SPLICING FACTOR GENE MUTATION-ASSOCIATED NEOANTIGENIC PEPTIDES AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This patent application claims the benefit of U.S. Provisional Application No. 63/556,177, filed on February 21, 2024, 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 January 6, 2025, is named 250298_000793_SL.xml and is 152,094 bytes in size.

TECHNICAL FIELD

[0003] The present disclosure relates to methods and compositions that involve isolated peptides derived from splicing factor gene mutation-associated tumors, and the use of such methods and compositions for treatment or prevention of cancer.

BACKGROUND

[0004] Splicing factor genes encode core components of the cellular pre-mRNA (precursor messenger ribonucleic acid) splicing machinery and are frequently mutated in tumors. Splicing factor gene mutations are commonly seen in a variety of hematologic and solid tumors, including, e.g., acral lentiginous melanoma (ALM), acute myeloid leukemia (AML), breast cancer (BRCA), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), esophageal carcinoma (ESCA), myeloproliferative neoplasms (MPN), myelodysplastic syndrome (MDS), mesothelioma (MESO), pancreatic ductal adenocarcinoma (PDAC), primary malignant melanoma (PMM), skin cutaneous melanoma (SKCM), uveal melanoma (UVM), and also pre-tumor syndromes, clonal hematopoiesis of indeterminate potential (CHIP), and clonal cytopenia of undetermined significance (CCUS). Functionally, hotspot mutations in splicing factor genes promote highly reproducible patterns of tumor-specific alternative aberrant pre-mRNA splicing as a result of systemic, aberrant 5′ and 3′ site selection.

[0005] Developing therapeutics that target splicing factor gene mutation-associated neoantigens is an appealing strategy for tumor targeting. However, there is an unmet need for the identification of splicing factor gene mutation-associated aberrant splicing targets that are commonly expressed in the context of splicing factor gene hotspot mutations, and at levels sufficient to be amenable to targeting by therapeutic cancer vaccines, immunotherapy, or other therapeutics that target these splicing factor gene-mutation associated neoantigens.

SUMMARY

[0006] As specified in the Background section above, there is a need in the art for development of methods to identify splicing factor gene mutation-associated aberrant splicing targets, i.e., splicing factor gene mutation-associated epitopes (e.g., neoepitopes) that could be helpful for patients suffering from a splicing factor gene mutation-associated tumor harboring a splicing factor gene mutation. In particular, the present application describes an experimental pipeline for identifying and prioritizing SF3B1 mutation-associated aberrant splicing neoepitopes for targeting via, among other things, a therapeutic cancer vaccine. The therapeutic vaccine can galvanize a polyclonal T cell response against a panel of aberrant splicing neoepitopes that are specifically expressed in SF3B1-mutated tumors. The present application addresses these and other needs.

[0007] In one aspect, provided herein is an isolated peptide comprising an amino acid sequence, wherein the amino acid sequence is at least 90% identical to the amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof, wherein the isolated peptide is about 8-200 amino acids in length.

[0008] In some embodiments, the isolated peptide comprises an amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169.

[0009] In some embodiments, the isolated peptide consists essentially of an amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169.

[0010] In some embodiments, the isolated peptide consists of an amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169.

[0011] In another aspect, provided herein is an isolated peptide comprising two or more amino acid sequences selected from any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof.

[0012] In some embodiments, the isolated peptide comprises a fragment of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169, wherein the fragment is 8-200 amino acids in length.

[0013] 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.

[0014] In some embodiments, the isolated peptide is produced by expression in a heterologous host cell.

[0015] In some embodiments, the isolated peptide is produced synthetically.

[0016] In some embodiments, the isolated peptide, or pharmaceutically acceptable salt thereof, or fragment or derivative thereof, induces an immune response specific for a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in a subject when presented in a complex with a major histocompatibility complex (MHC) molecule, or a fragment or derivative thereof, on the surface of an antigen presenting cell (APC).

[0017] In some embodiments, the tumor is uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

[0018] In some embodiments, the tumor is uveal melanoma.

[0019] In some embodiments, the one or more mutations of SF3B1 is capable of causing aberrant splicing during expression of a gene selected from ariadne RING-between-RING (RBR) E3 ubiquitin protein ligase 1 (ARIH1), chaperonin containing T-complex protein 1 (TCP1) subunit 2 (CCT2), cluster of differentiation 34 (CD34), component of oligomeric golgi complex 1 (COG1), dihydrolipoamide s-succinyltransferase (DLST), diphthamide biosynthesis 5 (DPH5), dishevelled segment polarity protein 2 (DVL2), endoplasmic reticulum (ER)-golgi intermediate compartment protein 3 (ERGIC3), HIG1 hypoxia inducible domain family member 2A (HIGD2A), interleukin 17 receptor C (IL17RC), minichromosome maintenance complex component 3 associated protein (MCM3AP), neuroepithelial cell transforming 1 (NET1), non-POU domain containing octamer binding (NONO), oxidase (cytochrome C) assembly 1-like (OXA1L), plexin B1 (PLXNB1), splicing factor 3a subunit 2 (SF3A2), stromal interaction molecule 1 (STIM1), serine/threonine kinase 24 (STK24), synovial apoptosis inhibitor 1, synoviolin (SYVN1), vacuolar protein sorting 33 homolog B (VPS33B), vacuolar protein sorting 51 homolog (VPS51), zinc finger DHHC-type containing 16 (ZDHHC16), ubiquitin specific peptidase 39 (USP39), bromodomain containing 9

(BRD9), Erythroferrone (ERFE), ubiquitin-like modifier activating enzyme 1 (UBA1), beta-site amyloid beta (β) precursor protein (APP)-cleaving enzyme 1 (BACE1), basal cell adhesion molecule (BCAM), DiGeorge syndrome critical region gene 2 (DGCR2), glycophorin C (GYPC), cluster of differentiation 81 (CD81), glycoprotein non-metastatic melanoma protein B (GPNMB), multiple epidermal growth factor (EGF) like domains 8 (MEGF8), nicastrin (NCSTN), plexin B3 (PLXNB3), amyloid beta (β) precursor protein (APP), calsyntenin 1 (CLSTN1), syndecan 3 (SDC3), protein tyrosine kinase 7 (PTK7), FXYD domain containing ion transport regulator 5 (FXYD5), neuregulin 3 (NRG3), carbonic anhydrase 14 (CA14), matrix metallopeptidase 14 (MMP14), solute carrier family 2 member 11 (SLC2A11), G protein-coupled receptor 143 (GPR143), tumor necrosis factor receptor superfamily member 14 (TNFRSF14), enolase 1 (ENO1), basigin (BSG, also known as CD147), transferrin receptor (TFRC), transmembrane protein 179B (TMEM179B), glutathione peroxidase 1 (GPX1), 5-hydroxytryptamine receptor 2B (HTR2B), and solute carrier family 3 member 2 (SLC3A2).

[0020] In another aspect, provided herein is a fusion protein comprising one or more isolated peptides described herein fused to one or more heterologous molecules.

[0021] In some embodiments, the one or more heterologous molecules enhance a peptide-specific immune response in a subject.

[0022] In some embodiments, the one or more heterologous molecules mediate peptide delivery to a specific site within a subject.

[0023] In some embodiments, the one or more heterologous molecules are an MHC molecule, or a fragment or derivative thereof.

[0024] In some embodiments, the isolated peptide, or pharmaceutically acceptable salt thereof, or a fragment or derivative thereof, induces an immune response specific for a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in the subject when presented in a complex with the MHC molecule, or a fragment or derivative thereof, on the surface of an antigen presenting cell (APC).

[0025] In some embodiments of any of the above-described fusion proteins, the tumor is uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

[0026] In some embodiments, the tumor is uveal melanoma.

[0027] In some embodiments, the one or more mutations of SF3B1 is capable of causing aberrant splicing during expression of a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

[0028] In another aspect, provided herein is a conjugate comprising one or more isolated peptides described herein, conjugated to one or more heterologous molecules.

[0029] In some embodiments, the one or more heterologous molecules enhance a peptide-specific immune response in a subject.

[0030] In some embodiments, the one or more heterologous molecules mediate peptide delivery to a specific site within a subject.

[0031] In some embodiments, the one or more heterologous molecules are an MHC molecule, or a fragment or derivative thereof.

[0032] In some embodiments, the isolated peptide, or pharmaceutically acceptable salt thereof, or a fragment or derivative thereof induces an immune response specific for a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in the subject when presented in a complex with an MHC molecule, or a fragment of derivative thereof, on the surface of an antigen presenting cell (APC).

[0033] In some embodiments, the tumor is uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

[0034] In some embodiments, the tumor is uveal melanoma.

[0035] In some embodiments, the one or more mutations of SF3B1 is capable of causing aberrant splicing during expression of a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

[0036] In some embodiments, the one or more peptides are conjugated to a particle.

[0037] In another aspect, provided herein is an oligomeric complex comprising two or more isolated peptides described herein.

[0038] 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.

[0039] In some embodiments, the MHC molecule, or the fragment or derivative thereof, is a class I MHC molecule.

[0040] In some embodiments, the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

[0041] In some embodiments, the MHC molecule, or the fragment or derivative thereof, is a class II MHC molecule.

[0042] In some embodiments, the class II MHC molecule is a class II HLA molecule.

[0043] 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.

[0044] In some embodiments, the MHC molecule, or the fragment or derivative thereof, is a class I MHC molecule.

[0045] In some embodiments, the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

[0046] In some embodiments, the MHC molecule, or the fragment or derivative thereof, is a class II MHC molecule.

[0047] In some embodiments, the class II MHC molecule is a class II HLA molecule.

[0048] In another aspect, provided herein is a conjugate comprising an isolated peptide described herein and an MHC molecule, or a fragment or derivative thereof.

[0049] In some embodiments, the MHC molecule, or the fragment or derivative thereof, is a class I MHC molecule.

[0050] In some embodiments, the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

[0051] In some embodiments, the MHC molecule, or the fragment or derivative thereof, is a class II MHC molecule.

[0052] In some embodiments, the class II MHC molecule is a class II HLA molecule.

[0053] 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.

[0054] In some embodiments, any of various pharmaceutical compositions described herein may further comprise an adjuvant.

[0055] 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.

[0056] In some embodiments, the isolated molecule is an antibody or an antigen-binding fragment thereof.

[0057] In some embodiments, the antibody is a bispecific antibody.

[0058] In some embodiments, the isolated molecule is an alternative scaffold.

[0059] In some embodiments, the isolated molecule is a chimeric antigen receptor (CAR).

[0060] In some embodiments, the isolated molecule is a T cell receptor (TCR).

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

[0062] In some embodiments, the isolated cell is an immune cell.

[0063] In some embodiments, the immune cell is a T cell, a natural killer (NK) cell, or a macrophage.

[0064] In another aspect, provided herein is an isolated cell comprising a TCR described herein.

[0065] In some embodiments, the isolated cell is an immune cell.

[0066] In some embodiments, the immune cell is a T cell, an NK cell, or a macrophage.

[0067] 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.

[0068] 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.

[0069] In some embodiments, the nucleotide sequence is operably linked to a promoter.

[0070] In some embodiments, an isolated polynucleotide described herein comprises deoxyribonucleic acid (DNA).

[0071] In some embodiments, an isolated polynucleotide described herein comprises ribonucleic acid (RNA).

[0072] In some embodiments, the RNA is messenger (mRNA).

[0073] In some embodiments, the RNA is self-replicating RNA.

[0074] In another aspect, provided herein is a vector comprising an isolated polynucleotide described herein.

[0075] In some embodiments, the vector is an expression vector.

[0076] In some embodiments, the vector is a viral vector.

[0077] In another aspect, provided herein is an isolated cell comprising an isolated polynucleotide described herein or a vector described herein.

[0078] In some embodiments, the isolated cell is a prokaryotic cell.

[0079] In some embodiments, the isolated cell is a eukaryotic cell.

[0080] In some embodiments, the isolated cell is an APC.

[0081] 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.

[0082] In some embodiments, the pharmaceutically acceptable carrier is a lipid nanoparticle carrier.

[0083] In another aspect, provided herein is a method of inducing an immune response against a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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) an oligomeric complex described herein;

e) a non-covalent complex described herein;

f) a pharmaceutical composition described herein;

g) an isolated molecule described herein;

h) an isolated cell described herein;

i) an isolated polynucleotide described herein; or

j) a vector described herein.

[0084] In another aspect, provided herein is a method of inducing an immune response against a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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.

[0085] In another aspect, provided herein is a method of inducing an immune response against a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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 an isolated peptide described herein in complex with an MHC molecule, or a fragment or derivative thereof.

[0086] In another aspect, provided herein is a method of treating a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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) an isolated molecule described herein;

h) an isolated cell described herein;

i) an isolated polynucleotide described herein; or

j) a vector described herein.

[0087] In another aspect, provided herein is a method of preventing or reducing the likelihood of a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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) an isolated molecule described herein;

h) an isolated cell described herein;

i) an isolated polynucleotide described herein; or

j) a vector described herein.

[0088] In another aspect, provided herein is a method of treating a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in a subject in need thereof, the method comprising administering to the subject an effective amount of one or more isolated peptides described herein.

[0089] In another aspect, provided herein is a method of preventing or reducing the likelihood of a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in a subject in need thereof, the method comprising administering to the subject an effective amount of one or more isolated peptides described herein.

[0090] In some embodiments, the isolated peptide, or pharmaceutically acceptable salt thereof, or a fragment or derivative thereof induces an immune response specific for a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in the subject when presented in a complex with an MHC molecule, or a fragment, or derivative thereof on the surface of an antigen presenting cell (APC).

[0091] In some embodiments, the tumor is a uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or endometrial cancer.

[0092] In some embodiments, the tumor is uveal melanoma.

[0093] In some embodiments, the one or more mutations of SF3B1 is capable of causing aberrant splicing during expression of a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

[0094] 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) an isolated 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.

[0095] In another aspect, provided herein is a computer implemented method for identifying a target peptide for cancer treatment for a subject in need thereof, the method comprising:

a) querying a first database comprising a plurality of first gene samples to identify a first gene associated with the plurality of first gene samples, the first gene having a prevalence of mutation beyond a predetermined threshold corresponding to a first cancer type;

b) filtering the plurality of first gene samples by gene samples that correspond to an exon splicing junction mutation;

c) identifying a second gene corresponding to the filtered plurality of first gene samples, wherein the second gene is associated with the first cancer type;

d) generating a peptide sequence based on the filtered plurality of first gene samples, wherein the generated peptide sequence is a sequence associated with the target peptide; and

e) verifying that the generated peptide sequence is associated with the second gene.

[0096] In some embodiments, the first gene is SF3B1.

[0097] In some embodiments, the second gene is ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

[0098] In some embodiments, the filtering the plurality of first gene samples comprises comparing a tumor sample of the first cancer type that is not associated with the exon splicing junction mutation to a tumor sample of the first cancer type that is associated with the exon splicing junction mutation.

[0099] In some embodiments, the filtering the plurality of first gene samples further comprises removing one or more gene samples of the plurality of first gene samples that are associated with copy number variations.

[00100] In some embodiments, a computer implemented method described herein further comprises:

f) categorizing gene samples of the filtered plurality of first gene samples as annotated or non-annotated based on a comparison of the gene samples of the filtered plurality of first gene samples to an annotated gene model,

wherein the second gene is categorized as non-annotated.

[00101] In some embodiments, a computer implemented method described herein further comprises further filtering the plurality of first gene samples by identifying gene samples of the plurality of first gene samples corresponding to a surface protein.

[00102] In some embodiments, the surface protein comprises one or more peptides that form an MHC-peptide complex with an MHC molecule.

[00103] In some embodiments, further filtering the plurality of first gene samples further comprises identifying exon splicing junction mutations that are not expressed in non-cancerous tissue.

[00104] In some embodiments, verifying that the generated peptide sequences are associated with the second gene further comprises performing mass spectrometry on an MHC-peptide complex associated with the cancer to confirm a presence of the generated peptide sequence.

[00105] In some embodiments, the first database comprises a Genome Reference Consortium (GRC) Human Build.

[00106] In another aspect, provided herein is a computer implemented method to determine viability of a cancer treatment for a subject in need thereof, the method comprising:

a) obtaining a ribonucleic acid (RNA) sequence from a sample obtained from the subject;

b) identifying a non-annotated exon splice junction in the RNA sequence from the sample;

c) obtaining an expression value based at least in part on a count of non-annotated exon splice junctions associated with a gene mutation of a target peptide of the cancer treatment;

d) comparing the expression value to a count of non-annotated exon splice junctions in a sample lacking the gene mutation; and

e) determining the viability of the cancer treatment for the subject based at least in part on the comparison of the expression value to the count of non-annotated exon splice junctions in the sample lacking the gene mutation.

[00107] In some embodiments, the subject has uveal melanoma, breast cancer, ovarian cancer, prostate cancer, skin cutaneous melanoma, acute myeloid leukemia, myelodysplastic syndrome, or chronic lymphocytic leukemia.

[00108] In some embodiments, the gene mutation comprises a mutation of SF3B1.

[00109] In some embodiments, the target peptide is derived from a second gene comprising ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

[00110] In another aspect, provided herein is a computer implemented method for identifying a target peptide for treatment of a disease, the method comprising:

a) identifying a splicing factor gene having a significant prevalence of mutation in the disease in a database, wherein the database includes genetic data of samples of the disease;

b) identifying differently expressed splicing junctions of the splicing factor gene based on a comparison of splicing junctions of genetic data of samples that have a mutation in the splicing factor gene compared to splicing junctions of genetic data of samples that lack a mutation in the splicing factor gene; and

c) determining the target peptide based at least in part on the differently expressed splicing junctions.

[00111] In some embodiments of the above-described computer implemented methods, the method further comprises:

d) identifying a disease-specific splicing junction based on one or more MHC-peptide complexes associated with the disease, such that the disease-specific splicing junction has minimal expression in non-diseased samples;

e) identifying a disease-specific mutated splicing junction as being both a disease-specific splicing junction and a differently expressed splicing junction; and

f) determining the target peptide based at least in part the disease-specific mutated splicing junction.

[00112] In some embodiments, the disease-specific mutated splicing junction comprises an SLC3A2-APIS splice variant junction.

[00113] In some embodiments, the target peptide is derived from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

[00114] In some embodiments, the splicing factor gene comprises SF3B1.

[00115] In another aspect, provided herein is a computer implemented method for identifying a target peptide for targeted therapy development, the method comprising:

a) identifying one or more candidate target peptides from differently expressed alternative splicing junction events in one or more patient samples based on a comparison of the one or more patient samples to a control (e.g., a wild-type control);

b) filtering the one or more candidate target peptides based at least in part on a raw read count to find disease-specific alternative splicing junction events resulting in a filtered list of the one or more candidate target peptides;

c) screening the filtered list of the one or more candidate target peptides based at least in part on protein function and structural domain to separate the filtered list of the one or more candidate target peptides into a first group comprising peptides derived from proteins expressed on a cell surface and a second group comprising peptides derived from proteins not expressed on the cell surface;

d) eliminating peptides with normal tissue expression from the first group based at least in part on a comparison of peptides of the first group to a normal tissue atlas;

e) retaining peptides with tumor type targeting in the first group based on a comparison of peptides of the first group to a tumor tissue atlas;

f) validating remaining peptides of the first group as comprising alternative splicing-derived peptide sequences based at least in part on mass-spectrometry of the remaining peptides of the first group;

g) eliminating peptides with normal tissue expression from the second group based on a comparison of peptides of the first group to a normal tissue atlas comprising a human thymus alternative splicing atlas database;

h) validating remaining peptides of the second group as comprising HLA presented alternative splicing-derived peptide sequences based at least in part on immunopeptidomics of the remaining peptides of the second group; and

i) prioritizing at least one target peptide comprising a validated peptide from the first group as identified through steps (a) through (f) and/or comprising a validated peptide from the second group as identified through steps (a) through (c), (g), and (h).

[00116] In some embodiments the at least one target peptide prioritized at step (i) is derived from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2, or SLC3A2.

[00117] In another aspect, provided herein is a method for identifying one or more target peptides for treatment of a disease, the method comprising:

a) identifying one or more alternatively spliced isoforms that are differentially expressed in RNA-seq samples derived from diseased cells to generate a working list of isoforms, wherein the diseased cells are associated with the disease for treatment, the one or more identified alternatively spliced isoforms are expressed at a higher frequency in a subset of the RNA-seq samples associated with a splicing factor gene mutation than in a subset of the RNA-seq samples that are not associated with the splicing factor gene mutation, and a targetable peptide sequence is associated with each identified isoform such that the targetable peptide sequence

distinguishes the identified isoform from other variants over which the identified isoform was differentially expressed;

b) performing one or more filtering steps on the working list of isoforms, wherein the working list of isoforms is updated after performance of each filtering step and the one or more filtering steps comprise at least one of the following:

1) filtering the working list of isoforms to filter out any isoforms that were expressed in the subset of RNA-seq samples not associated with the splicing factor gene mutation above a control sample threshold frequency, optionally wherein the control sample threshold frequency is 0%;

2) filtering the working list of isoforms to filter out any isoforms that are expressed in normal tissue above a normal tissue expression level threshold, optionally wherein the normal tissue expression level threshold is no detectable expression;

3) filtering the working list of isoforms to filter out any isoforms that are not associated with a gene that encodes a protein that is expressed on the cell surface;

4) filtering the working list of isoforms to filter out any isoforms that are not associated with a targetable peptide sequence that is aligned with at least a portion of an extracellular domain of a protein expressed on the cell surface;

5) filtering the working list of isoforms to filter out any isoforms that i) result in the loss of a transmembrane domain and/or an extracellular domain, and/or ii) result in the truncation of an extracellular domain; and

6) filtering the working list of isoforms to retain only isoforms for which the targetable pep-tide sequence introduces i) an insertion of an isoform-specific amino acid sequence into an extracellular domain of a protein expressed on the cell surface and/or ii) a deletion of one or more amino acids in the extracellular domain, wherein the insertion and/or the deletion distinguishes the identified isoform from other variants over which the identified isoform was differentially expressed in step (a); and

c) identifying the one or more target peptides for treatment of the disease from the working list of isoforms following performing the one or more filtering steps in step (b).

[00118] In some embodiments of the above-described method for identifying one or more target peptides for treatment of a disease:

(i) the method comprises filtering steps (b)(1) through (b)(2), optionally performed in that order;

(ii) the method comprises filtering steps (b)(1) through (b)(4), optionally performed in that order;

(iii) the method comprises filtering steps (b)(1) through (b)(5), optionally performed in that order; or

(iv) the method comprises filtering steps (b)(1) through (b)(6), optionally performed in that order.

[00119] In some embodiments, the isoform-specific amino acid sequence comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more contiguous amino acids.

[00120] In some embodiments, the deletion in step (b)(6) is of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more contiguous amino acids.

[00121] In some embodiments, the deletion in step (b)(6) is of no more than 10, 15, 20, 25, or 30 contiguous amino acids, optionally wherein the deletion does not occur at an N-terminal and/or a C-terminal portion of the extracellular domain.

[00122] In some embodiments, the isoform-specific amino acid sequence and/or deletion results from a frame shift relative to another isoform over which the identified isoform was differentially expressed in step (a).

[00123] In some embodiments, performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only identified isoforms which are expressed in the subset of RNA-seq samples associated with the splicing factor gene mutation above a mutation sample threshold frequency.

[00124] In some embodiments, performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only identified isoforms which are expressed in the subset of RNA-seq samples associated with the splicing factor gene mutation above a mutation sample expression level threshold, optionally wherein expression is measured by raw read counts or normalized expression values.

[00125] In some embodiments, performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only identified isoforms which are expressed in one or more additional diseases above a secondary disease threshold frequency, wherein the one or more additional diseases are different from the dis-ease which the diseased

cells were associated with in step (a), optionally wherein the disease for treatment and one or more additional diseases are different types of cancer.

[00126] In some embodiments, performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only identified isoforms which are expressed in one or more additional diseases above a secondary disease expression level threshold, wherein the one or more additional diseases are different from the disease which the diseased cells were associated with in step (a), optionally wherein the dis-ease for treatment and one or more additional diseases are different types of cancer.

[00127] In some embodiments, identifying the one or more alternatively spliced isoforms comprises aligning RNA-seq reads to a Genome Reference Consortium (GRC) Human Build and calling exon-exon junctions within the reads.

[00128] In some embodiments, the splicing factor gene mutation has a prevalence among the RNA-seq samples derived from diseased cells above a mutation prevalence threshold frequency, optionally wherein the method further comprises selecting the splicing factor gene mutation by querying a database of RNA-seq samples.

[00129] In some embodiments, the RNA-seq samples exclude any samples associated with a copy number variation of the splicing factor gene, optionally wherein step (a) comprises filtering out any RNA-seq samples that are associated with a copy number variation of the splicing factor gene.

[00130] In some embodiments, performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only isoforms that have been experimentally detected via mass spectrometry, optionally comprising performing the mass spectrometry.

[00131] In some embodiments of the above-described method for identifying one or more target peptides for treatment of a disease, the method further comprises generating a working list of peptide sequences for MHC binding, wherein each peptide sequence of the working list of peptide sequences for MHC binding (i) is a length suitable for binding an MHC molecule, (ii) is derivable from a targetable peptide sequence, and (iii) is not derivable from any other variants over which the corresponding identified isoform was differentially expressed in step (a), optionally wherein the working list of isoforms is filtered to retain only isoforms corresponding to a peptide sequence from the working list of peptide sequences for MHC binding.

[00132] In some embodiments, generating a working list of isoforms in step (a) comprises generating a working list of peptide sequences for MHC binding, wherein each peptide sequence of the working list of peptide sequences for MHC binding (i) is a length suitable for binding an MHC molecule, (ii) is derivable from a targetable peptide sequence, and (iii) is not derivable from any other variants over which the corresponding identified iso-form is differentially expressed.

[00133] In some embodiments, the MHC molecule is an MHC class I molecule, optionally wherein the length suitable for binding is 8-12, 8-10, or 9 amino acids in length.

[00134] In some embodiments, the MHC molecule is an MHC class II molecule, optionally wherein the length suitable for binding is 13-17 amino acids in length.

[00135] In some embodiments, performing the one or more filtering steps in step (b) further comprises filtering the list of working peptide sequences for MHC binding to retain only peptide sequences which are predicted to bind an MHC molecule.

[00136] In some embodiments, performing the one or more filtering steps in step (b) further comprises filtering the working list of peptide sequences for MHC binding to retain only peptide sequences that have been experimentally detected via mass spectrometry as bound to an MHC molecule, optionally comprising performing the mass spectrometry.

[00137] In some embodiments, the disease for treatment is a cancer.

[00138] In some embodiments, the disease for treatment is uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

[00139] In some embodiments, the splicing factor gene mutation is a mutation of splicing factor 3b subunit 1 (SF3B1).

[00140] In some embodiments, the one or more identified isoforms in step (a) are associated with a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

[00141] In some embodiments, the one or more identified target peptides in step (c) are associated with a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2,

ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

[00142] In some embodiments, the one or more identified isoforms in step (a) and/or the one or more identified target peptides in step (c) comprise an amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, 169.

[00143] In some embodiments, step (a) and/or one or more of the performed filtering steps in step (b), optionally each of the filtering steps performed in step (b), is computer implemented.

[00144] In some embodiments of the above-described method for identifying one or more target peptides for treatment of a disease, the method further comprises generating a peptide that comprises the targetable peptide sequence of one of the target peptides and/or generating a nucleic acid that encodes the same.

[00145] In another aspect, provided herein is a system comprising one or more processors and a memory storing processor-executable instructions thereon that, when executed by the one or more processors, cause the one or more processors to perform a method for identifying one or more target peptides for treatment of a disease described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[00146] **Fig. 1** is a flow diagram of a method disclosed herein for identifying a target peptide for cancer treatment.

[00147] **Fig. 2** is a flow diagram of method disclosed herein for identifying a target peptide for cancer treatment.

[00148] **Fig. 3** is a flow diagram of a method disclosed herein for determining viability of a cancer treatment for a subject in need of treatment.

[00149] **Fig. 4** is an illustration of a computing device configured to execute steps of a method presented herein.

[00150] **Fig. 5** is an illustration of a computing network configured to execute steps of methods presented herein.

[00151] **Fig. 6** is a flow diagram of a method of an example implementation of identifying a target peptide for cancer treatment.

[00152] **Fig.** 7 is a flow diagram of a method of an example implementation of determining viability of a cancer treatment for a subject in need of treatment.

[00153] **Fig. 8** is a heatmap plotting expression of SF3B1 mutant samples (MUT) and wild-type (WT).

[00154] **Fig. 9** is a heatmap plotting expression of SF3B1 mutant samples (MUT), wild-type (WT), and other splicing factors with mutations (other-MUT).

[00155] **Fig. 10** is a heatmap plotting expression of each of the 778 non-annotated junctions for 212 Melanoma patient-derived xenograft (PDX) samples.

[00156] **Fig. 11** is a heatmap plotting expression of 32 non-annotated junctions for the 69 UVM samples plotted in **Fig. 8**.

[00157] **Fig. 12** is a heatmap plotting expression of the 32 non-annotated junctions plotted in **Fig. 11** for 1000 breast cancer (BRCA) tumor samples, 19 with SF3B1 mutation (left in plot).

[00158] **Fig. 13** is a heatmap plotting expression of the 32 non-annotated junctions plotted in **Fig. 11** for 471 skin cancer melanoma tumor samples, 23 with SF3B1 mutation (left in plot).

[00159] **Fig. 14** is a plot of SYVN1-splice variant junctions for 10,550 samples from 33 cancer types (x-axis).

[00160] **Figs. 15A-15C** show a summary of SYVN1-splice variants in The Cancer Genome Atlas (TCGA). The plot depicted in **Fig. 15C** shows fold elevated expression of SYVN1-splice variants in mutated samples compared to wild-type (MUT/WT) samples, with the cancer types annotated along the x-axis.

[00161] **Fig. 16** is a plot of SYVN1-splice variant junctions in MDS for 81 samples with elevated expression in mutated (MUT) samples (left) compared to wild-type (WT) samples (right).

[00162] **Fig. 17** is a plot of SYVN1-splice variant junctions in chronic lymphocytic leukemia (CLL) for 84 samples with elevated expression in mutated (MUT) samples (left) compared to wild-type (WT) samples (right).

[00163] **Fig. 18** is a plot of OXA1L-splice variant junctions in TCGA for 10,550 samples from 33 cancer types (x-axis).

[00164] **Figs. 19A-19C** show a summary of OXA1L-splice variants in TCGA. The plot depicted in **Fig. 19C** shows fold elevated expression of OXA1L-splice variants in mutated samples compared to wild-type (MUT/WT) samples, with the cancer types annotated along the x-axis.

[00165] **Fig. 20** is a plot of OXA1L-splice variant junctions in CLL for 84 samples with elevated expression in mutated (MUT) samples (left) compared to wild-type (WT) samples (right).

[00166] **Fig. 21** is a plot of OXA1L-splice variant junctions in MDS for 81 samples with elevated expression in mutated (MUT) samples (left) compared to wild-type (WT) samples (right).

[00167] **Fig. 22** is a plot of NET1-splice variant junctions in TCGA for 10,550 samples from 33 cancer types (x-axis).

[00168] **Figs. 23A-23C** show a summary of NET1-splice variants in TCGA. The plot depicted in **Fig. 23C** shows fold elevated expression of NET1-splice variants in mutated samples compared to wild-type (MUT/WT) samples, with the cancer types annotated along the x-axis.

[00169] **Fig. 24** is a plot of NET1-splice variant junctions in CLL for 84 samples with elevated expression in mutated (MUT) samples (left) compared to wild-type (WT) samples (right).

[00170] **Fig. 25** is a plot of NET1-splice variant junctions in MDS for 81 samples with elevated expression in mutated (MUT) samples (left) compared to wild-type (WT) samples (right).

[00171] **Fig. 26A** is a flow diagram of a method for prioritizing target peptides from the pool of all identified peptides for treatment of a disease (e.g., cancer) as disclosed herein.

[00172] **Figs. 26B-26C** are flow diagrams showing another embodiment of the method for prioritizing target peptides from the pool of all identified peptides for treatment of a disease (e.g., cancer) as disclosed herein.

[00173] **Figs. 27A-27B** shows a summary of ENO1-splice variants in The Cancer Genome Atlas (TCGA) with elevated expression in mutated (MUT) samples compared to wild-type (WT) samples across various tumor types.

[00174] **Fig. 28** shows a plot of the SLC3A2-APIS splice variant junction for 10,550 samples from 33 cancer types (x-axis) as provided in The Cancer Genome Atlas (TCGA).

[00175] **Fig. 29** shows a summary of the SLC3A2-APIS splice variant junction in The Cancer Genome Atlas (TCGA). The plot shows fold elevated expression of SLC3A2-APIS splice variants in mutated samples compared to wild-type (MUT/WT) samples, with the cancer types annotated along the x-axis.

[00176] **Fig. 30** shows a plot of the SLC3A2-APIS splice variant junction in myelodysplastic syndrome (MDS) for 81 samples with elevated expression in mutated samples (left) compared to wild-type samples (right).

[00177] **Fig. 31** shows a plot of the SLC3A2-APIS splice variant junction in chronic lymphocytic leukemia (CLL) for 84 samples with elevated expression in mutated samples (left) compared to wild-type samples (right).

DETAILED DESCRIPTION

[00178] Splicing factor mutations are a class of driver mutations in human cancers that occur in a range of tumor types, including solid and hematological malignancies. Splicing factor mutations are typically heterozygous and occur in a mutually exclusive manner, thereby suggesting a common impact of these mutations on pre-mRNA splicing. The most commonly mutated splicing factor in cancer is SF3B1 (splicing factor 3b subunit 1, also known as SF3B155 or SAP155), which participates in the early steps of RNA splicing, e.g., 3′ splice-site recognition. Mutations in SF3B1 have been reported in a diverse series of tumor types, including acral lentiginous melanoma (ALM), acute myeloid leukemia (AML), breast cancer (BRCA), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), esophageal carcinoma (ESCA), myeloproliferative neoplasms (MPN), myelodysplastic syndrome (MDS), mesothelioma (MESO), pancreatic ductal adenocarcinoma (PDAC), primary malignant melanoma (PMM), skin cutaneous melanoma (SKCM), and also in pre-tumor syndromes such as clonal hematopoiesis of indeterminate potential (CHIP) and clonal cytopenia of undetermined significance (CCUS). SF3B1 mutations can lead to changes of the splice acceptor sites resulting in inclusion or exclusion of nucleic acids in the splice mRNA product, thereby leading to changes to the peptide product encoded by such alternatively spliced mRNA. The resultant additional or deleted amino acids, or the frame-shift, can generate shared neoepitopes common to tumors harboring SF3B1 mutations. These abnormal polypeptides can be loaded on the MHC-class I molecules, which may generate potential neoepitopes.

[00179] The present disclosure provides, among other things, isolated peptides derived from splicing factor gene mutation-associated tumors, and fragments or derivatives thereof. Various peptide-based molecules including complexes (e.g., 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 splicing factor gene mutation-associated tumors and/or for treatment or prevention of a splicing factor gene mutation-associated tumor and/or a splicing factor gene mutation-induced disease or disorder. In one aspect, the present disclosure also provides methods for identifying an immunogenic splicing factor gene mutation-associated neoantigenic peptide.

[00180] 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.

[00181] 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). Most T cells have an $\alpha\beta$ TCR, while a minority of T cells bear a $\gamma\delta$ 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).

[00182] 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.

[00183] Immunotherapeutic approaches for treating disease work to regulate T cell activity *in vivo*, e.g., to enhance anti-infection and antitumor 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.

[00184] The presentation of splicing factor gene mutation-associated neoantigenic peptides by MHC molecules on the surface of a cancer cell, and the recognition of these pMHC complexes by, and subsequent activation of, CD8+ cytotoxic T cells can provide an important mechanism for immunity-based protection against cancer. Cancer cells having splicing alterations induced by splicing factor gene mutation(s) can express various splicing factor gene mutation-associated antigens. Peptides derived from these antigens may be displayed on the cell surface in complex with MHC molecules. Detection of an MHC-presented splicing factor gene mutation-associated neoantigenic peptide by a T cell bearing the corresponding TCR can lead to targeted killing of the cancer cell.

[00185] The identification of splicing factor gene mutation-associated neoantigenic peptides presented on cancer cells expressing splicing factor gene mutation(s) can allow for the development of immunotherapeutic reagents designed to specifically target and destroy tumors associated with one or more mutations of splicing factor gene. Such reagents may be moieties that bind to the splicing factor gene mutation-associated neoantigenic 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.

[00186] The present disclosure, in part, is based on a proteogenomic approach that detects MHC-associated splicing factor gene mutation-associated neoantigenic peptides from cancer cells comprising splicing factor gene mutation(s). The repertoire of splicing factor gene mutation-

associated neoantigenic peptides can provide an accurate representation of splicing factor gene mutation-associated splicing epitopes in a population.

[00187] 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 splicing factor gene mutation-associated tumors.

[00188] The splicing factor gene mutation-associated neoantigenic peptides disclosed herein may be used in the development of vaccines or immunotherapeutics (such as antibodies (e.g., bispecific antibodies), engineered TCR- or CAR-based cellular therapies) for the treatment of splicing factor gene mutation-associated diseases or disorders including tumors, e.g., uveal melanoma, and to improve clinical surveillance to monitor the diversity of patient-specific neoantigenic epitopes associated with splicing factor gene mutations and to inform vaccine development. The proteogenomic discovery platform described herein provides a method for identifying splicing factor gene mutation-associated neoantigenic peptides as targets for anti-tumor related immunotherapy.

Definitions

[00189] 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.

[00190] 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.

[00191] 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.

[00192] The term "antigen" encompasses any agent (e.g., protein, peptide, polysaccharide, glycoprotein, glycolipid, nucleotide, portions thereof, etc., 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 the 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 an MHC may also be referred to as an "epitope," "neoantigenic epitope," or an "antigenic determinant." The terms "peptide," "antigenic determinant," "epitope," "neoantigenic epitope," 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.

[00193] 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 upon exposure to denaturing solvents, whereas epitopes formed by tertiary folding are typically lost upon treatment with denaturing solvent.

[00194] 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 that both classes of MHC can bind a core of about 9 amino acids (e.g., 5 to 17 amino acids) within peptides, the open-ended nature of the 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 peptide (e.g., 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 a 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.

[00195] "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, 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 (TCR). A pMHC complex encompasses cell-surface-expressed pMHC complexes and soluble pMHC complexes.

[00196] "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.

[00197] 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 (Treg) cells, 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.

[00198] 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.

[00199] The 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.

[00200] The term "specifically binds," "binds in a specific manner," "antigen-specific," or the like, indicates that the molecules involved in the specific binding can 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 a splicing factor gene mutation-associated neoantigenic 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 splicing factor gene mutation-associated neoantigenic 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 a splicing factor gene mutation-associated neoantigenic 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.

[00201] 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."

[00202] 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. Polynucleotides can 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.

[00203] 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 (e.g., Kozak consensus sequences); 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.

[00204] The term "isolated," as used herein, 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.

[00205] 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.

[00206] 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; (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.

[00207] 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.

[00208] 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. 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.

[00209] 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.

[00210] 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, intraarterial, 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.

[00211] 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

[00212] In one aspect, the present disclosure provides isolated peptides comprising an amino acid sequence derived from splicing factor 3B subunit 1 (SF3B1) mutation-associated neoantigens.

[00213] SF3B1 is a frequently mutated splicing factor gene in cancer. Specifically, SF3B1 is commonly mutated across a wide range of hematologic and solid tumors. Non-limiting examples of tumor types in which SF3B1 mutations are frequently observed include acral lentiginous melanoma (ALM), acute myeloid leukemia (AML), breast cancer (BRCA), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), esophageal carcinoma (ESCA), myeloproliferative neoplasms (MPN), myelodysplastic syndrome (MDS), mesothelioma (MESO), pancreatic ductal adenocarcinoma (PDAC), primary malignant melanoma (PMM), skin cutaneous melanoma (SKCM), uveal melanoma (UVM), and also pre-tumor syndromes, clonal hematopoiesis of indeterminate potential (CHIP) and clonal cytopenia of undetermined significance (CCUS). SF3B1 hotspot mutations can rewire the pre-mRNA splicing machinery in tumor cells. Rewiring of the pre-mRNA splicing machinery in tumor cells can result in the expression of numerous shared, tumor-specific, immunogenic polyepitope splicing neoantigens. Hotspot mutations result in reproducible patterns of alternative splicing in tumor cells, giving rise to shared, tumor-specific, poly-epitope splicing neoantigens.

[00214] SF3B1 protein (UniProtKB: 075533; SEQ ID NO: 87) in humans is encoded by the SF3B1 gene (Ensembl: ENSG00000115524). An example human SF3B1 protein sequence is the

amino acid sequence
MAKIAKTHEDIEAQIREIQGKKAALDEAQGVGLDSTGYYDQEIYGGSDSRFAGYVTSIA
ATELEDDDDDYSSSTSLLGQKKPGYHAPVALLNDIPQSTEQYDPFAEHRPPKIADREDEY
KKHRRTMIISPERLDPFADGGKTPDPKMNARTYMDVMREQHLTKEEREIRQQLAEKAK
AGELKVVNGAAASQPPSKRKRRWDQTADQTPGATPKKLSSWDQAETPGHTPSLRWDE
TPGRAKGSETPGATPGSKIWDPTPSHTPAGAATPGRGDTPGHATPGHGGATSSARKNR
WDETPKTERDTPGHGSGWAETPRTDRGGDSIGETPTPGASKRKSRWDETPASQMGGST
PVLTPGKTPIGTPAMNMATPTPGHIMSMTPEQLQAWRWEREIDERNRPLSDEELDAMFP
EGYKVLPPPAGYVPIRTPARKLTATPTPLGGMTGFHMQTEDRTMKSVNDQPSGNLPFLK
PDDIQYFDKLLVDVDESTLSPEEQKERKIMKLLLKIKNGTPPMRKAALRQITDKAREFGA
GPLFNQILPLLMSPTLEDQERHLLVKVIDRILYKLDDLVRPYVHKILVVIEPLLIDEDYYA
RVEGREIISNLAKAAGLATMISTMRPDIDNMDEYVRNTTARAFAVVASALGIPSLLPFLK
AVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSLVEIIEHGLVDEQQKVRTISALAIA
ALAEAATPYGIESFDSVLKPLWKGIRQHRGKGLAAFLKAIGYLIPLMDAEYANYYTREV
MLILIREFQSPDEEMKKIVLKVVKQCCGTDGVEANYIKTEILPPFFKHFWQHRMALDRR
NYRQLVDTTVELANKVGAAEIISRIVDDLKDEAEQYRKMVMETIEKIMGNLGAADIDHK
LEEQLIDGILYAFQEQTTEDSVMLNGFGTVVNALGKRVKPYLPQICGTVLWRLNNKSAK
VRQQAADLISRTAVVMKTCQEEKLMGHLGVVLYEYLGEEYPEVLGSILGALKAIVNVIG
MHKMTPPIKDLLPRLTPILKNRHEKVQENCIDLVGRIADRGAEYVSAREWMRICFELLEL
LKAHKKAIRRATVNTFGYIAKAIGPHDVLATLLNNLKVQERQNRVCTTVAIAIVAETCSP
FTVLPALMNEYRVPELNVQNGVLKSLSFLFEYIGEMGKDYIYAVTPLLEDALMDRDLVH
RQTASAVVQHMSLGVYGFGCEDSLNHLLNYVWPNVFETSPHVIQAVMGALEGLRVAIG
PCRMLQYCLQGLFHPARKVRDVYWKIYNSIYIGSQDALIAHYPRIYNDDKNTYIRYELD
YIL (SEQ ID NO: 87).

[00215] The SF3B1 gene encodes subunit 1 of the splicing factor 3b protein complex which, together with splicing factor 3a and a 12S RNA unit, forms the U2 small nuclear ribonucleoproteins complex (U2 snRNP). The splicing factor 3b/3a complex binds pre-mRNA upstream of the intron branch site in a sequence-independent manner and can anchor the U2 snRNP to the pre-mRNA. Splicing factor 3b is also a constituent of the minor U12-type spliceosome. The carboxy-terminal two-thirds of subunit 1 of the splicing factor 3b protein complex contains non-identical, tandem HEAT (huntingtin, elongation factor 3, protein phosphatase 2A, targets of

rapamycin 1) repeat domains that form rod-like, helical structures. Most mutations in SF3B1 are localized to its highly conserved C-terminal domain, which comprises 22 Huntington Elongation Factor 3 PR65/A TOR (HEAT) repeats (typically in the region corresponding to residues 622-781).

[00216] In some embodiments, a cancer-associated mutation in SF3B1 disclosed herein is a missense mutation within the major hotspots targeting the 5-9 HEAT repeat domains of the SF3B1 protein.

[00217] Non-limiting examples of SF3B1 mutations include mutations at amino acid positions E622, Y623, R625, N626, H662, T663, K666, K700, V701, K741, G742, D984, G740, and/or D894 of SEQ ID NO: 87. Additional non-limiting examples of SF3B1 mutations include mutations at amino acid positions K700, E622, R625, H662, and/or K666 of SEQ ID NO: 87. In various embodiments, mutations of SF3B1 can be found at amino acid positions R625, K700 (e.g., K700E), and/or K666 of SEQ ID NO: 87.

[00218] RNA splicing includes the removal of intronic sequences from pre-mRNA and the ligation of exons to generate mature mRNA. RNA splicing is carried out by the splicing machinery (spliceosome), which comprises five snRNPs, among other proteins. Introns contain a consensus sequence that includes the 5′ donor splice site (5′ss), branchpoint (BP) and 3′ acceptor splice site (3′ss) recognized by the U1 snRNP, SF1 protein and U2AF, respectively. U2AF is a heterodimer comprising U2AF2 (also called U2AF65) and U2AF1 (also called U2AF35), which recognize the poly-pyrimidine tract and the well-conserved AG dinucleotide sequence of 3′ss, respectively. Following binding to the 3′ss, U2AF participates in replacement of SF1 by U2 snRNP at the BP. Interaction between U1 and U2 snRNPs then triggers transesterification joining the 5′-end of the intron to the BP, most generally an adenosine located in a loosely defined consensus sequence approximately 25 nucleotides upstream of the 3′ss. The 5′ss and 3′ss are subsequently ligated together and the branched intron is excised.

[00219] In some embodiments, splicing abnormalities observed in SF3B1 hotspot mutation cells can include usage of an aberrant (or cryptic) 3′ss (also called AG′), typically located 10 to 30 bases upstream of normal 3′ss. Mutant SF3B1 may preferentially recognize alternative BPs upstream of the canonical sites; and the alternative 3′ss used in an SF3B1 hotspot mutations context may be less dependent on 25 U2AF.

[00220] Amino acid sequences derived from SF3B1 mutation-associated tumors 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. An SF3B1 mutation-associated tumor refers to a tumor associated with one or more mutations of SF3B1. The tumor and the one or more SF3B1 mutations (either in the tumor cells or other cells) can both be present in a subject. Non-limiting examples of SF3B1 mutation-associated tumors include adenoid cystic carcinoma, bladder urothelial carcinoma, breast cancer, chronic B-cell leukemia, chronic lymphocytic leukemia, chronic myelomonocytic leukemia or platelet derived growth factor receptor beta (PDGFRB)-associated chronic eosinophilic leukemia, hematological malignancies such as acute myeloid leukemia, liver cancer (e.g., hepatocellular carcinoma), lung cancer, melanoma, mesothelioma, mucosal melanoma, myelodysplastic myeloproliferative cancer, myeloid leukemia, myeloproliferative neoplasm, progesterone negative breast cancer, orbital melanoma, pancreatic adenocarcinoma, pulmonary adenocarcinoma, renal cell carcinoma, skin or cutaneous melanoma, and uveal melanoma. In some embodiments, the SF3B1 mutation-associated tumor can be uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or endometrial cancer.

[00221] In some embodiments, the amino acids that may be used in the present disclosure 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, 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.

[00222] 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: 1-37, 114-136, 160-164, and 169, 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: 1-37, 114-136, 160-164, and 169. In some embodiments,

an isolated peptide of the present disclosure consists essentially of an amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169. In some embodiments, the isolated peptide consists of an amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169. In some embodiments, the isolated peptide comprises two or more sequences selected from any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169, and a pharmaceutically acceptable salt thereof, and a fragment or derivative thereof.

[00223] A list of non-limiting examples of SF3B1 mutation-associated neoantigenic peptides is provided in **Table 1** below. Bold text denotes peptide sequences (neoantigenic peptides) caused by erroneous splicing associated with SF3B1 mutation. Specifically, bold text indicates insertion of the bolded amino acids. Single underlined text directly followed by italicized text indicates deletion of amino acids from between the single underlined and italicized portions of the sequence. Amino acid acids deleted from such sequences are displayed as double-underlined text within the corresponding wild-type (WT) sequence. Non-bold text denotes a wild-type sequence. "N/A" (i.e., not applicable) means a corresponding cognate WT sequence is not known to exist.

Table 1. Examples of SF3B1 Mutation-associated Neoantigenic Peptides

Gene	Peptide Sequence	SEQ ID NO:	Cognate WT Sequence	SEQ ID NO:
ARIH1	SNWIAANTK**FLFIGMSQMPCHN**	1	SNWIAANTK	91
ARIH1	**FLFIGMSQM**	2	SNWIAANTK	91
CCT2	HSEGNTTAGL**VNLFFFA**DMREGTIGDM	3	HSEGNTTAGLDMREGTIGDM	92
CCT2	TTAGL**VNL**	4	HSEGNTTAGLDMREGTIGDM	92
CCT2	GL**VNLFFFA**	5	HSEGNTTAGLDMREGTIGDM	92
CD34	RSWSPTGER**DAASLPKAPIEEASSHHSIIHQVL**	6	RSWSPTGER	93
CD34	**HHSIIHQVL**	7	RSWSPTGER	93
COG1	YQVWSSAVVK*SLLLDDAGSV*	8	YQVWSSAVVKVLIHGFTQSLLLDDAGSV	94
DLST	TEGDVRWEKA**FDCLFSCWRR**	9	TEGDVRWEKA	95
DPH5	KEQSLENLI**NFLPIGEGRSMNLHGI**	10	KEQSLENL	96
DVL2	PQAYFTLPR**T**YPGSSSMSTI	11	PQAYFTLPR	97
ERGIC3	PTVYMKVDGE**APLPPQ**VLRTNQFSVT	12	PTVYMKVDGEVLRTNQFSVT	98

HIGD2A	KTRENPVVPI**VALSPPQVAWPRR PPSPTASTPSTGATASALSS**	13	KTRENPVVPI	99
IL17RC	AALVQFGQSV**PSSPLSSQ**GSVVYD CFEA	14	AALVQFGQSVGS VVYDCFEA	100
MCM3A P	MKTSVTTSPQ**FLFYSNRVT**	15	MKTSVTTSPQ	101
NET1	HGELRSKSGH**RLFPISPETL**	16	HGELRSKSGH	102
NET1	**LFPISPETL**	17	HGELRSKSGH	102
NONO	HEHQVMLMRQ**VCCFFRFDEAPR RTSEDGRAAQPRGAKTKATGAQ VTFLEHFFPNNSKR**	18	HEHQVMLMRQ	103
NONO	MLMRQ**VCCF**	19	HEHQVMLMRQ	103
OXA1L	GLPWWGAIAA**FCIFQVQSLPAA**	20	GLPWWGAIAA	104
OXA1L	**FQVQSLPAA**	21	GLPWWGAIAA	104
PLXNB1	DAQEYEVSSS**LPCAYS**LVCITGAS GE	22	DAQEYEVSSLVC ITGASGE	105
SF3A2	AEPYETIAFK**EPSLSPALPGAEQR DRQGGGQVLDTLEPGDQAGEW LAPSLAPSHPALAALALP**	23	AEPYETIAFK	106
ENO1	QLADLYKSF**IKDYPDPFD**QDDWG AWQKF	24	QLADLYKSFIKD YPVVSIEDPFDQ DDWGAWQKF	107
STIM1	DDVDHKILTA**NSLSSVF**RQALSEV TAA	25	DDVDHKILTARQ ALSEVTAA	108
STK24	SEKKIHRDIK**GPPWLSFA**AANVLL SEHG	26	SEKKIHRDIKAA NVLLSEHG	109
SYVN1	YLTKSSPSMA**PLPTPRSCTSRPLS LSSFWAR**	27	YLTKSSPSMA	110
VPS33B	EIGHIFLLDR**GISVS**DVDFVTALCS	28	EIGHIFLLDRDVD FVTALCS	111
VPS33B	FLLDR**GISV**	29	EIGHIFLLDRDVD FVTALCS	111
VPS51	EDTTAIDVQ**YLFLPPGGAPVRRG CSQGPEQRLQQEDFLRVQQLSA AGPLRPQLYPQVWLVRGAPGGR DKRGRGCRAGSDSLSSPIQCPDG HQPLEQYPEAIL**	30	EDTTAIDVQ	112
VPS51	**FLRVQQLSA**	31	EDTTAIDVQ	112
ZDHHC1 6	SLVYLWFLCS**LRPSFLS**SVALALG AL	32	SLVYLWFLCSSV ALALGAL	113
USP39	**VLSPSHTSM**	33	N/A	-
USP39	CWLTLRNCVL**SPSHTSM**	34	N/A	-
BRD9	VDISMLSSLV**SFRACHLSRDAVLK SDRNLVSYYHGVCFVNLQISVLK LMLVSCA**	35	N/A	-

ERFE	KRSRGKAKKLK**VPFQ**FGLPGPPGPP	36	N/A	-
UBA1	DGVANALDNVDA**PLALCSPQACTWTAAVSTTGSHCWSQAHWAPKAMCRW**	37	N/A	-
BACE1	MTVGSPPQTL**VFQQTCRGHRGSVPLAPAAPQDPRSCLWSWW**	114	MTVGSPPQTLNILVDTGSS	137
BCAM	HDHYMLEWFL**VTMHDTRGRSPPYQLDSQGRLVLAEAQVGDERDYVCVVRAGAAGTAEATARLNVFGKCPRASPEGRQAGRGAGQGLLT**	115	HDHYMLEWFLTDRSGARPRLASA	138
DGCR2	EGRWEVAFK**APPFVSPT**GSSEVFLPPDPIF	116	EGRWEVAFKGSSEVFLPPDPIF	139
GYPC	MWSTRSPNSTAWPLSL**RIPFSMTSSPSLIRGWPLPP**EPDPGMASASTTMHTT	117	MWSTRSPNSTAWPLSLEPDPGMASASTTMHTT	140
BSG	SESVPPVTDWAWYKITDSEDK*GRSELHIENLNMEADPGQYRCNGTSSKGSD*	118	SESVPPVTDWAWYKITDSEDKALMNGSESRFFVSSSQGRSELHIENLNMEADPGQYRCNGTSSKGSDQAI	141
CD81	GFVNKDQ*ALQQAVVDDDANN*	119	GFVNKDQIAKDVKQFYDQALQQAVVDDDANN	142
GPNMB	**SLAFKDGKQCPDLRWLLGHICHCDLPLGVQIVFR**KHKEYNPIENSPGNVVRSKGLSVFLN	120	DPASPLRMANSALISVGCLAIFVTVISLLVYKKHKEYNPIENSPGNVVRSKGLSVFLNRAKAVFFPGNQEKDPLLKNQEFKGVS	143
MEGF8	PRSPEECRRLRTCSECLARHPRTLQPGDGE**PVFPHQ**ASTPRCKWCTNCPEGACIGRNGSC	121	MPRSPEECRRLRTCSECLARHPRTLQPGDGEASTPRCKWCTNCPEGACIGRNGSCTSENDCRINQREVFWAGNCSEAACGAADC	144
NCSTN	YALANLTGTVVNLTREQCQDPSKVPSENKD*GPLHSNETDRLPRCVRSTARLARALSPAFE*	122	YALANLTGTVVNLTREQCQDPSKVPSENKDLYEYSWVQGPLHSNETDRLPRCVRSTARLARALSPAFE	145

PLXNB3	GRAFADVQYAVSVASRPCNPEPSLYRTSAR**PRLS**RIVCVTSPAPNGTTGPVRVAIKSQPP	123	GRAFADVQYAVSVASRPCNPEPSLYRTSARIVCVTSPAPNGTTGPVRVAIKSQPP	146
APP	YTSIHHGVVE**LSSCFSG**	124	YTSIHHGVVEVDAAVTPEERHLSKMQQNGYENPTYKFFEQMQN	147
CLSTN1	KEPFTISVWMRHGPFGRKKETILCSSDKT**ETAFA**DMNRHHYSLYVHGCRLIFLFRQDPS	125	KEPFTISVWMRHGPFGRKKETILCSSDKTDMNRHHYSLYVHGCRLIFLFRQDPS	148
SDC3	RPVDLEGSGDDDSFPDDELDDLYSGSGSG**SLSLSPTPSD**FEQESGIETAMRFSPDVALA	126	RPVDLEGSGDDDSFPDDELDDLYSGSGSGYFEQESGIETAMRFSPDVALAVSTTPAVLPTTNIQPVGTPFEELPSERPTLEPAT	149
PTK7	VAPEDSGRYTCIAGNSCNIKHTEAPLYVVD**PSVCLS**DKPVPEESEGPGSPPPYKMIQTI	127	VAPEDSGRYTCIAGNSCNIKHTEAPLYVVDKPVPEESEGPGSPPPYKMIQTIGLSVGAAVAYII	150
FXYD5	IVGLILPTRG**SLWLQVGTDCGGGDRSREPKKEAAGMSRLEGTEAGPGWGQWSYPALTTW**	128	IVGLILPTRGQTLKDTTSSSSADSTIMDIQVPTRAPDAVYTELQPTSPTPTWPADETPQ	151
NRG3	KSHVQLQ**NKMSGF**	129	KSHVQLQNYSKVERHPVTALEKMMESSFVGP	152
CA14	SHLHEVRHK**VCSPQIRRPQCLPST**	130	SHLHEVRHKDQKTSVPPFN	153
MMP14	KDGKFVFFK**DVALLSSP**GDKHWVFDEA	131	KDGKFVFFKGDKHWVFDEA	154
SLC2A11	NLSIINAPTL**VHEHIEVLGRWCSWRGHGSL**	132	NLSIINAPTLHIQEFTNETW	155
GPR143	ILFQKTVTAVPS**VLHSGLFT**	133	ILFQKTVTAVASLLKGRQG	156
TNFRSF14	CLQCQMCDP**APPLGLQPWACARAGTAPGQRTPCVAAAQATSASSRTGTTAPRAALTPPPARARGCRRE**	134	CLQCQMCDPAMGLRASRNCSRTENAVCGCSPGHFC	157

	VSGGWRTPLPFPPWSPVPRCLEP QVSSTAWPAQGPW		IVQDGDHCAAC RA	
NRG3	LVKSHVQLQN**KMSGF**	135	LVKSHVQLQNY SKVERHPVT	158
HTR2B	**HAQQNNNETR**EAMWPLPLVL	136	PIALLTIMFEAM WPLPLVL	159
BSG	TDWAWYKITD**SEDKGRSE**LHIEN LNMEA	160	TDWAWYKITDS EDKALMNGSES RFFVSSSQGRSEL HIENLNMEA	165
TFRC	YTLIEKTMQN**HSGWSAMVRSWL TAASTSWVQAILLPQPPEELGLQ** VKHPVTGQFLY	161	YTLIEKTMQNVK HPVTGQFLY	166
TFRC	YTLIEKTMQN**MESSSVFLQHSGW SAMVRSWLTAASTSWVQAILLP QPPEELGLQ**VKHPVTGQFL	162	YTLIEKTMQNVK HPVTGQFLY	166
TMEM1 79B	VAAAAMTRTQ**LLSTGLLQW**	163	VAAAAMTRTQG SFSGRCPLYGVA TLNGSSLALSRPS APSLCYF	167
GPX1	GFPCNQFGHQ**LPFSPVGERQERR DSEFPQVRPAWWWVRAQLHAL REVRGERCGGAPSLRLPAGGPA SSQRRRHRAYDRPQAHHLVSGV SQRCCLELUEVPGGPURCAPTQ VQPPLPDHUHRAUHRSPAVSRA QLCLGRPSYPGCLAVAVLLSRG GFHLUGCFL**	164	GFPCNQFGHQEN AKNEEILNSLKY VRPGGGFEPNFM LFEKCEVNGAG AHPLFAFLREAL PAPSDDATALMT DPKLITWSPVCR NDVAWNFEKFL VGPDGVPLRRYS RRFQTIDIEPDIE ALLSQGPSCA	168
SLC3A2	FQGHGAGNLA**APIS**GLKGRLDYLS	169	FQGHGAGNLAG LKGRLDYLS	170

[00224] Antigenic fragments of the peptides disclosed herein (e.g., SEQ ID NOs: 1-37, 114-136, 160-164, or 169) may be identified based on informatics analysis and mass spectrometry.

[00225] 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.

[00226] 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.

[00227] 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 of how peptides bind to various MHC molecules, it is possible to identify the major MHC anchor amino acids, as well as the surface-exposed amino acids, that vary among different peptides.

[00228] In some embodiments, the length of an MHC-binding peptide disclosed herein can be from about 5 to about 40 amino acid residues, from about 6 to about 30 amino acid residues, from about 8 to about 20 amino acid residues, or from about 9 amino acid residues to about 11 amino acid residues. In some embodiments, an MHC-binding peptide disclosed herein can include peptides of any length, e.g., between 5 and 200 amino acids, in whole integer increments (e.g., 5, 6, 7, 8, 9 . . . 200 amino acids). While naturally MHC class II-bound peptides can vary from about 9-40 amino acids in length, in nearly all cases, the peptide can be truncated to an about 9-11 amino-long acid core, without loss of MHC binding activity or T cell recognition.

[00229] 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.

[00230] In some embodiments, the length of an isolated peptide described herein is about 8 to about 200 amino acids in length. In some embodiments, the isolated peptide described herein is about 8 to about 163 amino acids in length. In some embodiments, the isolated peptide described herein is about 8 to about 104 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 102 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 88 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 64 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 63 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 60 amino acids in

length. In some embodiments, the isolated peptide is about 8 to about 59 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 58 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 56 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 54 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 52 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 51 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 49 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 43 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 41 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 33 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 31 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 30 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 28 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 27 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 26 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 25 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 24 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 22 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 20 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 19 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 17 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 15 amino acids in length. In some embodiments, the isolated peptide is about 8 to about 13 amino acids in length. In some embodiments, the isolated peptide is about 8 amino acids in length. In some embodiments, the isolated peptide is about 9 amino acids in length.

[00231] In some embodiments, the isolated peptides described herein may be about 5-200 amino acids in length. For example, a peptide disclosed herein may be 5 amino acids, 6 amino acids, 7 amino acids, 8 amino acids, 9 amino acids, 10 amino acids, 11 amino acids, 12 amino acids, 13 amino acids, 14 amino acids, 15 amino acids, 16 amino acids, 17 amino acids, 18 amino acids, 19 amino acids, 20 amino acids, 21 amino acids, 22 amino acids, 23 amino acids, 24 amino acids, 25 amino acids, 26 amino acids, 27 amino acids, 28 amino acids, 29 amino acids, 30 amino

acids, 31 amino acids, 32 amino acids, 33 amino acids, 34 amino acids, 35 amino acids, 36 amino acids, 37 amino acids, 38 amino acids, 39 amino acids, 40 amino acids, 41 amino acids, 42 amino acids, 43 amino acids, 44 amino acids, 45 amino acids, 46 amino acids, 47 amino acids, 48 amino acids, 49 amino acids, 50 amino acids, 51 amino acids, 52 amino acids, 53 amino acids, 54 amino acids, 55 amino acids, 56 amino acids, 57 amino acids, 58 amino acids, 59 amino acids, 60 amino acids, 61 amino acids, 62 amino acids, 63 amino acids, 64 amino acids, 65 amino acids, 66 amino acids, 67 amino acids, 68 amino acids, 69 amino acids, 70 amino acids, 71 amino acids, 72 amino acids, 73 amino acids, 74 amino acids, 75 amino acids, 76 amino acids, 77 amino acids, 78 amino acids, 79 amino acids, 80 amino acids, 81 amino acids, 82 amino acids, 83 amino acids, 84 amino acids, 85 amino acids, 86 amino acids, 87 amino acids, 88 amino acids, 89 amino acids, 90 amino acids, 91 amino acids, 92 amino acids, 93 amino acids, 94 amino acids, 95 amino acids, 96 amino acids, 97 amino acids, 98 amino acids, 99 amino acids, 100 amino acids, 101 amino acids, 102 amino acids, 103 amino acids, 104 amino acids, 105 amino acids, 106 amino acids, 107 amino acids, 108 amino acids, 109 amino acids, 110 amino acids, 111 amino acids, 112 amino acids, 113 amino acids, 114 amino acids, 115 amino acids, 116 amino acids, 117 amino acids, 118 amino acids, 119 amino acids, 120 amino acids, 121 amino acids, 122 amino acids, 123 amino acids, 124 amino acids, 125 amino acids, 126 amino acids, 127 amino acids, 128 amino acids, 129 amino acids, 130 amino acids, 131 amino acids, 132 amino acids, 133 amino acids, 134 amino acids, 135 amino acids, 136 amino acids, 137 amino acids, 138 amino acids, 139 amino acids, 140 amino acids, 141 amino acids, 142 amino acids, 143 amino acids, 144 amino acids, 145 amino acids, 146 amino acids, 147 amino acids, 148 amino acids, 149 amino acids, 150 amino acids, 151 amino acids, 152 amino acids, 153 amino acids, 154 amino acids, 155 amino acids, 156 amino acids, 157 amino acids, 158 amino acids, 159 amino acids, 160 amino acids, 161 amino acids, 162 amino acids, 163 amino acids, 164 amino acids, 165 amino acids, 166 amino acids, 167 amino acids, 168 amino acids, 169 amino acids, 170 amino acids, 171 amino acids, 172 amino acids, 173 amino acids, 174 amino acids, 175 amino acids, 176 amino acids, 177 amino acids, 178 amino acids, 179 amino acids, 180 amino acids, 181 amino acids, 182 amino acids, 183 amino acids, 184 amino acids, 185 amino acids, 186 amino acids, 187 amino acids, 188 amino acids, 189 amino acids, 190 amino acids, 191 amino acids, 192 amino acids, 193 amino acids, 194 amino acids, 195 amino acids, 196 amino acids, 197 amino acids, 198 amino acids, 199 amino acids, or 200 amino acids in length.

[00232] In some embodiments, the length of an isolated peptide described herein is about 8 to about 163 amino acids in length.

[00233] 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.

[00234] In some embodiments, a peptide disclosed herein 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 bonds 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—.

[00235] 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*.

[00236] 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) can be used 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.

[00237] 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.

[00238] 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.

[00239] 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.

[00240] 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.

[00241] Peptides described herein may comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) 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, 4, 5, 6, 7, 8, 9, 10, or more) amino acids may be substituted and/or inserted and/or deleted from the sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169. Each substitution and/or insertion and/or deletion can take place at any position of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169.

[00242] 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 of SEQ ID NOs: 1-37, 114-136, 160-164, and 169. A peptide of the disclosure may comprise the amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169, except for one or more (e.g., 1, 2, 3, or 4) amino acid substitutions, insertions, and/or deletions.

[00243] 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.

[00244] 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.

[00245] 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.

[00246] 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.

[00247] 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.

[00248] 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.

[00249] The amino acid residues that do not substantially contribute to interactions with the T-cell receptor can be modified by replacement with other amino acids whose incorporation does not substantially affect T cell reactivity and does not eliminate binding to the relevant MHC.

[00250] The peptides may also comprise isosteres of two or more residues. An "isostere," as used herein, 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.

[00251] 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.

[00252] 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 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.

[00253] 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.

[00254] As the splicing factor gene (e.g., SF3B1) mutation-associated neoantigenic peptide is a mutant peptide, T cell lines against a natural (non-mutated) epitope can be 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. Splicing factor gene (e.g., SF3B1) mutation-associated neoantigenic 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 the 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 is used to define basic and useful characteristics of the splicing factor gene mutation-associated neoantigenic peptides:

[00255] 1. Minimal required concentration of splicing factor gene 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 peptide concentration; and

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

[00256] By way of a non-limiting example, splicing factor gene mutation-associated neoantigenic 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.

[00257] 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.

[00258] A peptide of the disclosure, or 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 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).

[00259] 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.

[00260] 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.

[00261] 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 NOs: 1-37, 114-136, 160-164, and 169, and a fragment or derivative thereof, and a pharmaceutically acceptable salt thereof. For example, such a multi-epitope peptide or polypeptide may comprise 2 to 67, 2 to 60, 2 to 55, 2 to 50, 2 to 45, 2 to 40, 2 to 37, 2 to 30, 5 to 25, 5 to 20, or 10 to 15 single-epitope peptides as described herein (e.g., SEQ ID NOs: 1-37, 114-136, 160-164, or 169, or a fragment or derivative thereof). The single-epitope peptides (e.g., SEQ ID NOs: 1-37, 114-136, 160-164, or 169, or a fragment or derivative thereof) may be arranged in any order, and may be identical or different.

[00262] 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, or 67 amino acid sequences selected from SEQ ID NOs: 1-37, 114-136, 160-164, and 169, and a derivative thereof, and a pharmaceutically acceptable salt thereof.

[00263] The single-epitope peptides may be linked via a linker. The linker may comprise of 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., any of those listed in **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, and more commonly three to six residues.

[00264] 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.

[00265] 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.

[00266] 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.

[00267] 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.

[00268] 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.

[00269] 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, the peptides may be provided in a form which is substantially free of other peptides or proteins.

Peptide-MHC (pMHC) Complexes Disclosed Herein

[00270] 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.

[00271] 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 (transmembrane) domain and a CYT (cytoplasmic) domain). The β chain contains two extracellular domains (β1 and β2), and two intracellular domains (a TM domain and CYT domain).

[00272] 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.

[00273] 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.

[00274] 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.

[00275] 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, incorporated herein by reference in its entirety for all purposes. If the peptide does not form a complex with an MHC molecule, then the MHC molecule will not refold. Refolding can be confirmed using an antibody that recognizes the MHC molecule in a folded state only. Alternatively, the ability of a peptide to stabilize the MHC molecule on the surface of transporter associated with antigen processing (TAP)-deficient cell lines (e.g., 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, can be assessed. Other biophysical methods to determine peptide-MHC molecule interaction parameters may also be utilized.

[00276] 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 MHC Binding and Non-binding (MHCBN) peptides database 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.

[00277] 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.

[00278] 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).

[00279] 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 a splicing factor gene mutation-associated tumor and/or a splicing factor gene mutation-induced disease or disorder. In some embodiments, the TCR includes an α-chain and a β-chain.

[00280] 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), and 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).

[00281] 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., B1-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.

[00282] 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.

[00283] 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, and 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.

[00284] 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.

[00285] 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.

[00286] 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.

[00287] 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.

[00288] 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.

[00289] 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.

[00290] 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, which are each herein incorporated by reference in their entirety. 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.

[00291] 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.

[00292] 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.

[00293] 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 200 contiguous amino acids. In

certain embodiments, the sequence can comprise from about 8 to about 195 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 190 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 185 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 180 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 175 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 170 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 165 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 160 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 155 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 150 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 145 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 140 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 135 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 130 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 125 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 120 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 115 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 110 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 105 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 100 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 95 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 90 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 85 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 80 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 75 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 70 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 65 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 60 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 55 contiguous

amino acids. In certain embodiments, the sequence can comprise from about 8 to about 50 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 45 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 40 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 35 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 30 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 25 contiguous amino acids. In certain embodiments, the sequence can comprise from about 8 to about 20 contiguous amino acids. 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. In certain embodiments, the sequence can comprise about 8 contiguous amino acids. In certain embodiments, the sequence can comprise about 9 contiguous amino acids. In certain embodiments, the sequence can comprise about 10 contiguous amino acids. In certain embodiments, the sequence can comprise about 11 contiguous amino acids. In certain embodiments, the sequence can comprise about 12 contiguous amino acids. In certain embodiments, the sequence can comprise about 13 contiguous amino acids. In certain embodiments, the sequence can comprise about 14 contiguous amino acids. In certain embodiments, the sequence can comprise about 15 contiguous amino acids.

[00294] 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.

[00295] 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.

[00296] 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: 89) and (GGGS)n (SEQ ID NO: 90), 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). 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. Exemplary linkers can comprise amino acid sequences including, but not limited to, those listed in **Table 2**.

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	39	GGGGGTGGTTCC	55	GGTGGCGGTAGT	71
GGSG	40	GGTGGGTCTGGG	56	GGGGGATCTGGT	72
GSGGS	41	GGGTCCGGGGGCTCC	57	GGCAGTGGCGGTAGC	73
GGSGG	42	GGTGGGAGCGGTGGT	58	GGAGGGAGTGGAGGG	74
GSGSG	43	GGCAGCGGAAGCGGA	59	GGGTCTGGCTCAGGC	75
GSGGG	44	GGGAGTGGGGGAGGT	60	GGTTCTGGCGGAGGT	76
GGGSG	45	GGTGGGGGAAGTGGA	61	GGTGGTGGGAGTGGA	77
GSSSG	46	GGCAGCTCATCTGGT	62	GGCTCAAGCAGTGGA	78
GCGASGGGGSGGGGS	47	GGATGTGGTGCATCTGGAGGGGGAGGCTCTGGGGGGGGTGGATCT	63	GGCTGTGGGGCTAGTGGGGGAGGTGGTAGTGGTGGTGGCGGTTCC	79
GCGASGGGGSGGGGS	48	GGGTGTGGTGCTAGTGGGGGTGGCGGATCAGGTGGAGGCGGGAGC	64	GGATGTGGGGCCTCAGGTGGGGGTGGCAGCGGTGGTGGAGGGTCA	80
GGGGSGGGGS	49	GGGGGCGGAGGATCTGGGGGAGGGGGATCA	65	GGTGGCGGGGGCTCTGGTGGAGGAGGATCT	81

GGGASGGGGS GGGGS	50	GGGGGGGGCGCT TCAGGCGGAGGT GGAAGTGGTGGA GGAGGT	66	GGAGGCGGCGCTT CTGGGGGCGGGGG TAGTGGGGGTGGA GGT	82
GGGGSGGGGS GGGGS	51	GGAGGGGGAGGT TCTGGCGGCGGG GGATCAGGAGGC GGTGGGAGC	67	GGTGGAGGTGGAA GTGGAGGAGGGG GATCAGGCGGAGG CGGGAGC	83
GGGASGGGGS	52	GGTGGGGGGGCG TCAGGTGGAGGC GGAAGT	68	GGAGGGGGAGCCT CTGGCGGTGGAGG ATCA	84
GGGGSGGGGS GGGGS	53	GGCGGCGGAGGT TCTGGTGGGGGT GGCAGTGGAGGA GGAGGCAGC	69	GGGGGAGGAGGC AGTGGAGGTGGGG GAAGTGGTGGAGG GGGGTCT	85
GGGGSGGGGS GGGGSGGGGS	54	GGAGGTGGAGGT AGTGGCGGTGGT GGGTCAGGGGGA GGCGGGTCCGGT GGCGGTGGGAGT	70	GGGGGTGGAGGAT CAGGAGGCGGTGG TTCTGGGGGAGGT GGATCCGGCGGGG GTGGTAGT	86

[00297] 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 (e.g., 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 (e.g., an addition or substitution mutation) 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.

[00298] 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.

[00299] Attaching the peptide to the MHC class I or MHC class II molecule via a flexible linker 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.

[00300] 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 (see, 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 (see, e.g., GenBank Accession No. AF072097.1).

[00301] 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 herein by reference in their entireties.

[00302] 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.

[00303] 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.

[00304] 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.

[00305] 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.

[00306] 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.

[00307] 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.

[00308] 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, the cells may present the complex at a level higher than they would in nature. Such cells may be obtained by pulsing the cells with one or more peptides (e.g., 2 to 67, 2 to 54, 2 to 37, 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 67, 2 to 54, 2 to 37, 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

[00309] 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.

[00310] In some embodiments, peptides are fused or conjugated to one or more heterologous molecules, which includes an MHC molecule (or fragments thereof).

[00311] 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.

[00312] 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: 171), 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.

[00313] 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 (SEQ ID NO: 171) tag, 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), a

yellow fluorescent protein (YFP), a red fluorescent protein (RFP), a cyan fluorescent protein (CFP), mCherry, tdTomato, a SUMO tag, and a ubiquitin tag.

[00314] In some embodiments, fusion proteins of the disclosure may comprise one or more epitopes that are 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 a PADRE peptide, tetanus toxoid peptide (830-843), and influenza haemagglutinin, HA(307-319).

[00315] 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), and 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.

[00316] 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.

[00317] 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 one or more 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.

[00318] Other suitable heterologous molecules include, but are not limited to, fluorescent labels, 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.

[00319] 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.

[00320] 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 enzyme-linked immunosorbent assay (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.

[00321] Peptides or pMHC complexes of the disclosure may be in multimeric form, for example, dimeric, tetrameric, pentameric, 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.

[00322] 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).

[00323] 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.

[00324] 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.

[00325] 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 dimethyl sulfoxide (DMSO). The formation of cysteine bonds can be controlled by the selective use of thiol-protecting groups during peptide synthesis.

[00326] 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 can connect 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., a CH3 domain), or a sequence that is substantially identical to the CH3. The hinge region may be Ig-derived and contribute 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 can 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 constructs are described in US 10,590,195, which is incorporated herein by reference in its entirety.

Nucleic Acids and Vectors

[00327] 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, oligomeric complexes), fusion proteins, or conjugates, comprising the peptides of the disclosure). The polynucleotide may be, for example, DNA, cDNA, PNA, or 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 the polynucleotide may or may not contain introns, as long as the polynucleotide codes for the peptide.

[00328] 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: 1-37, 114-136, 160-164, and 169, and 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: 1-37, 114-136, 160-164, and 169, and a fragment or derivative thereof.

[00329] In some embodiments, the polynucleotide described herein encodes more than one peptide selected from any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169, and fragments thereof. For example, the polynucleotide described herein may encode 2 to 67, 2 to 54, 2 to 37, 2 to 30, 5 to 25, 5 to 20, or 10 to 15 peptides as described herein (e.g., SEQ ID NOs: 1-37, 114-136, 160-164, or 169, or fragments thereof). The peptides may be arranged in any order, and may be identical or different.

[00330] 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, or 67 amino acid sequences selected from SEQ ID NOs: 1-37, 114-136, 160-164, and 169, and fragments thereof.

[00331] In some embodiments, the polynucleotide described herein is a DNA molecule.

[00332] 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.

[00333] In some embodiments, the polynucleotide described herein is an RNA molecule. For example, the RNA molecule may be mRNA or a self-replicating RNA.

[00334] A polynucleotide encoding RNA disclosed herein can be used to make a vaccine. RNA cannot integrate into the genome and, therefore, has 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.

[00335] 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.

[00336] 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.

[00337] 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.

[00338] 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.

[00339] 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.

[00340] 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, also known as 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.

[00341] 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.

[00342] In a further aspect, the disclosure provides a vector comprising a nucleic acid sequence described herein. 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.

[00343] 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, or the like.

[00344] 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.

[00345] The following vectors are provided by way of example: bacteriophages such as lambda (X) bacteriophage and 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.

[00346] 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 AAV 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 or bacterial spores.

[00347] As further examples, adenovirus vectors may be derived from human adenovirus (Ad), but also from adenoviruses that infect other species, such as a bovine adenovirus (e.g., bovine adenovirus 3, BAdV3), a canine adenovirus (e.g., CAdV2), a porcine adenovirus (e.g,. PAdV3 or 5), or adenoviruses that infect 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).

[00348] 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.

[00349] 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 (HEK) cells. Examples of insect cells include Sf9 cells, which can be transfected with baculovirus expression vectors.

[00350] Transformation of appropriate cell hosts with a DNA construct of the present disclosure can be accomplished by well-known methods that typically depend on the type of vector used. Successfully transformed cells, e.g., 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.

[00351] 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.

[00352] 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

[00353] Peptides, pMHC complexes, or other peptide-based molecules (such as oligomeric complexes, fusion proteins, or conjugates 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, a 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.

[00354] 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.

[00355] 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, and 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.

[00356] The candidate peptide binding moiety may be a peptide binding moiety of the type already described herein, such as an antibody or a TCR.

[00357] 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.

[00358] 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 the library may not correspond to those contained in a natural repertoire, for example, the TCRs may contain α- and β-chain pairing that would not be present *in vivo*, may contain non-natural mutations, and/or 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 magnetic bead or column matrix, and phage bound pMHC complexes can be 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.

[00359] Alternatively, antigen-binding T cells and TCRs can be isolated from fresh blood obtained from patients or healthy donors. Such method can involve stimulating T cells using autologous dendritic cells (DCs), followed by autologous B cells, and then pulsing 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 IFN-γ ELISpot assay). Activated cells may then be sorted by fluorescence-activated cell sorting (FACS) using labelled antibodies to detect intracellular cytokine production (e.g., IFN-γ), 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.

[00360] A peptide binding moiety disclosed herein can include, for example, without limitation, an antibody, a TCR, or a CAR.

[00361] 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".

[00362] 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 (e.g., cell or tissue that show higher expression of the gene(s) expressing neoantigenic peptides), or to another desired antigen associated with the same or similar disease or disorder.

[00363] 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.

[00364] 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.

[00365] 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 (see, e.g., WO94/13804, incorporated herein by reference in its entirety for all purposes). 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).

[00366] 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.

[00367] 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. Alternative scaffolds 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*.

[00368] 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.

[00369] 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)).

[00370] In some embodiments, the peptide binding moiety may be a TCR. TCRs can be described using the International Immunogenetics (IMGT) TCR nomenclature, and the IMGT public database of TCR sequences.

[00371] The TCRs of the present disclosure may be in any format. For example, the TCRs may be αβ heterodimers, or αα or ββ homodimers.

[00372] αβ 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.

[00373] 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 T cell receptor alpha constant (TRAC)/ T cell receptor beta constant (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.

[00374] 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.

[00375] 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.

[00376] TCRs of the disclosure may be engineered to include mutations. Methods for producing mutated, high affinity TCR variants include 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. In some cases, there may be between 1 and 15 mutations in the α- and or β- chain variable regions.

[00377] 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.

[00378] 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*.

[00379] 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.

[00380] In a further aspect, the disclosure provides nucleic acid encoding the TCR of the disclosure, a TCR expression vector comprising a 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 (ORF) 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.

[00381] 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., IL-2, IL-15), 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., IL-15 and/or IL-21; 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.

[00382] 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.

[00383] 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, can 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.

[00384] 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.

[00385] 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.

[00386] 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, FcγRIIIa receptor, or a functional fragment thereof.

[00387] 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, IL-2R gamma, IL-7Ra, 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 fragments thereof.

[00388] 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), and FcR β.

[00389] 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.

[00390] 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.

[00391] In another aspect, the disclosure further provides host 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 a T cell, an NK cell, or a macrophage. The host cell may be autologous or allogeneic with respective to the subject receiving the cell (as treatment).

[00392] 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.

[00393] 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 the desired cell or tissue or to another known cancer antigen.

Pharmaceutical Compositions, Dosage Forms, and Administration

[00394] 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.

[00395] 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, and the like.

[00396] Compositions of the disclosure may comprise multiple peptides, e.g., 2 to 67, 2 to 54, 2 to 37, 5 to 25, 5 to 20, or 10 to 15 peptides as described herein (e.g., SEQ ID NOs: 1-37, 114-136, 160-164, or 169, or a fragment or derivative thereof). 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, or 67 amino acid sequences selected from SEQ ID NOs: 1-37, 114-136, 160-164, and 169, and a fragment or derivative thereof, and a pharmaceutically acceptable salt thereof.

[00397] 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.

[00398] 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.

[00399] 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.

[00400] 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.

[00401] 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.

[00402] 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.

[00403] 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.

[00404] 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.

[00405] 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 one or more peptides of SEQ ID NOs: 1- 37, 114-136, 160-164, and 169, and fragments or variants thereof, may be developed to show a desired shelf stability at, e.g., −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.

[00406] 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.

[00407] 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.

[00408] 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 then slowly released from the microspheres and absorbed by the surrounding cells present in the selected tissue.

[00409] 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.

[00410] 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.

[00411] 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.

[00412] 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.

[00413] 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.

[00414] 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.

[00415] 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.

[00416] 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 condition in the subject (e.g., cancer).

[00417] 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.

[00418] 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.

[00419] 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.

[00420] 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.

[00421] 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 tubes, flasks, bottles, syringes, or other containers. 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.

[00422] 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.

[00423] 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.

[00424] Certain additional agents used in the combination therapies can be formulated and administered by any means known in the art.

[00425] 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.

[00426] 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 or prevention of a splicing factor gene mutation-associated tumor and/or a splicing factor gene mutation-induced disease or disorder. 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.

[00427] 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).

[00428] The pharmaceutical compositions of the disclosure may be administered directly into the patient, into the affected organ or systemically intradermal (i.d.), intramuscular (i.m.), subcutaneous (s.c.), intraperitoneal (i.p.), and intravenous (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 IL-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, and/or virus-like particles (VLPs). The peptide or peptide-based molecule may also be conjugated to a suitable carrier such as keyhole limpet hemocyanin (KLH) or mannan (see, e.g., WO 95/18145, incorporated herein by reference in its entirety, and Longenecker et al., 1993).

[00429] 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.

[00430] 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.

[00431] 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.

[00432] 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.

[00433] 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 ectodomains of multi-span integral membrane proteins, or antibodies directed to any of these molecules.

[00434] 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.

[00435] 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.

[00436] 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 example, 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 of

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).

[00437] Other molecules suitable for complexing with the polypeptide or polynucleotides of the disclosure include cationic molecules, such as polyamidoamine, dendritic polylysine, polyethylene irinine, polypropylene imine, polylysine, chitosan, DNA-gelatin coarcervates, DEAE dextran, dendrimers, and polyethylenimine (PEI).

[00438] In some embodiments, compositions of the present disclosure can be conjugated to nanoparticles. Nanoparticles that may be used for conjugation with compositions 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 a composition contained in a pharmaceutical composition for targeting cells (e.g., tumor cells or other cells).

Treatment Methods

[00439] 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 an SF3B1 mutation-associated diseases or disorders (e.g., an SF3B1 mutation-associated tumor). Such compositions may additionally, or alternatively, may be used to reduce the likelihood of a condition caused by an SF3B1 mutation.

[00440] 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.

[00441] 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.

[00442] In some embodiments, the contacting is *ex vivo*. In some embodiments, the contacting is *in vivo* in a subject (e.g., a human).

[00443] In some embodiments, the cell is a mammalian cell (e.g., a human cell).

[00444] 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.

[00445] The interaction of a T cell with a peptide 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., IL-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 the TCRs on the T cell. "T cell death" refers to the permanent cessation of substantially all functions of the T cell.

[00446] In another aspect, provided herein is a method of inducing an immune response against a SF3B1 mutation-associated tumor 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 complexes), fusion proteins, conjugates, nucleic acid molecules, vectors, cells, or binding moieties) of the present disclosure.

[00447] 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.

[00448] 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.

[00449] 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, ELISA, 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.

[00450] 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).

[00451] 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.

[00452] In a related aspect, disclosed herein is a method of treating an SF3B1 mutation-associated tumor 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 complexes), fusion proteins, conjugates, nucleic acid molecules, vectors, cells, binding moieties) of the present disclosure.

[00453] In a related aspect, disclosed herein is a method of preventing or reducing the likelihood of an SF3B1 mutation-associated tumor 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 complexes), fusion proteins, conjugates, nucleic acid molecules, vectors, cells, binding moieties) of the present disclosure.

[00454] The SF3B1 mutation-associated disease or disorder can be an SF3B1 mutation-associated tumor, e.g., a uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or endometrial cancer .

[00455] In some embodiments, an SF3B1 mutation-associated tumor can include chronic acral lentiginous melanoma (ALM), acute myeloid leukemia (AML), breast cancer (BRCA), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), esophageal carcinoma (ESCA), myeloproliferative neoplasms (MPN), myelodysplastic syndrome (MDS), mesothelioma (MESO), pancreatic ductal adenocarcinoma (PDAC), primary malignant melanoma (PMM), skin cutaneous melanoma (SKCM), uveal melanoma (UVM), and also pre-tumor syndromes, clonal

hematopoiesis of indeterminate potential (CHIP) and clonal cytopenia of undetermined significance (CCUS).

[00456] When the SF3B1 mutation-associated disease or disorder 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 with 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.

[00457] 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.

[00458] In some embodiments, the compositions and methods disclosed herein are useful to enhance the efficacy of vaccines directed to SF3B1 mutation-associated diseases or disorders (e.g., an SF3B1 mutation-associated tumor such as uveal melanoma). 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 uveal melanoma) is administered to the subject.

[00459] The compositions and methods described herein can be also administered in combination with an anti-tumor antibody (e.g., an antibody targeting a tumor such as uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or endometrial cancer), or an antibody directed at a pathogenic antigen or allergen.

[00460] 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 also be 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.

[00461] Therapeutic methods described herein can be combined with additional immunotherapies and therapies. For example, when used for treating cancer, lymphocytes (e.g., NK cells and T 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 compositions 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 compositions 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).

[00462] 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.

[00463] 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.

[00464] 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

[00465] 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.

[00466] 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.

[00467] 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.

[00468] 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.

[00469] 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).

[00470] 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.

[00471] 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.

[00472] 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.

[00473] 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 Identification of Target Peptide with Splice Junction Analysis

[00474] Targeted therapies are being developed that target surface molecules on malignant cells to selectively destroy the malignant cells while sparing healthy tissue. Identifying viable target surface molecules is an ongoing endeavor. Some of the research is aimed at developing targeted drugs that target proteins and fragments thereof (peptides) produced by specific gene mutations in cancer cells.

[00475] The present disclosure further comprises a method and system for identification of target peptides for targeted disease therapy. Some or all the steps of the method can be executed by a computational device or system. A suitable computational device and system are illustrated in **Figs. 4** and **5** as non-limiting examples. 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.

[00476] **Fig. 1** is a flow diagram of a method 100 for identifying a target peptide for cancer treatment as a non-limiting example.

[00477] At step 102, a first database comprising a plurality of first gene samples is queried to identify a first gene associated with the plurality of first gene samples. The first gene can have a prevalence of mutation beyond a predetermined threshold. The mutation can correspond to a first cancer type. In some embodiments, the first database can include a Genome Reference Consortium (GRC) Human Build or other similar database as understood by a person skilled in the art. In some embodiments, the first gene can include SF3B1, or another splicing factor gene as understood by a person skilled in the art. In some embodiments, the first cancer type can include uveal melanoma, breast cancer, ovarian cancer, prostate cancer, skin cutaneous melanoma, acute myeloid leukemia, myelodysplastic syndrome, or chronic lymphocytic leukemia.

[00478] At step 104, the plurality of first gene samples can be filtered by gene samples that correspond to an exon splicing junction mutation. Gene samples having an exon splicing junction mutation can be retained. In some embodiments, a tumor sample of the first cancer type that is not associated with the exon splicing junction mutation can be compared to a tumor sample of the first cancer type that is associated with the exon splicing junction mutation. In some embodiments, one or more gene samples of the plurality of first gene samples that are associated with copy number variations can be removed by filtering. In some embodiments, the plurality of first gene samples can be filtered by identifying gene samples that correspond to a surface protein and

retaining those identified gene samples. In some embodiments, the surface protein can include one or more peptides that form an MHC-peptide complex with an MHC molecule. In some embodiments, the plurality of first gene samples can be filtered to retain genes having exon splicing junction mutations that are not expressed in non-cancerous tissue.

[00479] At step 106, a second gene can be identified that corresponds to the filtered plurality of first gene samples. The second gene can be associated with the first cancer type. In some embodiments, gene samples of the filtered plurality of first gene samples can be categorized as being annotated or non-annotated based on a comparison of the gene samples to an annotated gene model. In some embodiments the second gene can be categorized as non-annotated.

[00480] At step 108, a peptide sequence can be generated based on the filtered plurality of first gene samples such that the generated peptide sequence is associated with a sequence of the target peptide.

[00481] At step 110, the generated peptide sequence can be verified as being associated with the second gene. In some embodiments, mass spectrometry can be performed on an MHC-peptide complex associated with the cancer to confirm a presence of the generated peptide sequence.

[00482] **Fig. 2** is a flow diagram of another method 200 for identifying a target peptide for disease treatment as a non-limiting example.

[00483] At step 202, a splicing factor gene can be identified that has a significant prevalence of mutation in the disease in a database. The database can include genetic data of samples of the disease. In some embodiments, the splicing gene factor includes SF3B1, or other splicing factor genes as understood by a person skilled in the art.

[00484] At step 204, differently expressed splicing junctions of the splicing factor gene can be identified based on a comparison of splicing junctions of genetic data of samples that have a mutation in the splicing factor gene compared to splicing junctions of genetic data of samples that lack a mutation in the splicing factor gene.

[00485] At optional step 206, a disease-specific splicing junction can be identified based on one or more MHC-peptide complexes associated with the disease, such that the disease-specific splicing junction has a minimal expression in non-diseased samples.

[00486] At optional step 208, a disease-specific mutated splicing junction can be identified as being both a disease-specific splicing junction and a differently expressed splicing junction. In

some embodiments, the disease-specific splicing junction comprises SLC3A2-APIS splice variant junction.

[00487] At step 210, the target peptide can be determined based at least in part on the differently expressed splicing junctions identified at step 204, and optionally based at least in part on the disease-specific mutated splicing junction identified at step 208.

[00488] In some embodiments, the target peptide is derived from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, BSG, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

Method and System for Prioritizing Target Peptides for Targeted Therapy Development

[00489] The present disclosure further comprises a method and system for prioritizing candidate target peptides for further development to create a targeted disease therapy, method 2600, as illustrated in **Fig. 26A**. **Figs. 26B-26C** are flow diagrams showing another embodiment of the method 2640 for prioritizing target peptides from the pool of all identified peptides for treatment of a disease (e.g., cancer). Some or all the steps of the method 2600 or method 2640 can be executed by a computational device or system. A suitable computational device and system are illustrated in **Figs. 4** and **5** as non-limiting examples. 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.

[00490] At step 2602 of method 2600 in **Fig. 26A**, one or more candidate target peptides can be identified from differently expressed alternative splicing junction events in one or more patient samples based at least in part on a comparison of the one or more patient samples to a control. In some embodiments, step 2602 can involve a method for identification of target peptide with splice junction analysis as disclosed herein, variations thereof, or alternatives thereto as understood by a person skilled in the art. For instance, step 2602 can be executed according to steps of the method 100 illustrated in **Fig. 1**, according to steps of the method 200 illustrated in

Fig. 2, or include a subset or combination steps of methods 100, 200 illustrated in **Fig. 1** and/or **Fig. 2**. Candidate target peptides can otherwise be identified at step 2602 according to target identification methods and examples presented herein, variations thereof, and alternatives thereto as understood by a person skilled in the pertinent art.

[00491] At step 2604, the one or more candidate target peptides identified at step 2602 are filtered based at least in part on a raw read count to find disease-specific alternative splicing junction events; thereby resulting in a filtered list of the one or more candidate target peptides.

[00492] At step 2606, the filtered list of the one or more candidate target peptides resulting from step 2604 can be screened based at least in part on protein function and structural domain. The filtered list can thus be further separated into two groups based on whether the target peptides resulted from alternative splicing junction events residing in a protein expressed on the cell surface. The first group comprises peptides derived from proteins expressed on a cell surface, while the second group comprises peptides derived from proteins not expressed on the cell surface. The two groups can then be evaluated separately.

[00493] The first group is evaluated according to steps 2610, 2612, and 2614.

[00494] At step 2610, peptides with normal tissue expression are eliminated from the first group based at least in part on a comparison of the peptides of the first group to a normal tissue atlas. In some embodiments, the normal tissue atlas is custom-built for the purpose of prioritizing candidate target peptides for further development to create a targeted disease therapy.

[00495] At step 2612, peptides with tumor type targeting are retained in the first group based at least in part on a comparison of peptides of the first group to a tumor-tissue atlas. In some embodiments, the tumor tissue atlas is custom-built for the purpose of prioritizing candidate target peptides for further development to create a targeted disease therapy.

[00496] At step 2614, remaining peptides of the first group are validated as comprising alternative splicing-derived peptide sequences based at least in part on mass-spectrometry of the remaining peptides of the first group.

[00497] The second group is evaluated according to steps 2620 and 2622.

[00498] At step 2620, peptides with normal tissue expression are eliminated from the second group based at least in part on a comparison of peptides of the first group to a normal tissue atlas comprising a human thymus alternative splicing atlas database.

[00499] At step 2622, remaining peptides of the second group can be validated as comprising HLA presented alternative splicing-derived peptide sequences based at least in part on immunopeptidomics.

[00500] At step 2630, the validated peptides of the first and/or second groups for can be prioritized for targeted therapy development.

[00501] In some embodiments the at least one target peptide prioritized at step (i) is derived from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

[00502] To summarize, candidate target peptides identified at step 2602, filtered based at least in part on disease-specific alternative splicing junction events at step 2604, divided into two groups based at least in part on protein function and structural domain at step 2606, then evaluated based on which group by fall into such that a first group having alternative splicing junction events follows steps 2610, 2612, and 2614 and such that a second group having proteins not expressed on a cell surface follows steps 2620 and 2622. Outputs from both groups can be prioritized at step 2630.

[00503] Various implementations of alternative splicing analysis as described herein may be used to identify one or more alternative splicing events or alternatively spliced isoforms that are differentially expressed between cells comprising a splicing factor gene mutation and those not comprising the splicing factor gene mutation. When such differential analysis is performed between diseased cells (e.g., cells derived from tumor samples of a certain cancer type) with and without the splicing factor gene mutation, disease-specific alternatively expressed isoforms that arise as a result of the splicing factor gene mutation, such as a mutation of SF3B1, or are otherwise correlated with the splicing factor gene mutation, can be identified. Accordingly, performing such analysis with regards to a disease that is correlated with a high prevalence of mutation with respect to a particular splicing factor gene or specific mutation thereof, may be used to identify target peptides that can be used to develop therapeutics that effectively target that disease. However, such analysis may identify isoforms that are not suitable or viable for therapeutic targeting to treat the

disease and/or may identify too many isoforms to feasibly pursue testing/development of each of the identified isoforms. Thus, one or more filtering steps, such as those described elsewhere herein, may be used identify or prioritize those isoforms that are the most suitable for deriving disease-specific target peptides.

[00504] **Figs. 26B-26C** include flow diagrams of a method 2640 for performing such an analysis.

[00505] At step 2642, according to certain aspects, the first step in such an analysis is to identify one or more differentially expressed isoforms that are associated with a splicing factor gene mutation in diseased tissue. As used herein, an "isoform" may refer to any of the multiple variant forms of a protein, or to any of the corresponding mRNA transcripts, that are produced from a gene as a result of an alternative splicing of pre-mRNA. An isoform, as used herein, may be identified by a full-length protein sequence (or corresponding mRNA sequence), or any portion thereof, which is sufficient to distinguish the isoform from other variants arising from the same alternative splicing event (i.e., the entire transcript sequence need not be determinable from the underlying sequencing data as long as the presence of alternative isoforms can be determined). One or more alternatively spliced isoforms that are differently expressed in RNA-seq samples derived from diseased cells are identified to generate a working list of isoforms wherein the diseased cells are associated with the disease for treatment, the one or more identified alternatively spliced isoforms are expressed at a higher frequency in a subset of the RNA-seq samples associated with a splicing factor gene mutation than in a subset of the RNA-seq samples that are not associated with the splicing factor gene mutation, and a targetable peptide sequence is associated with each identified isoform such that the targetable peptide sequence distinguishes the identified isoform from other variants over which the identified isoform was differentially expressed.

[00506] By comparing RNA sequencing (RNA-seq) data between at least two different data sets, a different rate or frequency of occurrence of one or more alternative splicing events may be uncovered. Differential expression analysis of alternative splicing events/alternatively spliced isoforms may reveal isoforms which are expressed at a higher frequency relative to alternative variants in one of the data sets (e.g., with statistical significance). The identification of one or more differentially expressed isoforms may involve a method for identification of a target peptide with splice junction analysis as disclosed herein, such as step 2602 of method 2600, method 100, method 200, or a subset and/or combination thereof, as well as variations thereof, or alternatives

thereto as understood by a person skilled in the art. Various methods, algorithms, and tools are known in the art for performing differential expression analysis of alternative splicing events, including, for example, STAR (Spliced Transcripts Alignment to a Reference), MAJIQ (Modeling Alternative Junction Inclusion Quantification), rMATS (replicate Multivariate Analysis of Transcript Splicing), FRASER (Find RAre Splicing Events in RNA-Seq), and JUM (Junction Usage Model). See, e.g., Dobin A, et al. Bioinformatics. 2013 Jan 1;29(1):15-21; Vaquero-Garcia J, et al. Nat Commun. 2023 Mar 3;14(1):1230; Shen S, et al. Proc Natl Acad Sci U S A. 2014 Dec 23;111(51):E5593-601; Mertes C, et al. Nat Commun. 2021 Jan 22;12(1):529; Proc Natl Acad Sci U S A. 2018 Aug 28;115(35):E8181-E8190; and Mehmood A, et al. Brief Bioinform. 2020 Dec 1;21(6):2052-2065. Various differential expression analysis methodologies may be implemented, for example, using OMICSOFT® (formerly ARRAY STUDIO) software (QIAGEN). According to some aspects, identifying one or more alternatively expressed isoforms comprises aligning or mapping RNA-seq reads to a reference genome, such as a Genome Reference Consortium (GRC) Human Build, and calling exon-exon junctions within the reads. The called junctions may then be quantitatively compared between two or more data sets to evaluate differential expression of alternatively spliced isoforms. The various steps of identifying differentially expressed isoforms may be computer implemented.

[00507] In specific instances, RNA-seq samples associated with a specific disease, such as a particular type of cancer, may be segregated between samples associated with any mutation in a particular splicing gene factor, such as SF3B1, or a specific mutation thereof (e.g., one or more of the SF3B1 mutations disclosed elsewhere herein) and samples not associated with the same. RNA-seq samples may be obtained, for example, from diseased cells or tissue (e.g., tumor cells or tissue) of different subjects/patients having, diagnosed with, or otherwise exhibiting the disease. Differential expression analysis of alternatively spliced isoforms may reveal certain isoforms that are correlated with mutation of the splicing factor gene in diseased cells and that may be used to target diseased cells. The presence or absence of the mutation in the splicing gene factor may be determined directly from the same RNA-seq data or from other data sets associated with the same samples or subjects/patients.

[00508] In various implementations, identifying one or more alternatively spliced isoforms may comprise identifying, selecting, or filtering a database of RNA-seq samples that are associated with a particular disease and/or that are associated with the presence and/or absence of a particular

mutation in a splicing factor gene or mutation status (e.g., mutated vs. wildtype). According to certain aspects, the splicing factor gene mutation has a prevalence among the RNA-seq samples derived from diseased cells above a mutation prevalence threshold frequency. The method of identifying one or more alternatively spliced isoforms that are differentially expressed may comprise selecting the splicing factor gene mutation by querying a database of RNA-seq samples (e.g., for splicing factor genes or specific mutations thereof that have a mutation prevalence above an input mutation prevalence threshold frequency). According to certain aspects, the mutation prevalence threshold frequency is at least about 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1.0%, 2.0%, 3.0%, 4.0%, 5.0%, 6.0%, 7.0%, 8.0%, 9.0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or higher. In certain instances, the mutation prevalence threshold frequency may be about 100%. According to certain aspects, the RNA-seq samples exclude any samples associated with a copy number variation of the splicing factor gene. The method of identifying one or more alternatively spliced isoforms that are differentially expressed may comprise filtering out any RNA-seq samples that are associated with a copy number variation of the splicing factor gene.

[00509] A working list of isoforms may be generated comprising isoforms that are expressed at a higher frequency in samples associated with the splicing factor gene mutation(s). A corresponding list of targetable peptide sequences may be generated from the working list of isoforms. In some instances, the working list of isoforms may be a working list of targetable peptide sequences. A targetable peptide sequence may comprise any peptide sequence that distinguishes the identified isoform from the working list of isoforms relative to the variant isoforms over which the identified isoform was found to be differentially expressed. Targetable peptide sequences may be derived directly from RNA-seq reads. In some instances, the targetable peptide sequence is the sequence of the full-length variant corresponding to the isoform or the mRNA transcript encoding the same. In some instances, the targetable peptide sequence is a shorter sequence derivable from the isoform's full length amino acid sequence (e.g., a domain, such as an extracellular domain), such that the full-length amino acid sequence comprises the targetable peptide sequence and the targetable peptide sequence is itself sufficient to distinguish the corresponding isoform from the variants over which the identified isoform was found to be differentially expressed. For example, the targetable peptide sequence may be a consensus sequence derived from RNA-seq data which is sufficient to identify a transcript as an alternatively spliced isoform. Targetable peptide sequences may be generated in a manner that prioritizes ideal

or desirable sequence lengths, as described elsewhere herein (e.g., isoform sequences may be truncated or minimized to a sequence that still satisfies the requirements of a targetable peptide sequence). In some instances, the method may be performed in a manner that associates a single targetable peptide sequence with each identified isoform. In some instances, an identified isoform may be associated with multiple targetable peptide sequences, which may or may not be overlapping. Generating the working list of isoforms and/or associated targetable peptide sequences may be computer implemented.

[00510] At step 2650, one or more filtering steps are performed on the working list of isoforms generated at step 2642. The working list of isoforms is updated after performance of each filtering step. Upon generating a working list of isoforms, one or more filtering steps, including those described elsewhere herein, may be used to identify or prioritize one or more target peptides, wherein the working list of isoforms (and associated targetable peptide sequences) is updated after each filtering step in a manner that identifies those isoforms that satisfy the filtering step's criterion/criteria. The working list of isoforms may be updated by filtering out from the list those isoforms that do not satisfy the criterion/criteria. For example, the isoforms that are filtered out may be physically removed or deleted from the working list. Additionally or alternatively, the working list may be updated with annotations that indicate which of the isoforms satisfy the criterion/criteria. The working list may be updated in a manner that allows computationally pulling from a list only those isoforms which satisfy one or more criteria and/or sorting of the list by satisfaction of the criterion/criteria of the one or more filtering steps (e.g., to rank the isoforms). Any of the filtering steps may be implemented in a manner that generates multiple different lists based on which isoforms do or do not satisfy a particular criterion. Performance of the one or more filtering steps may be computer implemented. In certain implementations, all of the performed filtering steps are computer implemented. Where multiple filtering steps are performed, the order of operation of the filtering steps may vary. In some instances, the filtering steps are performed such that filtering steps that are expected or more likely to minimize the number of target peptides relative to other filtering steps are performed earlier in the order of operations. In some instances, the filtering steps are performed such that filtering steps that exert less computational resources and/or require less computing time are performed earlier in the order of operations in order to minimize the number of isoforms in the working list of isoforms and/or the number of associated targetable peptides prior to completing filtering steps that require more computational resources

and/or more computing time. According to certain aspects, each target peptide identified or prioritized by the methods described herein can be associated with a targetable peptide sequence of one of the isoforms that is identified upon completion of the one or more filtering steps. For instance, the target peptide may have an amino acid sequence that comprises or consists of the corresponding targetable peptide sequence of an isoform from the working list of isoforms. In certain aspects, the amino acid sequence of the target peptide may, but is not required to be, longer than the corresponding targetable peptide sequence of an isoform from the working list of isoforms. For instance, the target peptide may contain extended portions of the isoform sequence that can ensure sequence uniqueness and/or provide the three-dimensional structure necessary to form, e.g., a neoantigen epitope. For each isoform that has the potential to give rise to a neoantigen epitope, it can be necessary to analyze at least a portion of the sequence through various filtering steps described herein. This analyzed portion can be distinct enough to differentiate the isoforms from other variants, such as those occurring in normal cells. Accordingly, each identified isoform can be associated with at least one targetable peptide sequence, wherein the targetable peptide sequence can be defined as a sequence unique enough to serve as a target.

[00511] **Fig. 26C** illustrates selected optional filtering steps that may be performed at step 2650 as part of method 2640 illustrated in **Fig. 26B**. The filtering steps can be performed in any combination and any order, and with complimentary steps not described herein as understood by a person skilled in the art and informed by the presented disclosure. For instance, step 2650 of **Fig. 26B** may include a subset or all of the filtering steps shown in **Fig. 26C**. Step 2650 may include filtering steps disclosed elsewhere herein or otherwise known to a person skilled in the pertinent art.

[00512] At optional step 2651, isoforms from the working list expressed in a subset of RNA-seq samples not associated with the splicing factor gene mutation above a control sample threshold frequency are filtered out of the working list. The one or more filtering steps may comprise filtering the working list of isoforms to filter out any isoforms that were expressed in the subset of RNA-seq samples not associated with the splicing factor gene mutation above a control sample threshold frequency. According to various aspects, the control sample threshold frequency may be about 0.00%, 0.01%, 0.02%, 0.03%, 0.04%, 0.05%, 0.10%, 0.20%, 0.30%, 0.40%, 0.50%, 1.00%, or 5.00%. According to certain aspects, the control sample threshold frequency is preferably 0% or about 0%. The one or more filtering steps may comprise filtering the working list of isoforms to

retain only identified isoforms which are expressed in the subset of RNA-seq samples associated with the splicing factor gene mutation above a mutation sample threshold frequency. Such isoforms are more likely to be commonly expressed in diseased cells of patients. According to certain aspects, the mutation sample threshold frequency is at least about 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1.0%, 2.0%, 3.0%, 4.0%, 5.0%, 6.0%, 7.0%, 8.0%, 9.0%, 10%, 20%, 30%, 40%, or 50%. The one or more filtering steps may comprise filtering the working list of isoforms to retain only identified isoforms which are expressed in one or more additional diseases above a secondary disease threshold frequency, wherein the one or more additional diseases are different from the disease which the diseased cells were associated with during the differential expression analysis. For example, the disease for treatment and the one or more additional diseases may be different types of cancer, such as the different types of cancers described elsewhere herein. According to certain aspects, the secondary disease threshold frequency is at least about 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1.0%, 2.0%, 3.0%, 4.0%, 5.0%, 6.0%, 7.0%, 8.0%, 9.0%, 10%, 20%, 30%, 40%, or 50%.

[00513] At optional step 2652, isoforms from the working list that are expressed in normal tissue above a normal tissue expression level threshold are filtered out from the working list. The one or more filtering steps may comprise filtering the working list of isoforms to filter out any isoforms that are expressed in normal tissue above a normal tissue expression level threshold. According to certain aspects, isoform expression in normal tissue can be assessed via the Gene Tissue Expression (GTEx) database (gtexportal.org/home/) or other suitable databases of gene expression or related data. For example, the GTEx database version 6, based on genome build GRCh38, contains gene expression data across 51 tissue types from 549 healthy donors. Gene expression may be queried for certain tissues. In some instances, the normal tissue expression level threshold may be assessed for only essential, normal tissues. Essential, normal tissues may include those tissues that cannot be sacrificed for the purpose of saving a patient's life, unlike non-essential tissues such as, e.g., breast, ovary, testes, etc. In some instances, the expression value within a normal tissue may be assessed based on percentile expression value across a number of samples for that tissue (e.g., the 95-, 96-, 97-, 98-, or 99- percentile expression value). The expression value within normal tissue may be assessed based on the tissue type exhibiting the highest level of expression. According to various aspects, the normal tissue expression level threshold may be about 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, or 1.0 transcripts per million (TPM). According

to certain aspects, the normal tissue expression level threshold is no detectable expression or about 0.0 TPM. The one or more filtering steps may comprise filtering the working list of isoforms to retain only identified isoforms which are expressed in the subset of RNA-seq samples associated with the splicing factor gene mutation above a mutation sample expression level threshold. The expression level in such samples may be assessed by raw read counts or normalized expression values as is known in the art. According to certain aspects, the mutation sample expression level threshold is at least about 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 TPM. The one or more filtering steps may comprise filtering the working list of isoforms to retain only identified isoforms, which are expressed in one or more additional diseases above a secondary disease expression level threshold, wherein the one or more additional diseases are different from the disease which the diseased cells were associated with during the differential expression analysis. For example, the disease for treatment and the one or more additional diseases may be different types of cancer, such as the different types of cancers described elsewhere herein. According to certain aspects, the secondary disease expression level threshold is at least about1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 TPM. The one or more filtering steps may comprise filtering the working list of isoforms to retain only isoforms that have been experimentally detected via mass spectrometry. According to certain aspects, the method of identifying one or more target peptides may comprise performing mass spectrometry to detect or validate the presence of the target peptides in diseased tissue.

[00514] At optional step 2653, isoforms from the working list that are not associated with a gene that encodes a protein that is expressed on the cell surface are filtered out from the working list. The one or more filtering steps may comprise filtering the working list of isoforms to filter out any isoforms that are not associated with a gene that encodes a protein that is expressed on the cell surface. The gene associated with an identified isoform may be determined by aligning the transcript (e.g., consensus sequence) used to identify the isoform to a reference genome as described elsewhere herein or to another canonical sequence. According to certain aspects, a gene may be determined to encode a protein that is expressed on the cell surface if the gene comprises an annotated extracellular domain or the gene is reported to be functionally associated with the plasma membrane.

[00515] At optional step 2654, isoforms from the working list that are not associated with a targetable peptide sequence that is aligned with at least a portion of an extracellular domain of a protein expressed on the cell surface are filtered out from the working list. In various implementations of the methods described herein for identifying one or more target peptides, target peptides which are expressed on the cell surface may be prioritized. Such target peptides may be readily accessible by certain types of therapeutics. Such isoforms are identified, for example, per various filtering steps described elsewhere herein. For instance, the one or more filtering steps may comprise filtering the working list of isoforms to filter out any isoforms that are not associated with a targetable peptide sequence that is aligned with at least a portion of an extracellular domain of a protein expressed on the cell surface. According to certain aspects, the targetable peptide sequence may be aligned with an extracellular domain if the targetable peptide sequence can be aligned with an annotated extracellular domain of a gene or otherwise aligned with a portion of the sequence predicted to form an extracellular portion of the protein. In some instances, the working list of isoforms is filtered to filter out any targetable peptide sequences that are not confined to an extracellular domain of a protein expressed on the cell surface. For example, domain prediction technologies and/or three-dimensional (3D) conformation modeling may be used to predict whether the targetable peptide sequence resides in an extracellular domain.

[00516] At optional step 2655, isoforms of the working list that i) result in the loss of a transmembrane domain and/or an extracellular domain, and/or ii) result in the truncation of an extracellular domain are filtered out from the working list. The one or more filtering steps may comprise filtering the working list of isoforms to filter out any isoforms that result in the loss of an extracellular domain. The one or more filtering steps may comprise filtering the working list of isoforms to filter out any isoforms that result in the loss of a transmembrane domain. Such isoforms may not be able to insert themselves in the transmembrane domain such that an extracellular domain is presented on the cell surface, available for targeting by certain therapeutics. The removal of a transmembrane and/or extracellular domain may be determined by aligning an isoform sequence to a reference genome or other annotated sequence (e.g., a canonical transcript) to identify missing segments of the peptide chain. According to certain aspects, isoforms that lack at least about 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of a domain (e.g., as determined by number of missing amino acids relative to the total number of amino acids in a domain) may be filtered from the working list of isoforms. According to certain aspects,

particularly where only a portion of the amino acids of a domain are missing in an isoform or annotated domains are not available, domain prediction technologies and/or 3D conformation modeling may be used to predict whether a domain is likely to be missing, such that the isoform may be filtered out. The one or more filtering steps may comprise filtering the working list of isoforms to filter out any isoforms that result in the truncation of an extracellular domain (i.e., the removal of amino acids from a free terminus of the peptide chain that is located in the extracellular domain). The loss of amino acids from a terminal portion of the peptide chain may be less likely to result in a targetable epitope that distinguishes the isoform from other variants. According to certain aspects, only isoforms which truncate the extracellular domain by about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 19, 20, 25, 30, 35, 40, 45, 50 or more amino acids are filtered from the working list of isoforms. According to certain aspects, only isoforms which truncate the extracellular domain by at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, or more (e.g., as determined by number of missing amino acids relative to the total number of amino acids in a domain) are filtered from the working list of isoforms.

[00517] At optional step 2656, the working list of isoforms may be filtered to retain only isoforms for which the targetable peptide sequence introduces i) an insertion of an isoform-specific amino acid sequence into an extracellular domain of a protein expressed on the cell surface and/or ii) a deletion of one or more amino acids in the extracellular domain. The one or more filtering steps may comprise filtering the working list of isoforms to retain only isoforms for which the targetable peptide sequences introduce an insertion of an isoform-specific amino acid sequence (e.g., a targetable peptide sequence) into an extracellular domain of a protein expressed on the cell surface and/or a deletion of one or more amino acids in the extracellular domain. The insertion of the isoform-specific amino acid sequence and/or the deletion may be defined as any event that distinguishes the identified isoform from other variants over which the identified isoform was differentially expressed. The position of an insertion or deletion of amino acids may be determined by aligning an isoform sequence to a reference genome or other annotated sequence (e.g., a canonical transcript) and then identifying the inserted or deleted amino acids to discern whether they fall within an extracellular domain. According to certain aspects, particularly where annotated domains are not available, domain prediction technologies and/or 3D conformation modeling may be used to predict whether an insertion or deletion of amino acids is likely to fall within an extracellular domain. According to certain aspects, the isoform-specific amino acid sequence

comprises at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more contiguous amino acids. According to certain aspects, the deletion is of at least about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more contiguous amino acids. According to certain aspects, the deletion is of no more than about 10, 15, 20, 25, or 30 contiguous amino acids. The deletion may be limited to deletions that do not occur at an N-terminal and/or a C-terminal portion of the extracellular domain. Such deletions may be less likely to result in targetable epitopes within an extracellular domain. According to certain aspects, the working list of isoforms may be filtered to retain only isoforms for which the insertion of the isoform-specific amino acid sequence and/or deletion results from a frame shift relative to another variant over which the identified isoform was differentially expressed. At step 2644 of the method 2640 in **Fig. 26B**, following performing the one or more filtering steps, one or more target peptides can be identified for treatment of the disease from the working list of isoforms. According to certain aspects, target peptides that can be presented by an MHC molecule (i.e., bound to or in complex with the binding groove of an MHC molecule) may be prioritized. Such target peptides may, for example, be recognizable by certain therapeutics which recognize pMHC complexes (e.g., TCRs and TCR-mimetic antibodies) and/or may be suitable for formulating as vaccines.

[00518] Accordingly, a method for identifying one or more target peptides may comprise generating a working list of peptide sequences for MHC binding. Each peptide sequence of the working list of peptide sequences for MHC binding may be a length suitable for binding an MHC molecule, such as an MHC class I molecule or MHC class II molecule, including suitable lengths described elsewhere herein. According to certain aspects, the MHC molecule is an MHC class I molecule, and the length suitable for binding is 8-12, 8-10, or 9 amino acids in length. According to certain other aspects, the MHC molecule is an MHC class II molecule and the length suitable for binding is 13-17 amino acids in length. Each peptide sequence of the working list of peptide sequences for MHC binding may be derivable from a targetable peptide sequence and not derivable from any other variants over which the corresponding identified isoform was differentially expressed. According to certain aspects, the working list of isoforms is filtered to retain only isoforms that correspond to a peptide sequence from the working list of peptide sequences for MHC binding. In some instances, the targetable peptide sequences associated with the working list of isoforms may be the working list of peptide sequences for MHC binding. For instance,

according to certain aspects, generating a working list of isoforms comprises generating a working list of peptide sequences for MHC binding.

[00519] According to certain aspects, identifying one or more target peptides may comprise filtering the working list of peptide sequences for MHC binding to retain only peptide sequences which are predicted to bind an MHC molecule. According to certain aspects, the binding affinity can be calculated using a computer-based model. As a non-limiting example, a commercial tool called NetMHCpan, which uses a machine learning model, utilizes a computer-based binding affinity model that may be suitable for computationally calculating the binding affinity. Any version of NetMHCpan may be used, for example, NetMHCpan4.0 or NetMHCpan4.1. Other non-limiting examples include MHCflurry (see, e.g., O'Donnell et al., Cell Syst. 2018 Jul 25;7(1):129-132.e4) and MixMHCPred (see, e.g., Boehm et al., BMC Bioinformatics volume 20, Article number: 7 (2019)). Predicting binding affinity estimates how tightly a peptide will bind to a specific MHC molecule. Binding affinity can measured as a predicted IC50 value in nanomolar (nM) range or a predicted percentile rank value. The predicted IC50 value attempts to approximate experimental IC50 value which measures the concentration of a competing peptide required to displace 50% of the binding of the peptide to the MHC molecule. According to certain aspects, if the predicted binding affinity $\leq$ 500 nM, a peptide is considered a binder. According to certain aspects, if the predicted binding affinity $\leq$ 50 nM, a peptide is considered a binder. The percentile rank value (e.g., %Rank_BA in NetMHCpan) normalizes the IC50 values across different MHC molecules. The rank is computed by comparing the predicted IC50 value of the peptide against the predicted IC50 values for a set of (on the order of 10^5) naturally occurring random peptides (see e.g., Jurtz V, et al. J Immunol. 2017 Nov 1;199(9):3360-3368). According to certain aspects, if the predicted percentile rank $\leq$ 2, a peptide is considered a binder. According to certain aspects, if the predicted percentile rank $\leq$ 0.5, a peptide is considered a binder.

[00520] According to certain aspects, identifying one or more target peptides may comprise filtering the working list of peptide sequences for MHC binding to retain only peptide sequences that have been experimentally detected or validated via mass spectrometry as bound to an MHC molecule. See, e.g., U.S. 11,747,340, the contents of which is incorporated herein by reference in its entirety for all purposes. According to certain aspects, the method of identifying one or more target peptides may comprise performing mass spectrometry to detect or validate the presence of the MHC-bound target peptides in diseased tissue.

[00521] According to various implementations of the methods described herein for identifying one or more target peptides for treatment of a disease, the disease for treatment may be a type of cancer. For example, the disease for treatment may be uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

[00522] According to various implementations of the methods described herein for identifying one or more target peptides for treatment of a disease, the splicing factor gene mutation is a mutation of splicing factor 3b subunit 1 (SF3B1). The one or more identified isoforms, before or after the completion of one or more of the filtering steps described herein (e.g., after the completion of multiple filtering steps), may comprise one or more isoforms associated with a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2. The working list of isoforms, before or after the completion of one or more of the filtering steps described herein (e.g., after the completion of multiple filtering steps), may comprise an isoform associated with a targetable peptide sequence comprising an amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169. In some instances, one or more of the targetable peptide sequences may be selected from the list of amino acid sequences comprising SEQ ID NOs: 1-37, 114-136, 160-164, and 169.

[00523] Various methods described herein relate to identifying and/or prioritizing (e.g., selecting/ranking/filtering) a list of target peptides. Any of these methods may further comprise synthesizing, or otherwise generating, one or more target peptides that have been identified or prioritized (e.g., a peptide comprising or consisting of a targetable sequence of an identified target peptide) or nucleic acids (e.g., vectors) encoding the same (e.g., for use in therapeutic development). Methods for peptide synthesis and nucleic acid synthesis include those known in the art and described herein. Any of these methods may further comprise isolating (e.g., purifying) a target peptide or nucleic acid encoding the same that has been identified or prioritized. Methods for peptide isolation and/or purification and/or nucleic acid isolation and/or purification include those known in the art and described herein. Any of these methods may further comprise

overexpressing a target peptide within a cell (e.g., a target peptide that is expressed on the cell surface). Any of these methods may further comprise loading a target peptide onto an MHC molecule or any suitable component thereof to form a pMHC complex as described elsewhere herein. Any of these methods may further comprise binding a target pMHC complex to an antigen-recognition molecule (e.g., an antibody, TCR, or CAR). For example, any of these methods may comprise screening a pMHC complex comprising a target peptide for binding to an antigen-recognition molecule. Any of these methods may comprise incubating and/or contacting a target peptide with one or more cells (e.g., pulsing cells (e.g., TAP-deficient cells, such as T2 cells) with a target peptide as described elsewhere herein). In a further aspect, provided herein are target peptides identified using the methods described herein. Accordingly, the present disclosure also provides libraries (e.g., splicing factor gene-specific libraries such as SF3B1-specific libraries, or mutation-specific libraries of the same) comprising one or more of the target peptides identified and/or prioritized using the methods described herein.

[00524] In a further aspect, any of the methods described herein for identifying and/or prioritizing a target peptide may comprise one or more steps towards developing a therapeutic, such as an antigen-recognition molecule (e.g., an antibody, TCR, or CAR) that recognizes one or more of the target peptides. For example, in order to generate one or more antigen-recognition molecules that recognize a target peptide, the method may further comprise immunizing a mouse (e.g., a genetically modified mouse having a genome that encodes human or humanized immunoglobulin heavy chain and/or light chain variable domains and/or human or humanized HLA molecules and β2m molecules) with an immunogen that comprises or encodes the target peptide. Antibodies, TCRs and/or CARs that recognize the target peptide may be screened for the development of an antigen-recognition molecule-based therapeutic. Similarly, the method may comprise screening a phage library expressing various peptide fragments for antigen-recognition molecules that bind to an antigen that comprises the target peptide (e.g., an immobilized antigen) to identify one or more antigen-recognition molecules that may be developed into an antigen-recognition molecule-based therapeutic. According to some aspects, a peptide-based or nucleic acid-based vaccine may comprise or encode the target peptide. The vaccine may be used to inoculate a subject against diseased cells (e.g., cancer cells) expressing the target peptide. Methods for formulating target peptides into immunogens and/or vaccines are well known in the art.

Method and System for Determining Viability of a Cancer Treatment

[00525] Another challenge to be addressed is to determine viability of a treatment prior to administration of the treatment to a patient. Presently, estimating likelihood of treatment viability for a given patient often involves comparing genetic markers of the patient to clinical trial data and other patient studies that show correlations between patient outcomes and the genetic markers.

[00526] The present disclosure further comprises a method and system for identification of target peptides for targeted disease therapy. Some or all the steps of the method can be executed by a computational device or system. A suitable computational device and system are illustrated in **Figs. 4** and **5** as non-limiting examples. 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.

[00527] **Fig. 3** is a flow diagram of a method 300 for determining viability of a cancer treatment for a subject in need thereof. In some embodiments, the subject has uveal melanoma, breast cancer, ovarian cancer, prostate cancer, skin cutaneous melanoma, acute myeloid leukemia, myelodysplastic syndrome, or chronic lymphocytic leukemia.

[00528] At step 302, an RNA sequence can be obtained from a sample obtained from the subject.

[00529] At step 304, a non-annotated exon splice junction in the RNA sequence from the sample can be identified.

[00530] At step 306, an expression value can be obtained such that the expression value is based at least in part on a count of non-annotated exon splice junctions associated with a gene mutation of a target peptide of the cancer treatment. In some embodiments, the gene mutation is a mutation of SF3B1 or other splicing factor gene as understood by a person skilled in the art. In some embodiments, the target peptide is derived from a second gene comprising ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

[00531] At step 308, the expression value can be compared to a count of non-annotated exon splice junctions in a sample lacking the gene mutation.

[00532] At step 310, the viability of the cancer treatment can be determined for the subject based at least in part on the comparison of the expression value to the count of non-annotated exon splice junctions in the sample lacking the gene mutation.

Computing Device and System

[00533] The present disclosure further comprises a computing device configured to perform computational processes disclosed herein.

[00534] **Fig. 4** illustrates a block diagram of an embodiment of a computing device 400 as a non-limiting example. As shown, computing device 400 may include one or more processor(s) 410, an input/output (I/O) device 420, a memory 430 containing an operating system ("OS") 440, a database 450, and a program 460.

[00535] Some or all of the steps of the computational methods disclosed herein can be executed by the computing device 400. The memory 430 can include instructions that, when executed by the processor 410, cause the processor to perform computational processes of the respective method 100, 200, 300. The I/O device 420 can be configured to communicate with one or more ancillary devices to receive inputs to perform certain steps of the respective method. The I/O device 420 can be configured to communicate with one or more external databases to perform certain steps of the respective method. Additionally, or alternatives, databases relied on to perform certain method steps can be stored whole, or in part in the database 450 of the computing device 400. The I/O device 420 can be configured to communicate with one or more ancillary devices to provide outputs of certain steps of the respective method.

[00536] Computing device 400 may be a single server or may be configured as a distributed computer system including multiple servers or computers that interoperate to perform one or more of the processes and functionalities associated with the disclosed embodiments. In some embodiments, computing device 400 may further include a peripheral interface, a transceiver, a mobile network interface in communication with processor 410, a bus configured to facilitate communication between the various components of computing device 400, and a power source configured to power one or more components of computing device 400. A peripheral interface may include the hardware, firmware and/or software that enables communication with various peripheral devices, such as media drives (e.g., magnetic disk, solid state, or optical disk drives),

other processing devices, or any other input source used in connection with the instant techniques. In some embodiments, a peripheral interface may include a serial port, a parallel port, a general-purpose input and output (GPIO) port, a game port, a universal serial bus (USB), a micro-USB port, a high-definition multimedia (HDMI) port, a video port, an audio port, a Bluetooth™ port, an NFC port, another like communication interface, or any combination thereof.

[00537] In some embodiments, a transceiver may be configured to communicate with compatible devices and ID tags when they are within a predetermined range. A transceiver may be compatible with one or more of: RFID, NFC, BluetoothTM, low-energy Bluetooth™ (BLE), WiFi™, ZigBee™, ABC protocols or similar technologies.

[00538] A mobile network interface may provide access to a cellular network, the Internet, or another wide-area network. In some embodiments, a mobile network interface may include hardware, firmware, and/or software that allows processor 410 to communicate with other devices via wired or wireless networks, whether local or wide area, private or public, as known in the art. A power source may be configured to provide an appropriate alternating current (AC) or direct current (DC) to power components.

[00539] Processor 410 may include one or more of a microprocessor, microcontroller, digital signal processor, co-processor or the like or combinations thereof capable of executing stored instructions and operating upon stored data. Memory 430 may include, in some implementations, one or more suitable types of memory (e.g., volatile or non-volatile memory, random access memory (RAM), read only memory (ROM), programmable read-only memory (PROM), erasable programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), magnetic disks, optical disks, floppy disks, hard disks, removable cartridges, flash memory, a redundant array of independent disks (RAID), and the like) for storing files, including an operating system, application programs (including, e.g., a web browser application, a widget or gadget engine, or other applications, as necessary), executable instructions, and data. In one embodiment, the processing techniques described herein are implemented as a combination of executable instructions and data within memory 430.

[00540] Processor 410 may be one or more known processing devices, such as a microprocessor from the Pentium™ family manufactured by Intel™ or the Turion™ family manufactured by AMD™. Processor 410 may constitute a single core or multiple core processor that executes parallel processes simultaneously. For example, processor 410 may be a single core

processor that is configured with virtual processing technologies. In certain embodiments, processor 410 may use logical processors to simultaneously execute and control multiple processes. Processor 410 may implement virtual machine technologies, or other similar known technologies to provide the ability to execute, control, run, manipulate, store, etc. multiple software processes, applications, programs, etc. Other types of processor arrangements could be implemented that provide for the capabilities disclosed herein as understood by a person skilled in the pertinent art.

[00541] Computing device 400 may include one or more storage devices configured to store information used by processor 410 (or other components) to perform certain functions related to the disclosed embodiments. In one example, computing device 400 may include memory 430 that includes instructions to enable processor 410 to execute one or more applications, such as server applications, network communication processes, and any other type of application or software known to be available on computer systems. Alternatively, the instructions, application programs, etc., may be stored in an external storage or available from a memory over a network. The one or more storage devices may be a volatile or non-volatile, magnetic, semiconductor, tape, optical, removable, non-removable, or other type of storage device or tangible computer-readable medium.

[00542] In one embodiment, computing device 400 may include memory 430 that includes instructions that, when executed by processor 410, perform one or more processes consistent with the functionalities disclosed herein. Methods, systems, and articles of manufacture consistent with disclosed embodiments are not limited to separate programs or computers configured to perform dedicated tasks. For example, computing device 400 may include memory 430 that may include one or more programs 460 to perform one or more functions of the disclosed embodiments. Moreover, processor 410 may execute one or more programs 460 located remotely from computing device 400. For example, computing device 400 may access one or more remote programs 460, that, when executed, perform functions related to disclosed embodiments.

[00543] Memory 430 may include one or more memory devices that store data and instructions used to perform one or more features of the disclosed embodiments. Memory 430 may also include any combination of one or more databases controlled by memory controller devices (e.g., server(s), etc.) or software, such as document management systems, Microsoft™ SQL databases, SharePoint™ databases, Oracle™ databases, Sybase™ databases, or other relational databases. Memory 430 may include software components that, when executed by processor 410,

perform one or more processes consistent with the disclosed embodiments. In some embodiments, memory 430 may include database 450 for storing related data to enable computing device 400 to perform one or more of the processes and functionalities associated with the disclosed embodiments.

[00544] Computing device 400 may also be communicatively connected to one or more memory devices (e.g., databases (not shown)) locally or through a network. The remote memory devices may be configured to store information and may be accessed and/or managed by computing device 400. By way of example, the remote memory devices may be document management systems, Microsoft™ SQL database, SharePoint™ databases, Oracle™ databases, Sybase™ databases, or other relational databases. Systems and methods consistent with disclosed embodiments, however, are not limited to separate databases or even to the use of a database.

[00545] Computing device 400 may also include one or more I/O devices 420 that may include one or more interfaces for receiving signals or input from devices and providing signals or output to one or more devices that allow data to be received and/or transmitted by computing device 400. For example, computing device 400 may include interface components, which may provide interfaces to one or more input devices, such as one or more keyboards, mouse devices, touch screens, track pads, trackballs, scroll wheels, digital cameras, microphones, sensors, and the like, that enable computing device 400 to receive data from one or more users (such as via user device 130).

[00546] In example embodiments of the disclosed technology, computing device 400 may include any number of hardware and/or software applications that are executed to facilitate any of the operations. The one or more I/O interfaces may be utilized to receive or collect data and/or user instructions from a wide variety of input devices. Received data may be processed by one or more computer processors as desired in various implementations of the disclosed technology and/or stored in one or more memory devices.

[00547] While computing device 400 has been described as one form for implementing the techniques described herein, other, functionally equivalent techniques may be employed as understood by a person skilled in the pertinent art. For example, as known in the art, some or all of the functionality implemented via executable instructions may also be implemented using firmware and/or hardware devices such as application specific integrated circuits (ASICs),

programmable logic arrays, state machines, etc. Furthermore, other implementations may include a greater or lesser number of components than those illustrated.

[00548] The present disclosure further comprises a computing network or system configured to perform computational processes disclosed herein.

[00549] **Fig. 5** illustrates a block diagram of an embodiment of a computing network 500 including computing device(s) 510, server(s) 520, memory store(s) 540, and a network 530 facilitating communication between each.

[00550] Some or all of the steps of the computational methods disclosed herein can be executed by the computing network 500. Steps of a computational method can be executed across multiple computing devices 510, servers 520, or a combination of computing devices 510 and servers 520 as understood by a person skilled in the pertinent art. Preferably, a primary set of instructions to execute the respective method is stored on a single computing device 510 or server 520. The primary set of instructions can cause the single computing device 510 or server 520 to communicate over the network 530 to retrieve data from one or more databases 540 and/or perform computational analysis through one or more servers 520 to accomplish the steps of the respective method.

[00551] Network 530 may be of any suitable type, including individual connections via the internet such as cellular or WiFi™ networks. In some embodiments, network 530 may connect terminals, services, and mobile devices using direct connections such as radio-frequency identification (RFID), near-field communication (NFC), Bluetooth™, low-energy Bluetooth™ (BLE), WiFi™, ZigBee™, ambient backscatter communications (ABC) protocols, USB, WAN, or LAN. Because the information transmitted may be personal or confidential, security concerns may dictate one or more of these types of connections be encrypted or otherwise secured. In some embodiments, however, the information being transmitted may be less personal, and therefore the network connections may be selected for convenience over security.

[00552] Certain embodiments and implementations of the disclosed technology are described above with reference to block and flow diagrams of systems and methods and/or computer program products according to example embodiments or implementations of the disclosed technology. It will be understood that one or more blocks of the block diagrams and flow diagrams, and combinations of blocks in the block diagrams and flow diagrams, respectively, can be implemented by computer-executable program instructions. Likewise, some blocks of the block

diagrams and flow diagrams may not necessarily need to 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.

[00553] 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 specified in the flow diagram block or blocks. 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 specified in the flow diagram block or blocks.

[00554] As an example, 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 specified in the flow diagram block or blocks. 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 specified in the flow diagram block or blocks.

[00555] Accordingly, blocks of the block diagrams and flow diagrams 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 each block of the block diagrams and flow diagrams, and combinations of blocks in the block diagrams and flow diagrams, 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.

[00556] Certain implementations of the disclosed technology are described above with reference to 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. Additionally, implementations of the disclosed technology can be utilized with internet of things (IoT) devices, smart televisions and media devices, appliances, automobiles, toys, and voice command devices, along with peripherals that interface with these devices.

EXAMPLES

[00557] 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.

Example 1. Determining a target gene for immunotherapy using splicing junctions

[00558] **Fig. 6** is a flow diagram of a method 600 used to determine a target gene for immunotherapy as non-limiting example.

[00559] At step 602, raw RNA-seq data (fastq files) were aligned to human genome build hg19 (i.e., B37). Another suitable build (e.g., B38) can be utilized as understood by a person skilled in the art.

[00560] At step 604, exon splicing junctions with annotated gene model were called with the human genome build, and junctions were categorized into: a) annotated junctions; or b) non-annotated junctions. The human genome build was queried to find locations of the exon splicing junctions within the human genome build, then these locations were compared to an annotated

gene model to categorize the exon splicing junctions as: a) annotated junctions; or b) non-annotated junctions.

[00561] At step 606, a differential splicing junction analysis was performed in which tumor samples were divided into two groups: a) with SF3B1 mutation; or b) without SF3B1 mutation. Differently spliced non-annotated junctions were compared between the two groups. The SF3B1 mutation was utilized due to a high prevalence of mutation in The Cancer Genome Atlas Program database. In some embodiments, an alternative splicing factor gene can be used in place of SF3B1 as understood by a person skilled in the art.

[00562] At step 608, the resulting non-annotated splice junctions were filtered by: a) whether the splice junction is from a surface protein; or b) whether a high percent of the splice junction were found in tumor samples with SF3B1 mutation; or c) whether low or no expression counts of the splice junction were found in normal tissue.

[00563] At step 610, target peptide sequences were generated based on selected non-annotated junction and the sequences were confirmed with mass spectrometry.

Example 2. Diagnostics to determine viability of cancer treatment

[00564] **Fig. 7** is a flow diagram of a method 700 to determine viability of a cancer treatment that targets a target peptide derived from a splicing factor gene mutation as a non-limiting example.

[00565] At step 702, raw RNA-seq data (fastq files) were aligned to human genome build hg19 (i.e., B37). Another suitable build (e.g., B38) can be utilized as understood by a person skilled in the art.

[00566] At step 704, exon splicing junctions with annotated gene model were called with the human genome build. The human genome build was queried to find locations of the exon splicing junctions within the human genome build.

[00567] At step 706, a heatmap was generated with previously identified differentially spliced junctions. The previously identified differentially spliced junctions were identified via the method 600 illustrated in **Fig. 6**. In some embodiments, differentially spliced junctions suitable for use in step 706 can be identified through other methods including those presented in this disclosure, variations thereof, and alternatives thereto.

[00568] The patterns in the heatmap generated at step 706 were examined. The heatmap may resemble any of the example heatmaps illustrated in, e.g., **Figs. 8-13**, or variations thereof as

understood by a person skilled in the art. Samples with SF3B1 mutation had a higher expression of differently spliced junctions (e.g., non-annotated junctions in **Figs. 8-13**) compared to SF3B1 wild-type samples and were visualized in the heat map by a distinctive intensity (e.g., darker intensity in **Figs. 8-13**). Thus, samples with SF3B1 mutation were identified by the splicing pattern of the sample without performing exome-seq to detect SF3B1 somatic mutation.

[00569] At step 708, samples with a higher expression of SF3BI mutations at the differently spliced junctions compared to wild-type were identified as candidates for treatment.

Example 3. Splicing factor analysis to identify target peptides for cancer treatment

[00570] Target peptides were identified through an unbiased data-driven methodology including an analysis of splicing factors with a focus on tumor samples.

[00571] First, a splicing factor gene with a high prevalence of mutation was identified. The Cancer Genome Atlas (TCGA) database was queried for splicing factor mutations, and SF3B1 was identified as the most prevalent mutated splicing factor gene in the database. A method disclosed herein can be applied to analysis of mutations associated with other splicing factor genes.

[00572] Next, a cancer type with prevalent mutations within the mutated splicing factor gene was identified. An analysis across cancer types identified the SF3B1 as most prevalent in Uveal Melanoma (UVM) in TCGA. A method disclosed herein can be applied to analysis of other disease, cancer, or tumor types.

[00573] There were 80 UVM tumor samples in TCGA at the time of analysis. Exon junction data, including junction counts, junction coordinates, gene name, mutation status, and copy number variation (CNV) status of these UVM samples were extracted from the Array Studio TCGA Land B37.

[00574] To minimize confounding effects, samples with SF3B1 copy number changes were removed. This resulted in 69 remaining UVM samples. Among those, 17 samples were with SF3B1 somatic mutations and were annotated as mutant (MUT), and the remaining 52 samples were annotated as wild-type (WT).

[00575] One challenge that was overcome in this analysis was the size of the data. Unlike gene expression data which contain only 30k to 50k annotated genes, each gene can have multiple, dozens, or scores of splice junctions. Therefore, each sample could easily have tens of millions of splice junction data points. The data size informed which statistical test was to be applied to the

comparison between SF3B1 wild-type samples and SF3B1 mutant samples. Based on the splicing analysis results, the analysis next focused on non-annotated in-frame junctions first, which effectively trimmed the data to about half the size. A total of 67,956 non-annotated in-frame junctions were used for the comparison of UVM tumor samples with an SF3B1 somatic mutation versus without an SF3B1 somatic mutation. DESeq2 test was performed to detect differentially spliced non-annotated in-frame junctions between SF3B1 MUT versus WT. A stringent threshold was applied to the differential splicing comparison to minimize false positives. Filtering parameters included: False Discovery Rate (FDR) < 0.00001 and junction counts >=10. This resulted in 778 non-annotated junctions spanning 670 genes that were differentially spliced between SF3B1 MUT versus WT in 69 UVM tumor samples.

[00576] **Fig. 8** is an illustration of a heatmap plotting expression of mutated non-annotated junctions for mutant and wild-type SF3B1 genes. Each column in the heatmap represents a sample of the 69 UVM samples. In the heatmap, expression in SF3B1 mutated samples (MUT) are on the left and expression in wild-type samples (WT) are on the right. Expression of mutated non-annotated junctions is indicated by intensity, with higher expression indicated by darker intensity. Each row in the heat map corresponds to a non-annotated junction of the 778 non-annotated junctions of the 670 genes. **Fig. 8** shows that samples with SF3B1 mutations have a higher expression mutated non-annotated junctions compared expression of mutated non-annotated junctions in SF3B1 wild-type samples.

[00577] **Fig. 9** is an illustration of a heatmap plotting expression of SF3B1 mutant samples (MUT), wild-type (WT) samples, and samples with other splicing factors with mutations (other-MUT). Each column in the heat map represents a sample in the 80 UVM tumor samples before copy number changes were removed. In the heatmap, MUT samples are on the left, WT samples are on the right, and other-MUT are in the middle. Note that samples with other splicing factor mutations (other-MUT) do not show the distinct pattern of samples with SF3B1 mutations.

[00578] **Figs. 8** and **9** also show that of UVM SF3B1 WT samples (WT), one sample (LF MUT) had a splicing pattern similar to the SF3B1 MUT samples. Upon further analysis, although TCGA indicated that the one sample was a WT sample, that sample contained an SF3B1 mutation with a low mutation frequency (LF MUT). This indicates that using these 778 non-annotated splicing junctions as a splicing signature, just as a gene signature, one can detect samples with an

SF3B1 mutation. In some instances, the splicing signature may be more sensitive than a gene signature.

[00579] **Fig. 10** is an illustration of a heatmap plotting expression of each of the 778 non-annotated junctions for 212 Melanoma PDX samples. Samples having an SF3B1 mutation (as determined by TCGA) are on the left and wild-type (as determined by TCGA) are on the right. By applying the 778 non-annotated in-frame splice junctions to 212 Melanoma PDX samples, additional samples originally not annotated as SF3B1 mutant were detected with convincing pattern. For instance, the SF3B1 genes represented by the three vertical lines on the right, wild-type portion of the heatmap in **Fig. 10** (as indicated by arrows) were originally not annotated as SF3B1 mutant. Because these three vertical lines were so easily identifiable upon creation of the heatmap, the SF3B1 samples indicated by these lines were investigated further, and it was found that these three samples were, in fact, SF3B1 mutants (R625C, R549C, D621N). This demonstrates that the heatmap can be used to identify mutated SF3B1 genes that were previously misidentified as wild-type. The non-annotated junctions were applied to other tumor types and the pattern still held, meaning the non-annotated junctions can be used to successfully identify SF3B1 mutations across multiple tumor types.

[00580] **Fig. 11** is an illustration of a heatmap plotting expression of 32 non-annotated junctions for the 69 UVM samples plotted in **Fig. 8**. Even when the list of the junctions was trimmed down, the pattern still held, meaning that 32 is a sufficient number of non-annotated junctions which can be used to identify SF3B1 mutants.

[00581] **Fig. 12** is an illustration of a heatmap plotting expression of the 32 non-annotated junctions plotted in **Fig. 11** for 1000 breast cancer (BRCA) tumor samples, 19 with SF3B1 mutations (left in plot).

[00582] **Fig. 13** is an illustration of a heatmap plotting expression of the 32 non-annotated junctions plotted in **Fig. 11** for 471 skin cancer melanoma tumor samples, 23 with SF3B1 mutations (left in plot).

[00583] **Figs. 12** and **13** show that using the criteria of membrane proteins, the list of 778 novel junctions can be slimmed down to 32 and the same pattern will still show up in multiple cancer types, such as, for example, acral lentiginous melanoma (ALM), acute myeloid leukemia (AML), breast cancer (BRCA), chronic lymphocytic leukemia (CLL), chronic myeloid leukemia (CML), esophageal carcinoma (ESCA), myeloproliferative neoplasms (MPN), myelodysplastic

syndrome (MDS), mesothelioma (MESO), pancreatic ductal adenocarcinoma (PDAC), primary malignant melanoma (PMM), skin cutaneous melanoma (SKCM), uveal melanoma (UVM), and also pre-tumor syndromes, clonal hematopoiesis of indeterminate potential (CHIP) and clonal cytopenia of undetermined significance (CCUS), etc.

[00584] By plotting top differentially expressed non-annotated in-frame junctions between wild-type and mutant SF3B1 samples in UVM, a pattern emerged. Non-annotated in-frame junctions were up-regulated in SF3B1 mutant samples compared to wild-type SF3B1 UVM tumor samples, especially in hotspot mutation samples. These results suggest that if the gene with erroneous splicing is tumor-associated, then the gene could be a target for cancer therapy.

[00585] Differentially spliced novel in-frame junctions can be used as a classifier, feature selector, or gene signature to identify tumor samples with an SF3B1 mutation.

[00586] A series of criteria were developed to trim down the list of genes, as described below.

[00587] First, surface proteins were chosen to get the list of 32 non-annotated junctions.

[00588] Second, looking at each cancer, the percentage of SF3B1 mutated samples carrying those non-annotated splice junction was checked. **Table 3** shows the percent of SF3B1-mutated samples carrying the 32 non-annotated junctions for UVM, BRCA, SKCM, ovarian cancer (OV), and chronic lymphocytic leukemia (CLL). The table below shows SLC3A2 as a top potential target, referred to herein as SLC3A2-APIS splice variant junction.

Table 3. Percent of SF3B1-mutated Samples Carrying the 32 Non-annotated Junctions

Gene	UVM	BRCA	SKCM	OV	CLL
SLC3A2	100	89.5	80	66.7	100
BACE1	100	21.1	60	50	-
BSG	100	73.6	66.7	33.3	14.3
PLXNB1	100	73.7	60	66.7	0
IL17RC	100	78.9	60	50	-
GPNMB	100	94.7	100	-	57
CD81	94.1	73.7	46.7	33.3	100
BCAM	94.1	100	53.3	100	

DGCR2	94.1	84.2	53.3	33.3	14.3
TNFRSF14	88.2	10.5	40	0	100
APP	88.2	89.5	60	33.3	0
GPR143	88.2	-	33.3	-	-
FXYD5	82.4	31.6	66.7	50	100
GYPC	82.4	10.5	33.3	16.7	0
PTK7	82.4	26.3	26.7	33.3	0
CLSTN1	76.5	68.4	40	66.7	28.6
MEGF8	76.5	26.3	40	33.3	28.6
NCSTN	70.6	57.9	46.7	33.3	57
SLC2A11	70.6	84.2	46.7	33.3	0
SDC3	64.7	36.8	40	16.7	0
MMP14	58.8	21.1	40	33.3	28.6
CA14	52.9	-	13.3	0	-
PLXNB3	52.9	31.6	40	0	0
HTR2B	17.6	26.3	33.3	33.3	57
NRG3	5.9	5.3	40	0	100

[00589] Third, expression of SLC3A2-APIS splice variant junction in normal tissues was checked. Minimal expression of SLC3A2-APIS splice variant junction was found in normal tissues of 16,964 samples, which were from 30 tissue types from GTEx.

[00590] SLC3A2-APIS splice variant junction was compared to blood samples in GTEx, and samples with high expression of genes with advanced age and potential clonal hematopoiesis of indeterminate origin (CHIP) carriers, were found.

[00591] **Fig. 28** shows a plot of SLC3A2-APIS splice variant junction in TCGA for 10,550 samples from 33 cancer types (x-axis) as provided in TCGA. SLC3A2-APIS expression was elevated over wild-type in many of the cancer types. Non-TCGA datasets of other cancer types, such as CLL and MDS were queried and found elevated expression of SLC3A2-APIS over wild-type.

[00592] **Fig. 29** shows a summary of SLC3A2-APIS splice variant junction in TCGA. The plot shows fold elevated expression of SLC3A2-APIS splice variants in mutated samples compared to wild-type (MUT/WT) samples, with the cancer types annotated along the x-axis.

[00593] **Fig. 30** shows a plot of SLC3A2-APIS splice variant junction in MDS for 81 samples with elevated expression in mutated samples (left) compared to wild-type samples (right).

[00594] **Fig. 31** shows a plot of SLC3A2-APIS splice variant junction in CLL for 84 samples with elevated expression in mutated samples (left) compared to wild-type samples (right).

[00595] The data demonstrate that the same novel in-frame splicing junction of SLC3A2 had elevated expression in SF3B1 mutated cells.

[00596] **Figs. 14**, **18**, and **22** show plots of SYVN1, OXA1L, and NET1 splice variant junctions in TCGA for 10,550 samples from 33 cancer types (x-axis), respectively.

[00597] **Figs. 15A-15C**, **19A-19C**, **23A-23C**, and **27A-27B** summarize SYVN1, OXA1L, NET1, and ENO1 splice variants in TCGA, respectively. The plots depicted in Figs. 15C, 19C, and 23C show fold elevated expression of SYVN1, OXA1L, NET1, and ENO1 splice variants in mutated samples compared to wild-type (MUT/WT) samples, with the cancer types annotated along the x-axis, respectively.

[00598] **Figs. 16**, **21**, and **25** show plots of SYVN1, OXA1L, and NET1 splice variant junctions in MDS for 81 samples with elevated expression in mutated samples (left) compared to wild-type samples (right), respectively.

[00599] **Figs. 17**, **20**, and **24** show plots of SYVN1, OXA1L, and NET1-splice variant junction in CLL for 84 samples with elevated expression in mutated samples (left) compared to wild-type samples (right), respectively.

[00600] The data described herein demonstrate that the same novel in-frame splicing junction of SYVN1, OXA1L, and NET1 had elevated expression in SF3B1 mutated cells.

[00601] Viability of splice variant as an oncology target was validated. Fluorescence-activated cell sorting (FACS) sorted single cells from cell line Mel202 that carries the SF3B1 mutation were used. A full-length whole transcriptome sequencing on single cells successfully showed that the R625G mutation and splice variant junctions are observed concurrently 100% of the time. Therefore, all samples with SF3B1 hotspot mutation have the splice variant junction.

[00602] Ultimately, the process described in the foregoing example can lead to discovery of antigens that can be used as drug targets. Another application of the present Example is to identify additional SF3B1 (or other splicing factor gene) mutations.

[00603] The present Example can be used to determine cancer type-specific splicing junction signatures, and to intersect or merge such signatures to determine conserved or comprehensive splicing junction signatures to identify potential drug targets.

Example 4. Identification of SF3B1 mutation-associated neoantigens or neoepitopes

[00604] The present Example relates to nucleic acid assay methods and compositions for identifying SF3B1 mutation-associated splicing neoantigens or neoepitopes.

[00605] RNA or DNA is sequenced from a cancer patient's tumor and/or healthy tissue to identify SF3B1 mutation-associated splicing variants from sequence datasets.

[00606] Tumor material is assessed for SF3B1 mutation-associated splicing variants by sequencing such as by RNA sequencing. Sequencing data are used to investigate splice variants common to multiple tumor types (e.g., ALM, AML, BRCA, CLL, CML, ESCA, MPN, MDS, MESO, PDAC, PMM, SKCM, UVM, pre-tumor syndromes, CHIP, and/or CCUS). Peptide stretches comprising variants predicted to give rise to multiple high-quality neoantigens are created *in silico* and are filtered through the application of prediction algorithms or used to identify MHC-associated neoepitopes via mass spectrometry data. Where appropriate, suitable MHC-binding SF3B1 mutation-associated neoepitope sets are used to identify physiologically relevant SF3B1 mutation-associated neoepitope-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.

[00607] SF3B1 mutation-associated neoepitopes are determined by sequencing the genome and/or exome transcriptome of tumor tissue and/or healthy tissue from a cancer patient using next generation sequencing (NGS) approaches. For example, genes selected based on the presence of alterations induced by SF3B1 mutations and ability to act as an antigen are sequenced using RNA sequencing (RNA-Seq).

[00608] Assays are employed to identify SF3B1 mutation-associated neoepitopes in biological samples. Suitable assays used to identify SF3B1 mutation-associated neoepitopes 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.

[00609] The sequencing data derived from determining the presence of SF3B1 mutation-associated neoepitopes in a cancer patient is analyzed to predict personal SF3B1 mutation-associated aberrant splicing 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 SF3B1mutation-associated splicing neoantigens.

[00610] SF3B1 mutation-associated tumor antigens are assessed by their affinity to MHC molecules.

[00611] 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 SF3B1 mutation-associated epitopes for each patient is identified and prioritized. SF3B1 mutation-associated epitope prediction algorithms are used to predict binding of candidate peptides to MHC class I molecules or MHC class II molecules.

[00612] 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 as BIMAS, SYFPEITHI, or Rankpep.

[00613] 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 SF3B1 mutation-associated antigen. Such algorithms predict peptide binding to different MHC class I variants based on artificial neural networks (ANN), providing predicted IC50 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.

[00614] Proteomic-based methods for identifying tumor specific SF3B1 mutation-associated antigens such as immunopeptidomics are used. Protein sequencing of enzymatic digests using multidimensional MS techniques including tandem mass spectrometry (MS/MS) is used to identify SF3B1 mutation-associated 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 SF3B1 mutation-associated antigens. In some instances, meta-shotgun protein sequencing is used to identify expressed SF3B1 mutation-associated antigens.

[00615] Tumor specific SF3B1 mutation-associated antigens are identified using MHC multimers to identify SF3B1 mutation-associated antigen-specific T cell responses. For example, high-throughput analysis of SF3B1 mutation-associated 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 SF3B1 mutation-associated antigens, or as a secondary screening protocol to assess SF3B1 mutation-associated antigens to which a patient may have already been exposed, which may support the selection of candidate SF3B1 mutation-associated 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 SF3B1 mutation-associated antigens.

[00616] Pulsing antigen presenting cells (APCs) with relatively long synthetic peptides that encompass minimal T cell epitopes is used to identify SF3B1 mutation-associated epitopes. Nonsynonymous mutated epitopes are identified in tumors by evaluating the response of CD4+ tumor infiltrating lymphocytes (TIL) 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 SF3B1 mutation-associated epitope-reactive T cells isolated from a tumor associated with SF3B1 mutations.

[00617] A tandem minigene screening approach is used to identify SF3B1 mutation-associated 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, SF3B1 mutation-associated epitopes are identified in cancer patients, e.g., hepatocellular carcinoma (HCC) patients.

[00618] SF3B1 mutation-associated 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 5. Determination of MHC binding capacity

[00619] 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 horseradish peroxidase (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 6. Preparation of peptide-MHC (pMHC) complexes

[00620] This Example relates to a method for the preparation of soluble recombinant HLA loaded with an SF3B1 mutation-associated neoantigenic peptide.

[00621] 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.

[00622] 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.

[00623] 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.

[00624] 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.

[00625] 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.

[00626] 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 7. Identification of T cell receptors (TCRs) that bind to pMHC complexes

[00627] 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.

[00628] 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 indicates that the TCR is not highly cross reactive.

[00629] 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.

[00630] 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 8. Characterization of binding to MHC and stability of pMHC complex

[00631] *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 SF3B1 mutation-associated neoantigenic 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.

[00632] *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.

[00633] 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 are 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.

[00634] 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 a fluorometer. Dose response curves are generated. The amount of peptide required to achieve 50% of the maximum is recorded as ED50 value.

[00635] 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).

[00636] 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 9. Determination of responses against tumor cells

[00637] A suitable number of groups of mice are immunized with a plasmid expressing SF3B1 mutation-associated neoantigenic peptides disclosed herein by direct inoculation. By way of a non-limiting example, mice are vaccinated with pDNA (or RNA) encoding SF3B1-mutation associated splicing neoantigen, and/or SF3B1 mutation-associated neoantigenic peptides. Vaccination can occur via a suitable route of administration, e.g., via electroporation for pDNA, and any of various routes for RNA or protein (e.g., intramuscular, intravenous, intranodal). Mice are inoculated 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 can be 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. Immune response is monitored (e.g., by ELISpot assays) and the ability to reject tumors is tested. In some cases, splenocytes are stimulated *ex vivo* with POI and tested against labeled tumor cells at various E:T ratios.

Example 10. *In vivo* assessment of enhanced immunity against SF3B1 mutation-associated neoantigenic peptides

[00638] A suitable number of groups of mice are immunized with a plasmid expressing SF3B1 mutation-associated neoantigenic 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.

[00639] To evaluate the *in vivo* response against SF3B1 mutation-associated neoantigenic 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 peripheral blood mononuclear cells (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 11. Testing of POI's increased immunogenicity

[00640] POIs are used in *in vitro* for immunization of blood to generate cytotoxic T lymphocytes (CTLs).

[00641] 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 12. $CD8^+$ T cell responses against peptides

[00642] 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 IFN-γ secretion.

Example 13. Cytokine production assay

[00643] For cytokine (e.g., IL-2 and IFN-γ) 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 14. Enhancement of a peptide-specific $CD8^+$ T cell immunity in a peptide-presenting tumor model

[00644] For *in vivo* studies, the ability of a SF3B1 mutation-associated neoantigenic peptide of interest (POI) 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.

[00645] 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 SF3B1 mutation-associated neoantigenic 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.

[00646] Below are additional examples of methods that may be used in accordance with the disclosure.

[00647] *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.

[00648] *Pulsing of APCs with peptides.* APC are incubated with POI 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.

[00649] *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.

[00650] 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.

[00651] 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.

[00652] *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^{TM} 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.

[00653] *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.

[00654] 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.

[00655] 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.

[00656] *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.

[00657] *Peptide immunization of mice.* For immunization, a suitable amount of a SF3B1 mutation-associated neoantigenic 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).

[00658] *Quantification of cytotoxicity activity by flow cytometry.*

[00659] 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 SF3B1 mutation-associated neoantigenic 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).

[00660] 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.

[00661] Flow cytometry staining and acquisition: Cells are transferred to multi-well cell culture plate, washed in FACS staining buffer and stained suitable fluorescently-labeled fluorescent agents for detection of expressed transcripts encoding SF3B1 splicing neoantigens 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.

[00662] *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.

* * *

[00663] 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.

[00664] 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

What is claimed is:

1. An isolated peptide comprising an amino acid sequence, wherein the amino acid sequence is at least 90% identical to the amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169, or a pharmaceutically acceptable salt thereof, or a fragment or derivative thereof, wherein the isolated peptide is about 8-200 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: 1-37, 114-136, 160-164, and 169.

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: 1-37, 114-136, 160-164, and 169.

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: 1-37, 114-136, 160-164, and 169.

5. An isolated peptide comprising two or more amino acid sequences selected from any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169 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 a fragment of any one of SEQ ID NOs: 1-37, 114-136, 160-164, and 169, wherein the fragment is 8-200 amino acids in length.

7. The isolated peptide of any one of claims 1-6, 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.

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

9. The isolated peptide of any one of claims 1-7, wherein the isolated peptide is produced synthetically.

10. The isolated peptide of any one of claims 1-9, wherein the isolated peptide, or pharmaceutically acceptable salt thereof, or fragment or derivative thereof induces an immune response specific for a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in a subject when presented in a complex with a major histocompatibility complex (MHC) molecule, or a fragment or derivative thereof, on the surface of an antigen presenting cell (APC).

11. The isolated peptide of claim 10, wherein the tumor is uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

12. The isolated peptide of claim 11, wherein the tumor is uveal melanoma.

13. The isolated peptide of any one of claims 10-12, wherein the one or more mutations of SF3B1 is capable of causing aberrant splicing during expression of a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

14. A fusion protein comprising one or more isolated peptides of any one of claims 1-13, fused to one or more heterologous molecules.

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

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

17. The fusion protein 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 fusion protein of any one of claims 14-17, wherein the isolated peptide, or pharmaceutically acceptable salt thereof, or a fragment or derivative thereof, induces an immune response specific for a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in the subject when presented in a complex with the MHC molecule, or a fragment or derivative thereof, on the surface of an antigen presenting cell (APC).

19. The fusion protein of claim 18, wherein the tumor is uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

20. The fusion protein of claim 19, wherein the tumor is uveal melanoma.

21. The fusion protein of any one of claims 18-20, wherein the one or more mutations of SF3B1 is capable of causing aberrant splicing during expression of a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

22. A conjugate comprising one or more isolated peptides of any one of claims 1-13, conjugated to one or more heterologous molecules.

23. The conjugate of claim 22, wherein the one or more heterologous molecules enhance a peptide-specific immune response in a subject.

24. The conjugate of claim 22, wherein the one or more heterologous molecules mediate peptide delivery to a specific site within a subject.

25. The conjugate of any one of claims 22-24, wherein the one or more heterologous molecules are an MHC molecule, or a fragment or derivative thereof.

26. The conjugate of any one of claims 22-25, wherein the isolated peptide, or pharmaceutically acceptable salt thereof, or a fragment or derivative thereof induces an immune response specific for a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in the subject when presented in a complex with an MHC molecule, or a fragment of derivative thereof, on the surface of an antigen presenting cell (APC).

27. The conjugate of claim 26, wherein the tumor is uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

28. The conjugate of claim 27, wherein the tumor is uveal melanoma.

29. The conjugate of any one of claims 26-28, wherein the one or more mutations of SF3B1 is capable of causing aberrant splicing during expression of a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

30. The conjugate of any one of claims 22-29, wherein the one or more peptides are conjugated to a particle.

31. An oligomeric complex comprising two or more isolated peptides of any one of claims 1-13.

32. A non-covalent complex comprising the isolated peptide of any one of claims 1-13 and an MHC molecule, or a fragment or derivative thereof.

33. The non-covalent complex of claim 32, wherein the MHC molecule, or the fragment or derivative thereof, is a class I MHC molecule.

34. The non-covalent complex of claim 33, wherein the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

35. The non-covalent complex of claim 32, wherein the MHC molecule, or the fragment or derivative thereof, is a class II MHC molecule.

36. The non-covalent complex of claim 35, wherein the class II MHC molecule is a class II HLA molecule.

37. A fusion protein comprising the isolated peptide of any one of claims 1-13, and an MHC molecule, or a fragment or derivative thereof.

38. The fusion protein of claim 37, wherein the MHC molecule, or the fragment or derivative thereof, is a class I MHC molecule.

39. The fusion protein of claim 38, wherein the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

40. The fusion protein of claim 37, wherein the MHC molecule, or the fragment or derivative thereof, is a class II MHC molecule.

41. The fusion protein of claim 40, wherein the class II MHC molecule is a class II HLA molecule.

42. A conjugate comprising the isolated peptide of any one of claims 1-13 and an MHC molecule, or a fragment or derivative thereof.

43. The conjugate of claim 42, wherein the MHC molecule, or the fragment or derivative thereof, is a class I MHC molecule.

44. The conjugate of claim 43, wherein the class I MHC molecule is a class I human leukocyte antigen (HLA) molecule.

45. The conjugate of claim 42, wherein the MHC molecule, or the fragment or derivative thereof, is a class II MHC molecule.

46. The conjugate of claim 45, wherein the class II MHC molecule is a class II HLA molecule.

47. A pharmaceutical composition comprising (i) one or more isolated peptides of any one of claims 1-13, one or more fusion proteins of any one of claims 14-21 and 37-41, one or more conjugates of any one of claims 22-30 and 42-46, one or more oligomeric complexes of claim 31, or one or more non-covalent complexes of any one of claims 32-36, or any combination thereof; and (ii) a pharmaceutically acceptable carrier or excipient.

48. The pharmaceutical composition of claim 47, further comprising an adjuvant.

49. An isolated molecule that binds the isolated peptide of any one of claims 1-13, the fusion protein of any one of claims 14-21 and 37-41, the conjugate of any one of claims 22-30 and 42-46, the oligomeric complex of claim 31, or the non-covalent complex of any one of claims 32-36.

50. The isolated molecule of claim 49, wherein the isolated molecule is an antibody or an antigen-binding fragment thereof.

51. The isolated molecule of claim 50, wherein the antibody is a bispecific antibody.

52. The isolated molecule of claim 49, wherein the isolated molecule is an alternative scaffold.

53. The isolated molecule of claim 49, wherein the isolated molecule is a chimeric antigen receptor (CAR).

54. The isolated molecule of claim 49, wherein the isolated molecule is a T cell receptor (TCR).

55. An isolated cell comprising the CAR of claim 53.

56. The isolated cell of claim 55, wherein the isolated cell is an immune cell.

57. The isolated cell of claim 56, wherein the immune cell is a T cell, a natural killer (NK) cell, or a macrophage.

58. An isolated cell comprising the TCR of claim 54.

59. The isolated cell of claim 58, wherein the isolated cell is an immune cell.

60. The isolated cell of claim 59, wherein the immune cell is a T cell, an NK cell, or a macrophage.

61. A pharmaceutical composition comprising (i) the isolated molecule of any one of claims 49-54, or the isolated cell of any one of claims 55-60; and (ii) a pharmaceutically acceptable carrier or excipient.

62. An isolated polynucleotide comprising a nucleotide sequence encoding one or more isolated peptides of any one of claims 1-13 or the fusion protein of any one of claims 14-21 and 37-41.

63. The isolated polynucleotide of claim 62, wherein the nucleotide sequence is operably linked to a promoter.

64. The isolated polynucleotide of claim 62 or 63, wherein the isolated polynucleotide comprises deoxyribonucleic acid (DNA).

65. The isolated polynucleotide of claim 62 or 63, wherein the isolated polynucleotide comprises ribonucleic acid (RNA).

66. The isolated polynucleotide of claim 65, wherein the RNA is messenger RNA (mRNA).

67. The isolated polynucleotide of claim 65, wherein the RNA is self-replicating RNA.

68. A vector comprising the isolated polynucleotide of any one of claims 62-67.

69. The vector of claim 68, wherein the vector is an expression vector.

70. The vector of claim 68 or 69, wherein the vector is a viral vector.

71. An isolated cell comprising the isolated polynucleotide of any one of claims 62-67 or the vector of any one of claims 68-70.

72. The isolated cell of claim 71, wherein the isolated cell is a prokaryotic cell.

73. The isolated cell of claim 71, wherein the isolated cell is a eukaryotic cell.

74. The isolated cell of claim 73, wherein the isolated cell is an APC.

75. A pharmaceutical composition comprising (i) the isolated polynucleotide of any one of claims 62-67 or the vector of any one of claims 68-70; and (ii) a pharmaceutically acceptable carrier or excipient.

76. The pharmaceutical composition of claim 75, wherein the pharmaceutically acceptable carrier is a lipid nanoparticle carrier.

77. A method of inducing an immune response against a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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 1-13;
b) the fusion protein of any one of claims 14-21 and 37-41;
c) the conjugate of any one of claims 22-30 and 42-46;
d) the oligomeric complex of claim 31;
e) the non-covalent complex of any one of claims 32-36;
f) the pharmaceutical composition of any one of claims 47, 48, 61, 75, and 76;
g) the isolated molecule of any one of claims 49-54;
h) the isolated cell of any one of claims 55-60 and 71-74;
i) the isolated polynucleotide of any one of claims 62-67; or
j) the vector of any one of claims 68-70.

78. A method of inducing an immune response against a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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 1-13.

79. A method of inducing an immune response against a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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 1-13 in complex with an MHC molecule, or a fragment or derivative thereof.

80. A method of treating a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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 1-13;
b) the fusion protein of any one of claims 14-21 and 37-41;
c) the conjugate of any one of claims 22-30 and 42-46;
d) the oligomeric complex of claim 31;
e) the non-covalent complex of any one of claims 32-36;
f) the pharmaceutical composition of any one of claims 47, 48, 61, 75, and 76;
g) the isolated molecule of any one of claims 49-54;
h) the isolated cell of any one of claims 55-60 and 71-74;
i) the isolated polynucleotide of any one of claims 62-67; or
j) the vector of any one of claims 68-70.

81. A method of preventing or reducing the likelihood of a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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 1-13;
b) the fusion protein of any one of claims 14-21 and 37-41;
c) the conjugate of any one of claims 22-31 and 42-46;
d) the oligomeric complex of claim 31;
e) the non-covalent complex of any one of claims 32-36;
f) the pharmaceutical composition of any one of claims 47, 48, 61, 75, and 76;
g) the isolated molecule of any one of claims 49-54;
h) the isolated cell of any one of claims 55-60 and 71-74;
i) the isolated polynucleotide of any one of claims 62-67; or
j) the vector of any one of claims 68-70.

82. A method of treating a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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 1-13.

83. A method of preventing or reducing the likelihood of a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) 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 1-13.

84. The method of any one of claims 77-83, wherein the isolated peptide, or pharmaceutically acceptable salt thereof, or a fragment or derivative thereof induces an immune response specific for a tumor associated with one or more mutations of splicing factor 3b subunit 1 (SF3B1) in the subject when presented in a complex with an MHC molecule, or a fragment, or derivative thereof on the surface of an antigen presenting cell (APC).

85. The method of any one of claims 77- 84, wherein the tumor is a uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or endometrial cancer.

86. The method of claim 85, wherein the tumor is uveal melanoma.

87. The method of any one of claims 77-86, wherein the one or more mutations of SF3B1 is capable of causing aberrant splicing during expression of a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

88. A kit comprising:

(i) a) one or more isolated peptides of any one of claims 1-13;

b) the fusion protein of any one of claims 14-21 and 37-41;

c) the conjugate of any one of claims 22-30 and 42-46;

d) the oligomeric complex of claim 31;

e) the non-covalent complex of any one of claims 32-36;

f) the pharmaceutical composition of any one of claims 47, 48, 61, 75, and 76;

g) the isolated molecule of any one of claims 49-54;

h) the isolated cell of any one of claims 55-60 and 71-74;

i) the isolated polynucleotide of any one of claims 62-67; or

j) the vector of any one of claims 68-70; and

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

89. A computer implemented method for identifying a target peptide for cancer treatment for a subject in need thereof, the method comprising:

a) querying a first database comprising a plurality of first gene samples to identify a first gene associated with the plurality of first gene samples, the first gene having a prevalence of mutation beyond a predetermined threshold corresponding to a first cancer type;

b) filtering the plurality of first gene samples by gene samples that correspond to an exon splicing junction mutation;

c) identifying a second gene corresponding to the filtered plurality of first gene samples, wherein the second gene is associated with the first cancer type;

d) generating a peptide sequence based on the filtered plurality of first gene samples, wherein the generated peptide sequence is a sequence associated with the target peptide; and

e) verifying that the generated peptide sequence is associated with the second gene.

90. The computer implemented method of claim 89, wherein the first gene is SF3B1.

91. The computer implemented method of claim 89, wherein the second gene is ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16,

USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

92. The computer implemented method of claim 89, wherein the filtering the plurality of first gene samples comprises comparing a tumor sample of the first cancer type that is not associated with the exon splicing junction mutation to a tumor sample of the first cancer type that is associated with the exon splicing junction mutation.

93. The computer implemented method of claim 89, wherein the filtering the plurality of first gene samples further comprises removing one or more gene samples of the plurality of first gene samples that are associated with copy number variations.

94. The computer implemented method of claim 89, further comprising:
f) categorizing gene samples of the filtered plurality of first gene samples as annotated or non-annotated based on a comparison of the gene samples of the filtered plurality of first gene samples to an annotated gene model,
wherein the second gene is categorized as non-annotated.

95. The computer implemented method of claim 89, further comprising further filtering the plurality of first gene samples by identifying gene samples of the plurality of first gene samples corresponding to a surface protein.

96. The computer implemented method of claim 95, wherein the surface protein comprises one or more peptides that form an MHC-peptide complex with an MHC molecule.

97. The computer implemented method of claim 95, wherein further filtering the plurality of first gene samples further comprises identifying exon splicing junction mutations that are not expressed in non-cancerous tissue.

98. The computer implemented method of claim 89, wherein verifying that the generated peptide sequences are associated with the second gene further comprises performing mass spectrometry on an MHC-peptide complex associated with the cancer to confirm a presence of the generated peptide sequence.

99. The computer implemented method of claim 89, wherein the first database comprises a Genome Reference Consortium (GRC) Human Build.

100. A computer implemented method to determine viability of a cancer treatment for a subject in need thereof, the method comprising:
a) obtaining a ribonucleic acid (RNA) sequence from a sample obtained from the subject;
b) identifying a non-annotated exon splice junction in the RNA sequence from the sample;
c) obtaining an expression value based at least in part on a count of non-annotated exon splice junctions associated with a gene mutation of a target peptide of the cancer treatment;
d) comparing the expression value to a count of non-annotated exon splice junctions in a sample lacking the gene mutation; and
e) determining the viability of the cancer treatment for the subject based at least in part on the comparison of the expression value to the count of non-annotated exon splice junctions in the sample lacking the gene mutation.

101. The computer implemented method of claim 100, wherein the subject has uveal melanoma, breast cancer, ovarian cancer, prostate cancer, skin cutaneous melanoma, acute myeloid leukemia, myelodysplastic syndrome, or chronic lymphocytic leukemia.

102. The computer implemented method of claim 100, wherein the gene mutation comprises a mutation of SF3B1.

103. The computer implemented method of claim 102, wherein the target peptide is derived from a second gene comprising ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1,

BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

104. A computer implemented method for identifying a target peptide for treatment of a disease, the method comprising:
a) identifying a splicing factor gene having a significant prevalence of mutation in the disease in a database, wherein the database includes genetic data of samples of the disease;
b) identifying differently expressed splicing junctions of the splicing factor gene based on a comparison of splicing junctions of genetic data of samples that have a mutation in the splicing factor gene compared to splicing junctions of genetic data of samples that lack a mutation in the splicing factor gene; and
c) determining the target peptide based at least in part on the differently expressed splicing junctions.

105. The method of claim 104, further comprising:
d) identifying a disease-specific splicing junction based on one or more MHC-peptide complexes associated with the disease, such that the disease-specific splicing junction has minimal expression in non-diseased samples;
e) identifying a disease-specific mutated splicing junction as being both a disease-specific splicing junction and a differently expressed splicing junction; and
f) determining the target peptide based at least in part the disease-specific mutated splicing junction.

106. The computer implemented method of claim 104, wherein the disease-specific mutated splicing junction comprises SLC3A2-APIS splice variant junction.

107. The computer implemented method of claim 104, wherein the target peptide is derived from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81,

GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

108. The computer implemented method of claim 104, wherein the splicing factor gene comprises SF3B1.

109. A computer implemented method for prioritizing candidate target peptides for antigen development, the method comprising:

a) identifying one or more candidate target peptides from differently expressed alternative splicing junction events in one or more patient samples based at least in part on a comparison of the one or more patient samples to a control;

b) filtering the one or more candidate target peptides based at least in part on a raw read count to find disease-specific alternative splicing junction events resulting in a filtered list of the one or more candidate target peptides;

c) screening the filtered list of the one or more candidate target peptides based at least in part on protein function and structural domain to separate the filtered list of the one or more candidate target peptides into a first group comprising peptides derived from proteins expressed on a cell surface and a second group comprising peptides derived from proteins not expressed on the cell surface;

d) eliminating peptides with normal tissue expression from the first group based at least in part on a comparison of peptides of the first group to a normal tissue atlas;

e) retaining peptides with tumor type targeting in the first group based at least in part on a comparison of peptides of the first group to a tumor tissue atlas;

f) validating remaining peptides of the first group as comprising alternative splicing-derived peptide sequences based at least in part on mass-spectrometry of the remaining peptides of the first group;

g) eliminating peptides with normal tissue expression from the second group based at least in part on a comparison of peptides of the first group to a normal tissue atlas comprising a human thymus alternative splicing atlas database;

h) validating remaining peptides of the second group as comprising HLA presented alternative splicing-derived peptide sequences based at least in part on immunopeptidomics of the remaining peptides of the second group; and

i) prioritizing at least one target peptide comprising a validated peptide from the first group as identified through steps (a) through (f) and/or comprising a validated peptide from the second group as identified through steps (a) through (c), (g), and (h).

110. The computer implemented method of claim 109, wherein the at least one target peptide prioritized at step (i) is derived from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, or SLC3A2.

111. A method for identifying one or more target peptides for treatment of a disease, the method comprising:

a) identifying one or more alternatively spliced isoforms that are differentially expressed in RNA-seq samples derived from diseased cells to generate a working list of isoforms, wherein the diseased cells are associated with the disease for treatment, the one or more identified alternatively spliced isoforms are expressed at a higher frequency in a subset of the RNA-seq samples associated with a splicing factor gene mutation than in a subset of the RNA-seq samples that are not associated with the splicing factor gene mutation, and a targetable peptide sequence is associated with each identified isoform such that the targetable peptide sequence distinguishes the identified isoform from other variants over which the identified isoform was differentially expressed;

b) performing one or more filtering steps on the working list of isoforms, wherein the working list of isoforms is updated after performance of each filtering step and the one or more filtering steps comprise at least one of the following:

1) filtering the working list of isoforms to filter out any isoforms that were expressed in the subset of RNA-seq samples not associated with the splicing factor gene mutation above a

control sample threshold frequency, optionally wherein the control sample threshold frequency is 0%;

2) filtering the working list of isoforms to filter out any isoforms that are expressed in normal tissue above a normal tissue expression level threshold, optionally wherein the normal tissue expression level threshold is no detectable expression;

3) filtering the working list of isoforms to filter out any isoforms that are not associated with a gene that encodes a protein that is expressed on the cell surface;

4) filtering the working list of isoforms to filter out any isoforms that are not associated with a targetable peptide sequence that is aligned with at least a portion of an extracellular domain of a protein expressed on the cell surface;

5) filtering the working list of isoforms to filter out any isoforms that i) result in the loss of a transmembrane domain and/or an extracellular domain, and/or ii) result in the truncation of an extracellular domain; and

6) filtering the working list of isoforms to retain only isoforms for which the targetable peptide sequence introduces i) an insertion of an isoform-specific amino acid sequence into an extracellular domain of a protein expressed on the cell surface and/or ii) a deletion of one or more amino acids in the extracellular domain, wherein the insertion and/or the deletion distinguishes the identified isoform from other variants over which the identified isoform was differentially expressed in step (a); and

c) identifying the one or more target peptides for treatment of the disease from the working list of isoforms following performing the one or more filtering steps in step (b).

112. The method of claim 111, wherein:

(i) the method comprises filtering steps (b)(1) through (b)(2), optionally performed in that order;

(ii) the method comprises filtering steps (b)(1) through (b)(4), optionally performed in that order;

(iii) the method comprises filtering steps (b)(1) through (b)(5), optionally performed in that order; or

(iv) the method comprises filtering steps (b)(1) through (b)(6), optionally performed in that order.

113. The method of any one of claims 111- 112, wherein the isoform-specific amino acid sequence comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more contiguous amino acids.

114. The method of any one of claims 111-113, wherein the deletion in step (b)(6) is of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more contiguous amino acids.

115. The method of any one of claims 111-114, wherein the deletion in step (b)(6) is of no more than 10, 15, 20, 25, or 30 contiguous amino acids, optionally wherein the deletion does not occur at an N-terminal and/or a C-terminal portion of the extracellular domain.

116. The method of any one of claims 111-115, wherein the isoform-specific amino acid sequence and/or deletion results from a frame shift relative to another isoform over which the identified isoform was differentially expressed in step (a).

117. The method of any one of claims 111-116, wherein performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only identified isoforms which are expressed in the subset of RNA-seq samples associated with the splicing factor gene mutation above a mutation sample threshold frequency.

118. The method of any one of claims 111-117, wherein performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only identified isoforms which are expressed in the subset of RNA-seq samples associated with the splicing factor gene mutation above a mutation sample expression level threshold, optionally wherein expression is measured by raw read counts or normalized expression values.

119. The method of any one of claims 111-118, wherein performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only identified isoforms which are expressed in one or more additional diseases above a secondary

disease threshold frequency, wherein the one or more additional diseases are different from the disease which the diseased cells were associated with in step (a), optionally wherein the disease for treatment and one or more additional diseases are different types of cancer.

120. The method of any one of claims 111-119, wherein performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only identified isoforms which are expressed in one or more additional diseases above a secondary disease expression level threshold, wherein the one or more additional diseases are different from the disease which the diseased cells were associated with in step (a), optionally wherein the disease for treatment and one or more additional diseases are different types of cancer.

121. The method of any one of claims 111-120, wherein identifying the one or more alternatively spliced isoforms comprises aligning RNA-seq reads to a Genome Reference Consortium (GRC) Human Build and calling exon-exon junctions within the reads.

122. The method of any one of claims 111-121, wherein the splicing factor gene mutation has a prevalence among the RNA-seq samples derived from diseased cells above a mutation prevalence threshold frequency, optionally wherein the method further comprises selecting the splicing factor gene mutation by querying a database of RNA-seq samples.

123. The method of any one of claims 111-122, wherein the RNA-seq samples exclude any samples associated with a copy number variation of the splicing factor gene, optionally wherein step (a) comprises filtering out any RNA-seq samples that are associated with a copy number variation of the splicing factor gene.

124. The method of any one of claims 111-123, wherein performing the one or more filtering steps in step (b) further comprises filtering the working list of isoforms to retain only isoforms that have been experimentally detected via mass spectrometry, optionally comprising performing the mass spectrometry.

125. The method of any one of claims 111-124, further comprising generating a working list of peptide sequences for MHC binding, wherein each peptide sequence of the working list of peptide sequences for MHC binding (i) is a length suitable for binding an MHC molecule, (ii) is derivable from a targetable peptide sequence, and (iii) is not derivable from any other variants over which the corresponding identified isoform was differentially expressed in step (a), optionally wherein the working list of isoforms is filtered to retain only isoforms corresponding to a peptide sequence from the working list of peptide sequences for MHC binding.

126. The method of any one of claims 111-124, wherein generating a working list of isoforms in step (a) comprises generating a working list of peptide sequences for MHC binding, wherein each peptide sequence of the working list of peptide sequences for MHC binding (i) is a length suitable for binding an MHC molecule, (ii) is derivable from a targetable peptide sequence, and (iii) is not derivable from any other variants over which the corresponding identified isoform is differentially expressed.

127. The method of claim 125 or 126, wherein the MHC molecule is an MHC class I molecule, optionally wherein the length suitable for binding is 8-12, 8-10, or 9 amino acids in length.

128. The method of claim 125 or 126, wherein the MHC molecule is an MHC class II molecule, optionally wherein the length suitable for binding is 13-17 amino acids in length.

129. The method of any one of claims 125-128, wherein performing the one or more filtering steps in step (b) further comprises filtering the list of working peptide sequences for MHC binding to retain only peptide sequences which are predicted to bind an MHC molecule.

130. The method of any one of claims 125-129, wherein performing the one or more filtering steps in step (b) further comprises filtering the working list of peptide sequences for MHC binding to retain only peptide sequences that have been experimentally detected via mass

spectrometry as bound to an MHC molecule, optionally comprising performing the mass spectrometry.

131. The method of any one of claims 111-130, wherein the disease for treatment is a cancer.

132. The method of any one of claims 111-131, wherein the disease for treatment is uveal melanoma, hematological malignancy, breast cancer, skin melanoma, renal cell carcinoma, pulmonary adenocarcinoma, hepatocarcinoma, pancreatic carcinoma, or an endometrial cancer.

133. The method of any one of claims 111-132, wherein the splicing factor gene mutation is a mutation of splicing factor 3b subunit 1 (SF3B1).

134. The method of any one of claims 111-133, wherein the one or more identified isoforms in step (a) are associated with a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, or UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

135. The method of any one of claims 111-134, wherein the one or more identified target peptides in step (c) are associated with a gene selected from ARIH1, CCT2, CD34, COG1, DLST, DPH5, DVL2, ERGIC3, HIGD2A, IL17RC, MCM3AP, NET1, NONO, OXA1L, PLXNB1, SF3A2, STIM1, STK24, SYVN1, VPS33B, VPS51, ZDHHC16, USP39, BRD9, ERFE, UBA1, BACE1, BCAM, DGCR2, GYPC, CD81, GPNMB, MEGF8, NCSTN, PLXNB3, APP, CLSTN1, SDC3, PTK7, FXYD5, NRG3, CA14, MMP14, SLC2A11, GPR143, TNFRSF14, ENO1, BSG, TFRC, TMEM179B, GPX1, HTR2B, and SLC3A2.

136. The method of any one of claims 111-135, wherein the one or more identified isoforms in step (a) and/or the one or more identified target peptides in step (c) comprise an amino acid sequence of any one of SEQ ID NOs: 1-37, 114-136, 160-164, 169.

137. The method of any one of claims 111-136, wherein step (a) and/or one or more of the performed filtering steps in step (b), optionally each of the filtering steps performed in step (b), is computer implemented.

138. The method of any one of claims 111-137, further comprising generating a peptide that comprises the targetable peptide sequence of one of the target peptides and/or generating a nucleic acid that encodes the same.

139. A system comprising one or more processors and a memory storing processor-executable instructions thereon that, when executed by the one or more processors, cause the one or more processors to perform the method of any of claims 111-138.