molecule.

68. A method of treating a HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of:
a) one or more isolated peptides of any one of claims 1-9;
b) the fusion protein of any one of claims 10-13 and 25-29;
c) the conjugate of any one of claims 14-18 and 30-34;
d) the oligomeric complex of claim 19;
e) the non-covalent complex of any one of claims 20-24;

f) the pharmaceutical composition of any one of claims 35, 36, 49, 63, and 64;
g) the molecule of any one of claims 37-42;
h) the isolated cell of any one of claims 43-48 and 59-62;
i) the isolated polynucleotide of any one of claims 50-55; or
j) the vector of any one of claims 56-58.

69. A method of preventing or reducing the likelihood of a HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of:
a) one or more isolated peptides of any one of claims 1-9;
b) the fusion protein of any one of claims 10-13 and 25-29;
c) the conjugate of any one of claims 14-18 and 30-34;
d) the oligomeric complex of claim 19;
e) the non-covalent complex of any one of claims 20-24;
f) the pharmaceutical composition of any one of claims 35, 36, 49, 63, and 64;
g) the molecule of any one of claims 37-42;
h) the isolated cell of any one of claims 43-48 and 59-62;
i) the isolated polynucleotide of any one of claims 50-55; or
j) the vector of any one of claims 56-58.

70. A method of treating a HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of one or more isolated peptides of any one of claims 1-9.

71. A method of preventing or reducing the likelihood of a HTLV-1-induced disease or disorder in a subject in need thereof, the method comprising administering to the subject an effective amount of one or more isolated peptides of any one of claims 1-9.

72. The method of any one of claims 69-71, wherein the HTLV-1-induced disease or disorder is adult T-cell leukemia/lymphoma (ATL) or HTLV-1-myelopathy/tropical spastic paraparesis (HAM/TSP).

73. A kit comprising:

(i) a) one or more isolated peptides of any one of claims 1-9;

b) the fusion protein of any one of claims 10-13 and 25-29;

c) the conjugate of any one of claims 14-18 and 30-34;

d) the oligomeric complex of claim 19;

e) the non-covalent complex of any one of claims 20-24;

f) the pharmaceutical composition of any one of claims 35, 36, 49, 63, and 64;

g) the molecule of any one of claims 37-42;

h) the isolated cell of any one of claims 43-48 and 59-62;

i) the isolated polynucleotide of any one of claims 50-55; or

j) the vector of any one of claims 56-58; and

(ii) packaging and/or instructions for use for the same.