

PEPTIDE-MHC COMPACKTS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/651,639, filed on April 2, 2018, which is hereby incorporated by reference in its entirety for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted via EFS-Web and is hereby incorporated herein by reference in its entirety. Said ASCII copy, created on Month XX, 20XX, is named XXXXXUS_sequencelisting.txt, and is X,XXX,XXX bytes in size.

BACKGROUND

[0003] T cells are the primary mediators of adaptive immunity. Directed by the specificity of each T cell's unique T cell receptor (TCR), T cells regulate autoimmunity, help activate B cells and innate effectors, and directly kill infected and cancerous cells in a precisely targeted manner. Each TCR recognizes a ligand presented by a major histocompatibility complex (MHC) molecule on target cells. Identification of relevant peptide-MHC complex ligands plays a role in understanding immune responses to tumors and pathogens. MHC complex ligands are also valuable for understanding responses to self and dietary antigens. This understanding enables clinically beneficial immunotherapies (e.g. TCR gene transfer and vaccines) that initiate, amplify, or attenuate immune responses to target antigens.

[0004] Mutated 'neoepitopes' are important targets of endogenous and engineered immune responses to cancer. Neoepitope-reactive TILs are present in the endogenous repertoire and regress tumors upon adoptive transfer. Likewise, tumor mutational burden predicts the clinical effectiveness of CTLA-4 or PD-1 blockade, suggesting these checkpoint inhibition strategies affect tumor regression by unleashing neoepitope-reactive T cells. Because neoepitopes result from somatic mutation in tumor cells, they are not generally presented by thymic epithelial cells to induce central tolerance. Thus, T cell responses directed at these neoepitopes are tumor-specific, likely high-affinity, and patient-specific (i.e. private). From a clinical standpoint, this presents an opportunity and a challenge: neoepitopes are excellent targets for immunotherapy, but TCR isolation methods should be sufficiently high-throughput to enable therapeutic application on a clinically-useful scale.

[0005] There is an unmet need for rapid and robust TCR ligand discovery technologies for both basic and translational research. Peptide-MHC multimers enable sorting of T cells

according to the antigenic specificity of their TCRs, an important step in isolating tumor-specific TCRs for gene therapy. A typical current peptide-MHC production protocol begins with solid-phase synthesis of the peptide ligand(s) of interest. In parallel, the universal β_2 -microglobulin and relevant MHC class I molecules are heterologously expressed in *E. coli*, yielding misfolded inclusion bodies. Each peptide is added to a refolding reaction containing β_2 M and the relevant MHC I molecule. Finally, the portion of ternary complex that refolds correctly can be purified and formulated for use in Peptide-MHC multimer production. To facilitate parallel production of a particular MHC molecule with many different peptide ligands, Schumacher and colleagues devised a photocleavable peptide that binds a particular MHC molecule as a conditional ligand. A single refolding reaction is performed to generate that MHC molecule bound to its conditional ligand. Upon exposure to UV light, the conditional ligand is cleaved and exchanged for a desired peptide present in excess. Many such exchange reactions can be performed in parallel, enabling the construction of a pMHC library for that particular MHC allele. Even so, this state-of-the-art technology has challenging limitations. First, the production, purification, and refolding of MHC molecules expressed in *E. coli* inclusion bodies is laborious and produces low yields of properly folded peptide-MHC complex. Second, the turnaround time (weeks) for commercial peptide synthesis is at odds with timescales optimal in the context of personalized on-demand TCR gene therapies directed at patient-specific neoepitopes. Third, many predicted ligands cannot be used to screen T cells through this approach because the biophysical properties (e.g. hydrophobicity) of the peptide precludes its synthesis or exchange. Fourth, exchange efficiency is generally poor (<50% exchange efficiency for a majority of predicted HLA-binding peptides). The resulting mixture of properly folded exchanged MHC and misfolded unliganded MHC results in multimer staining with low signal to noise, an issue that is exacerbated when screening T cells with a multiplexed pool of peptide-MHC reagents. Fifth, the design and validation of conditional ligands for each new MHC allele is a laborious and non-robust undertaking. As the MHC locus is the most multi-allelic locus in the human genome, this is a major hindrance to implementing neoepitope-targeted gene therapies across patients of diverse MHC haplotypes. Together, these limitations underscore the need for novel technologies in this field. Disclosed herein are various compositions and processes for producing peptide-MHC multimers that address these limitations.

SUMMARY

[0006] The methods and compositions described herein enable rapid identification of antigens targeted by an immune response and isolation of the T cells and thus the TCR sequences mediating that response. This rapid identification is a goal of research and drug development in the fields of infectious disease, tumor immunology, autoimmunity, and immunotherapy. The compositions may be applied to identify and to isolate antigen-specific T cells and their encoded MHC I-restricted TCRs for drug development. The identification and characterization of antigen-specific T cells may also be applied in the diagnosis, monitoring and prognosis of immune responses and disease in the context of (pre- and post-) therapeutic treatment. The compACT approach may also be applied to identify the targets of MHC II-restricted T cell responses. Such responses are central to autoimmune disease, and are emerging as important components of cancer immunotherapy. In addition, the efficiency of compACT protein secretion from mammalian cells expression data generated in the method described herein is potentially useful as a surrogate metric of proper protein folding and thus the affinity of each neoepitope ligand for its cognate MHC molecule. This technology may then afford the unique ability to refine the MHC binding algorithms for less-well studied MHC alleles, improving other targeted immunotherapies such as neoepitope vaccines. As such, the compositions disclosed herein will have broad application to human health and clear commercial potential.

[0007] In one aspect, disclosed herein is a polynucleotide molecule comprising, in a 5' to 3' orientation, (i) a first universal target sequence, (ii) a nucleotide sequence encoding an antigenic peptide, (iii) a second universal target sequence that is distinct from the first universal target sequence, (iv) a Beta 2 Microglobulin (β 2M) sequence, and (v) an Multiple Histocompatibility Complex (MHC) allele sequence. In one embodiment, the polynucleotide molecule comprises, in a 5' to 3' orientation, (i) a promoter sequence, (ii) the first universal target sequence comprising the sequence shown in SEQ ID NO: 3, (iii) the nucleotide sequence encoding an antigenic peptide, wherein the antigenic peptide is a tumor neoantigen, (iv) the second universal target sequence comprising the sequence shown in SEQ ID NO: 4, (v) the β 2M sequence comprising the sequence shown in SEQ ID NO: 106, (vi) the MHC allele sequence comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOs: 109-174, (vii) a first affinity tag sequence comprising the sequence shown in SEQ ID NO: 29, (viii) a protease cleavage site sequence comprising the sequence shown in SEQ ID NO: 31, and (ix) a second affinity tag sequence comprising the sequence shown in SEQ ID NO: 35.

[0008] In one embodiment, the antigenic peptide is selected from the group consisting of a tumor antigen, a neoantigen, a tumor neoantigen, a viral antigen, bacterial antigen, phosphoantigen, and a microbial antigen. In another embodiment, the antigenic peptide is a neoantigen.

[0009] In some embodiments a neoantigen is selected by analyzing tumor sequencing data from a subject to identify one or more somatic mutations. In one embodiment, the analyzing is performed using an in silico predictive algorithm. In another embodiment, the predictive algorithm further comprises an MHC binding algorithm to predict binding between the neoantigen and an MHC allele.

[0010] In some embodiments, the MHC allele is a mammalian MHC allele. In some embodiments, the MHC allele is a human MHC allele. In some embodiments, the MHC allele is a class I HLA allele. In other embodiments, the HLA allele comprises an HLA-A, HLA-B, or HLA-C allele. The HLA allele is an HLA allele of the subject.

[0011] In some embodiments, the HLA allele is selected from the group consisting of: HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*11:01, HLA-A*23:01, HLA-A*30:01, HLA-A*33:03, HLA-A*25:01, HLA-A*26:01, HLA-A*29:02, HLA-A*68:02, HLA-B*07:02, HLA-B*14:02, HLA-B*18:01, HLA-B*27:02, HLA-B*39:01, HLA-B*40:01, HLA-B*44:02, HLA-B*46:01, HLA-B*50:01, HLA-B*57:01, HLA-B*58:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:03, HLA-B*35:01, HLA-B*40:02, HLA-B*42:01, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*13:02, HLA-B*15:07, HLA-B*27:05, HLA-B*35:03, HLA-B*37:01, HLA-B*38:01, HLA-B*41:02, HLA-B*44:05, HLA-B*49:01, HLA-B*52:01, HLA-B*55:01, HLA-C*02:02, HLA-C*03:04, HLA-C*05:01, HLA-C*07:01, HLA-C*01:02, HLA-C*04:01, HLA-C*06:02, HLA-C*07:02, HLA-C*16:01, HLA-C*03:03, HLA-C*07:04, HLA-C*08:01, HLA-C*08:02, HLA-C*12:02, HLA-C*12:03, HLA-C*14:02, HLA-C*15:02, and HLA-C*17:01.

[0012] In some embodiments, the HLA allele comprises a sequence selected from the group consisting of SEQ ID NOs: 109-174.

[0013] In some embodiments, the β 2M allele is a mammalian β 2M allele. In some embodiments, the β 2M allele is a human β 2M allele. In some embodiments, the β 2M allele comprises the sequence shown in SEQ ID NO: 106.

[0014] In some embodiments, the nucleotide sequence encoding an antigenic peptide is between 20-60, between 20-30, between 25-35, between 20-45, between 30-45, between 40-

60, or between 45-60 nucleotides in length. In one embodiment, the nucleotide sequence encoding an antigenic peptide is between 20-30 nucleotides in length.

[0015] In some embodiments, the first universal target sequence is between 4-50, between 4-15, between 15-40, between 15-35, between 15-30, between 20-40, between 25-40, or between 30-40 nucleotides in length. In one embodiment, the first universal target sequence is between 25-35 nucleotides in length. In another embodiment, the first universal target sequence is at least about 15 nucleotides in length. In another embodiment, wherein the first universal target sequence is between 4-6 nucleotides in length. In some embodiments, the second universal target sequence is between 4-50, between 4-15, between 15-40, between 15-35, between 15-30, between 20-40, between 25-40, or between 30-40 nucleotides in length. In one embodiment, the second universal target sequence is between 25-35 nucleotides in length. In another embodiment, the second universal target sequence is at least about 15 nucleotides in length. In another embodiment, the second universal target sequence is between 4-6 nucleotides in length.

[0016] In some embodiments, the first and second universal target sequences comprise polymerase chain reaction (PCR) primer target sequences. In other embodiments, the first and second universal target sequences comprise restriction enzyme cleavage sites.

[0017] In some embodiments, the sequence of the first universal target comprises the sequence shown in SEQ ID NO: 3. In some embodiments, the sequence of the second universal target comprises the sequence shown in SEQ ID NO: 4.

[0018] In some embodiments, the first universal target sequence further comprises a signal sequence. In some embodiments, the signal sequence encodes a signal sequence comprising a Human Growth Hormone signal sequence, a hIG1 Kappa light chain signal sequence, a Beta 2 microglobulin, signal sequence, or an IL2 signal sequence. In some embodiments, the signal sequence comprises a sequence comprising the sequence shown in SEQ ID NOs: 1, 24, 26, or 28. In one embodiment, the signal sequence encodes Human Growth Hormone (HGH). In some embodiments, the signal sequence comprises the sequence shown in SEQ ID NO: 1. In one embodiment, the signal sequence is between 40-90, 40-60, 45-70, 50-80, 60-90, 55-70, 60-80, or 70-80 nucleotides in length.

[0019] In some embodiments, the 3' end of the polynucleotide sequence further comprises, in a 5' to 3' orientation, a purification cluster sequence comprising (i) a first affinity tag sequence, (ii) a protease cleavage site sequence, and (iii) a second affinity tag sequence. In some embodiments, the first and second affinity tags are selected from the group consisting of: AviTag, streptavidin-tag, polyhistidine (His6)-tag, FLAG-tag, HA-tag, and Myc-tag. In

some embodiments, the first affinity tags comprises a sequence comprising the sequence shown in SEQ ID NO: 29, and the second affinity tag comprises a sequence comprising the sequence shown in SEQ ID NO: 33.

[0020] In some embodiments, the protease cleavage site sequence is selected from the group consisting of: a TEV cleavage site sequence, a thrombin cleavage site sequence, a Factor Xa cleavage site sequence, an enteropeptidase cleavage site sequence, and a rhinovirus 3C protease cleavage site sequence. In one embodiment, the protease cleavage site sequence comprises a TEV cleavage site sequence shown in SEQ ID NO: 31. In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for a TEV cleavage site, and the second affinity tag encodes for a His6 peptide. In one embodiment, the first affinity tag sequence comprises the sequence shown in SEQ ID NO: 29, the protease cleavage site comprises the sequence shown in SEQ ID NO: 31, and the second affinity tag comprises the sequence shown in SEQ ID NO: 33.

[0021] In some embodiments, the second universal target sequence further comprises a first linker sequence. In one embodiment, the first linker sequence comprises the sequence shown in SEQ ID NO: 10, 14, 16, or 18.

[0022] In some embodiments, the polynucleotide further comprises a second linker sequence between the β 2M sequence and the MHC allele sequence. In one embodiment, the second linker sequence comprises the sequence shown in SEQ ID NO: 10 or 20.

[0023] In some embodiments, the polynucleotide further comprises a third linker sequence between the MHC allele sequence and the first affinity tag. In one embodiment, the third linker sequence comprises the sequence shown in SEQ ID NO: 12 or 22.

[0024] The 5' end of the polynucleotide sequence may further comprise a promoter sequence linked to the 5' end of the first universal target. In some embodiments, the promoter sequence is a CMV, an EF1 α , or an SV40 promoter. In one embodiment, the promoter sequence is a CMV promoter. In one embodiment, the promoter sequence is a CMV promoter sequence.

[0025] The 3' end of the polynucleotide sequence may further comprise a polyA sequence. In some embodiments, the polyA sequence is an SV40, hGH, BHG, or rbGlob polyA sequence. In one embodiment, the polyA sequence comprises a bGH polyA sequence as shown in SEQ ID NO: 179.

[0026] In one embodiment, the polynucleotide molecule comprises, in a 5' to 3' orientation, (i) a promoter sequence, (ii) a first universal target sequence, (iii) a nucleotide sequence encoding an antigenic peptide, (iv) a second universal target sequence that is distinct from the first universal target sequence, (v) a β 2M sequence, (vi) an MHC allele sequence, (vi) a first

affinity tag sequence, (vii) a protease cleavage site sequence, and (viii) a second affinity tag sequence.

[0027] In another aspect, described herein is an expression construct comprising the polynucleotide molecule as disclosed herein. In one embodiment, the expression construct comprises a plasmid or a viral vector.

[0028] In another aspect, described herein is a host cell comprising the polynucleotide molecule or the expression construct as previously described. In one embodiment, the polynucleotide is integrated into the cell genome. In another embodiment, the polynucleotide is extrachromosomal. In some embodiments, the host cell is a mammalian cell. In some embodiments, the host cell is a human cell. In one embodiment, the cell is a stem cell, a tumor cell, an immortalized cell, or a fetal cell. In some embodiments, the host cell is a prokaryotic cell. In some embodiments, the cell is an *Escherichia coli* cell. In some embodiments, the cell expresses a BirA protein or fragment thereof.

[0029] In yet another aspect, described herein is a library comprising the polynucleotide molecule or expression construct as described herein, wherein the library comprises greater than or equal to two distinct polynucleotide molecules, wherein each distinct polynucleotide molecule comprises (i) the first universal sequence, (ii) the nucleotide sequence encoding a antigenic peptide, wherein the nucleotide sequence is not the same for each of the greater than or equal to two polynucleotide molecules (iii) the second universal target sequence, (iv) the β 2M sequence, and (v) the MHC allele sequence. In one embodiment, the MHC allele sequence is not the same for each of the greater than or equal to two polynucleotide molecules. In another embodiment, the library comprises 20 to 500 distinct polynucleotide molecules. In some embodiments, the library comprises at least 66 distinct polynucleotide molecules.

[0030] In some embodiments, the library comprises at least the HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*11:01, HLA-A*23:01, HLA-A*30:01, HLA-A*33:03, HLA-A*25:01, HLA-A*26:01, HLA-A*29:02, HLA-A*68:02, HLA-B*07:02, HLA-B*14:02, HLA-B*18:01, HLA-B*27:02, HLA-B*39:01, HLA-B*40:01, HLA-B*44:02, HLA-B*46:01, HLA-B*50:01, HLA-B*57:01, HLA-B*58:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:03, HLA-B*35:01, HLA-B*40:02, HLA-B*42:01, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*13:02, HLA-B*15:07, HLA-B*27:05, HLA-B*35:03, HLA-B*37:01, HLA-B*38:01, HLA-B*41:02, HLA-B*44:05, HLA-B*49:01, HLA-B*52:01, HLA-B*55:01, HLA-C*02:02, HLA-C*03:04, HLA-C*05:01,

HLA-C*07:01, HLA-C*01:02, HLA-C*04:01, HLA-C*06:02, HLA-C*07:02, HLA-C*16:01, HLA-C*03:03, HLA-C*07:04, HLA-C*08:01, HLA-C*08:02, HLA-C*12:02, HLA-C*12:03, HLA-C*14:02, HLA-C*15:02, and HLA-C*17:01 alleles. In some embodiments, the library comprises at least the sequences shown in SEQ ID NOs: 109-174.

[0031] In another aspect, described herein is a polypeptide comprising, in an amino to carboxyl terminus orientation, (i) a first universal target peptide, (ii) an antigenic peptide, (iii) a second universal target peptide that is distinct from the first universal target peptide, (iv) a β 2M peptide, and (v) an MHC peptide. In some embodiments, the antigenic peptide is selected from the group consisting of a tumor antigen, a neoantigen, a tumor neoantigen, a viral antigen, bacterial antigen, phosphoantigen, and a microbial antigen. In one embodiment, the antigenic peptide is a neoantigen.

[0032] In some embodiments, the polypeptide comprises in a 5' to 3' orientation, (i) a promoter peptide comprising the sequence shown in SEQ ID NO: 2, (ii) the first universal target peptide, (iii) the nucleotide sequence encoding an antigenic peptide, wherein the antigenic peptide is a tumor neoantigen, (iv) the second universal target peptide comprising the sequence shown in SEQ ID NO: 15 or 17, (v) the β 2M peptide comprising the sequence shown in SEQ ID NO: 105, (vi) the MHC peptide comprising a sequence shown in SEQ ID NOs: 38-103, (vii) a first affinity peptide comprising the sequence shown in SEQ ID NO: 30, (viii) a protease cleavage site peptide comprising the sequence shown in SEQ ID NO: 32, and (ix) a second affinity peptide comprising the sequence shown in SEQ ID NO: 36.

[0033] In some embodiments the neoantigen is selected by analyzing tumor sequencing data from a subject to identify one or more somatic mutations. In one embodiment, the analyzing is performed using an in silico predictive algorithm. In another embodiment, the predictive algorithm further comprises an MHC binding algorithm to predict binding between the neoantigen and an MHC peptide.

[0034] In one embodiment, the antigenic peptide is 7-15 amino acids, 7-10, 8-9, 7, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids in length.

[0035] In some embodiments, the MHC peptide is a mammalian MHC peptide. In some embodiments, the MHC peptide is a human MHC peptide. In some embodiments, the MHC peptide is a class I HLA peptide. In other embodiments, the HLA peptide comprises an HLA-A, HLA-B, or HLA-C peptide. In one embodiment, the HLA peptide is an HLA peptide of the subject. In one embodiment, the HLA peptide comprises a Y84A or a Y84C mutation.

[0036] In some embodiments, the β 2M peptide is a mammalian β 2M peptide. In some embodiments, the β 2M peptide is a human β 2M peptide. In some embodiments, the β 2M

peptide comprises the sequence shown in SEQ ID NO: 105. In another embodiment, the β 2M peptide comprises an S88C mutation. In some embodiments, the β 2M peptide comprises the sequence shown in SEQ ID NO: 107, comprising an S88C mutation at amino acid 88.

[0037] In some embodiments, the HLA peptide comprises HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*11:01, HLA-A*23:01, HLA-A*30:01, HLA-A*33:03, HLA-A*25:01, HLA-A*26:01, HLA-A*29:02, HLA-A*68:02, HLA-B*07:02, HLA-B*14:02, HLA-B*18:01, HLA-B*27:02, HLA-B*39:01, HLA-B*40:01, HLA-B*44:02, HLA-B*46:01, HLA-B*50:01, HLA-B*57:01, HLA-B*58:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:03, HLA-B*35:01, HLA-B*40:02, HLA-B*42:01, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*13:02, HLA-B*15:07, HLA-B*27:05, HLA-B*35:03, HLA-B*37:01, HLA-B*38:01, HLA-B*41:02, HLA-B*44:05, HLA-B*49:01, HLA-B*52:01, HLA-B*55:01, HLA-C*02:02, HLA-C*03:04, HLA-C*05:01, HLA-C*07:01, HLA-C*01:02, HLA-C*04:01, HLA-C*06:02, HLA-C*07:02, HLA-C*16:01, HLA-C*03:03, HLA-C*07:04, HLA-C*08:01, HLA-C*08:02, HLA-C*12:02, HLA-C*12:03, HLA-C*14:02, HLA-C*15:02, or HLA-C*17:01. In some embodiments, the HLA peptide comprises a sequence selected from the group consisting of SEQ ID NOs: 38-103.

[0038] In some embodiments, the first universal target peptide further comprises a signal peptide. In one embodiment, the signal sequence is between 15-45, between 15-30, between 20-45, between 20-30, or between 30-45 amino acids in length. In one embodiment, the signal peptide comprises Human Growth Hormone. In some embodiments, the signal peptide comprising a Human Growth Hormone signal peptide, a hIG1 Kappa light chain signal peptide, a Beta 2 microglobulin signal peptide, or an IL2 signal peptide. In some embodiments, the signal peptide comprises a sequence comprising the sequence shown in SEQ ID NOs: 2, 23, 25, or 27. In some embodiments, the signal peptide comprises the Human Growth Hormone (HGH) signal peptide sequence shown in SEQ ID NO: 2. In one embodiment, the second universal target peptide comprises the sequence GGGGSGGGGSGGGGS. In one embodiment, the second universal target peptide comprising the sequence shown in SEQ ID NO: 15 or 17.

[0039] In some embodiments, the carboxyl terminus of the polypeptide may further comprise, in an amino to carboxylic terminus orientation, a purification cluster comprising (i) a first affinity peptide, (ii) a protease cleavage site, and (iii) a second affinity peptide. In some embodiments, the first and second affinity peptides are selected from the group consisting of AviTag, strep-tag, polyhistidine (His6)-tag, FLAG-tag, HA-tag, and/or Myc-tag. In some

embodiments, the first affinity peptide comprises a sequence comprising SEQ ID NO: 30 and the second affinity peptide comprises a sequence comprising SEQ ID NO: 34.

[0040] In some embodiments, the protease cleavage site is a TEV cleavage site, a thrombin cleavage site, a Factor Xa cleavage site, an enteropeptidase cleavage site, or a rhinovirus 3C protease cleavage site. In some embodiments, the purification cluster comprises an AviTag epitope, a TEV cleavage site, and a His6 epitope. In some embodiments, the protease cleavage site comprises the TEV cleavage site sequence shown in SEQ ID NO: 32. In some embodiments, the purification cluster comprises the AviTag peptide sequence shown in SEQ ID NO: 30, the TEV cleavage site sequence shown in SEQ ID NO: 32, and the His6 peptide sequence shown in SEQ ID NOs: 34 or 36. In some embodiments, the purification cluster comprises two or more copies of the His6 peptide sequence shown in SEQ ID NO: 34. In some embodiments, the purification cluster comprises the sequence shown in SEQ ID NO: 37.

[0041] In some embodiments, the second universal target peptide further comprises a linker comprising the sequence shown in SEQ ID NO: 9, 11, 15, or 17.

[0042] In some embodiments, the polypeptide comprises a second linker sequence between the β 2M sequence and the MHC allele sequence. In some embodiments, the second linker sequence comprises the sequence shown in SEQ ID NO: 9 or 19. In some embodiments, the polypeptide further comprising a third linker sequence between the MHC allele sequence and the first affinity tag. In some embodiments, the third linker sequence comprises the sequence shown in SEQ ID NO: 13 or 21. In some embodiments, the polypeptide is biotinylated.

[0043] In one embodiment, the polypeptide comprises in an amino to carboxyl terminus orientation, (i) the first universal target peptide, (ii) the antigenic peptide, (iii) the second universal target peptide, (iv) the β 2M peptide, (v) the MHC peptide, (vi) the first affinity tag peptide, (vii) the protease cleavage site, and (viii) the second affinity tag peptide. In some embodiments, the polypeptide is biotinylated.

[0044] In another aspect, disclosed herein is a library comprises greater than or equal to two distinct polypeptide molecules, wherein the distinct polypeptide molecule comprises (i) the first universal target peptide, (ii) the antigenic peptide, wherein the antigenic peptide is not the same for each of the greater than or equal to two polypeptide molecules, (iii) the second universal target peptide that is distinct from the first universal target peptide, (iv) the β 2M peptide, and (v) the MHC peptide. In one embodiment, the MHC peptide is not the same for each of the greater than or equal to two polypeptide molecules. In another embodiment, the

library comprises 20 to 500 distinct polypeptide molecules. In some embodiments, the library comprises at least 66 distinct polypeptides.

[0045] In some embodiments, the library comprises HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*11:01, HLA-A*23:01, HLA-A*30:01, HLA-A*33:03, HLA-A*25:01, HLA-A*26:01, HLA-A*29:02, HLA-A*68:02, HLA-B*07:02, HLA-B*14:02, HLA-B*18:01, HLA-B*27:02, HLA-B*39:01, HLA-B*40:01, HLA-B*44:02, HLA-B*46:01, HLA-B*50:01, HLA-B*57:01, HLA-B*58:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:03, HLA-B*35:01, HLA-B*40:02, HLA-B*42:01, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*13:02, HLA-B*15:07, HLA-B*27:05, HLA-B*35:03, HLA-B*37:01, HLA-B*38:01, HLA-B*41:02, HLA-B*44:05, HLA-B*49:01, HLA-B*52:01, HLA-B*55:01, HLA-C*02:02, HLA-C*03:04, HLA-C*05:01, HLA-C*07:01, HLA-C*01:02, HLA-C*04:01, HLA-C*06:02, HLA-C*07:02, HLA-C*16:01, HLA-C*03:03, HLA-C*07:04, HLA-C*08:01, HLA-C*08:02, HLA-C*12:02, HLA-C*12:03, HLA-C*14:02, HLA-C*15:02, and HLA-C*17:01 polypeptides.

[0046] In some embodiments, the polypeptide is attached to a particle, wherein the particle is a surface, a nanoparticle, a bead, or a polymer. In one embodiment, the polypeptide is attached via a linker to the particle. In one embodiment, the particle is a nanoparticle and the nanoparticle is a magnetic nanoparticle or a polystyrene nanoparticle. In one embodiment, the magnetic nanoparticle comprises magnetic iron oxide. In another embodiment, the bead is an agarose bead or a sepharose bead. In some embodiments, the polypeptide attached to the particle further comprises a fluorophore. In one embodiment, the fluorophore is attached, with or without a linker, to the particle.

[0047] In some embodiments, a library of polypeptides is attached to particles wherein the library comprises greater than or equal to two distinct polypeptide molecules, wherein the antigenic peptide is not the same for each of the greater than or equal to two polypeptide molecules, and wherein each distinct polypeptide is attached to a particle. In one embodiment, the MHC peptide is not the same for each of the greater than or equal to two polypeptide molecules. In one embodiment, the library further comprises a unique defined barcode sequence operably associated with the identity of each distinct polypeptide. In another embodiment, the library comprises 20 to 500 distinct polypeptides. In another embodiment, the library comprises at least 66 distinct polypeptides.

[0048] In another aspect, described herein is a kit comprising the composition of any of the above embodiments and instructions for use.

[0049] In another aspect, disclosed herein is a method of manufacturing a polynucleotide molecule comprising the steps of: (a) obtaining a first polynucleotide sequence comprising, in a 5' to 3' orientation, (i) a first universal target sequence comprising a restriction site, (ii) a second universal target sequence comprising a restriction site that is not the same as the first universal target sequence, (iii) a β 2M sequence, and (iv) an MHC allele sequence; (b) obtaining a second polynucleotide comprising, in a 5' to 3' orientation (i) a portion of the first universal target sequence, (i) a sequence encoding an antigenic peptide, and (iii) a portion of the second universal target sequence; (c) obtaining a third polynucleotide comprising, in a 5' to 3' orientation, (i) the reverse complement of a portion of the second universal target sequence, (ii) the reverse complement of the sequence encoding an antigenic peptide, and (iii) the reverse complement of a portion of the first universal target sequence; (d) mixing the second and third polynucleotides together such that the complementary sequences anneal; (e) performing a restriction digest on the first polynucleotide sequence with at least one restriction enzyme; and (f) ligating the first and second polynucleotides together by mixing the digested first polynucleotide and annealed second and third polynucleotides together with DNA ligation reaction reagents.

[0050] In some embodiments, the method further comprises phosphorylating the 5' nucleotide of the second and third polynucleotides after the complementary sequences anneal, wherein the phosphorylating comprises incubating the first or second and third polynucleotides with an enzyme. In some embodiments, the enzyme is T4 kinase. In some embodiments, the first polynucleotide is an expression construct. In some embodiments, comprising inserting the manufactured polynucleotide into an expression construct.

[0051] In some embodiments, the first polynucleotide further comprises a promoter sequence. In one embodiment, the promoter sequence is a CMV, an EF1 α , or an SV40 promoter.

[0052] In another embodiment, the 3' end of the second polynucleotide sequence further comprises, in a 5' to 3' orientation, a purification cluster sequence comprising (i) a first affinity tag sequence, (ii) a protease cleavage site sequence, and (iii) a second affinity tag sequence.

[0053] In one embodiment, the 3' end of the second polynucleotide sequence further comprises a polyA sequence.

[0054] In some embodiments, obtaining the third and fourth polynucleotide sequence steps further comprises obtaining a predictive data set of a tumor sequencing data from a subject. In one embodiment, obtaining the third and fourth polynucleotide sequence steps further

comprises predicting the sequence encoding the antigenic peptide from the tumor sequencing data. In another embodiment, obtaining the third and fourth polynucleotide sequence steps further comprises synthesizing the polynucleotide encoding the antigenic peptide based on the predicted sequence.

[0055] In some embodiments, the method further comprises manufacturing a library comprising greater than or equal to two distinct polynucleotide molecules, wherein the distinct polynucleotide molecule comprises (i) the first universal target sequence, (ii) the nucleotide sequence encoding an antigenic peptide, wherein the antigenic peptide is not the same for each of the greater than or equal to two polypeptide molecules, (iii) the second universal target sequence that is distinct from the first universal target sequence, (iv) the β 2M peptide, and (v) the MHC allele sequence. In one embodiment, the MHC allele is not the same for each of the greater than or equal to two polynucleotide molecules. In another embodiment, the library comprises 40 to 500 distinct polynucleotide molecules.

[0056] In some embodiments, the method further comprises inserting the manufactured polynucleotide into an expression construct. In one embodiment, the method further comprises manufacturing a library comprising greater than or equal to two expression constructs comprising distinct polynucleotide molecules, wherein the distinct polynucleotide molecule comprises (i) the first universal target sequence, (ii) the nucleotide sequence encoding an antigenic peptide, wherein the nucleotide sequence is not the same for each of the greater than or equal to two polynucleotide molecules, (iii) the second universal target sequence that is distinct from the first universal target sequence, (iv) the β 2M sequence, and (v) the MHC allele sequence. In one embodiment, the MHC allele is not the same for each of the greater than or equal to two polynucleotide molecules. In another embodiment, the library comprises 20 to 500 distinct expression constructs. In some embodiments, the library comprises at least 66 distinct polynucleotide molecules. In some embodiments, the MHC allele is selected from the sequences shown in the group consisting of SEQ ID NOs: 109-174.

[0057] In some embodiments, the method further comprises expressing a polypeptide molecule from the polynucleotide molecule. In some embodiments, the polynucleotide is transfected or transduced into a cell. In one embodiment, the polynucleotide is integrated into the cell genome. In another embodiment, the polynucleotide remains extrachromosomal in the cell. In some embodiments, the host cell is a mammalian cell or a human cell. In one embodiment, the cell is a stem cell, a tumor cell, an immortalized cell, or a fetal cell. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is an *Escherichia coli* cell.

[0058] In one embodiment, the method further comprises quantifying the expression of the polyprotein. In some embodiments, the expressed and non-expressed polyprotein sequences are used to refine the antigenic peptide prediction analysis step of the method. In one embodiment, the method further comprises biotinylating the polyprotein.

[0059] In some embodiments, the method comprises purifying the polyprotein. The purifying step may comprise affinity chromatography. In one embodiment, the affinity chromatography comprises immobilized metal affinity chromatography, comprising a support, a chelator and a divalent metal. The chelator for the affinity chromatography can be nitrolotriacetic acid (NTA) or iminodiacetic acid (IDA). In some embodiments the divalent metal is selected from the group consisting of: Nickel (Ni), Cobalt (Co), Copper (Cu), and Iron (Fe). In one embodiment, the chelator is NTA and the divalent metal is Ni. In other embodiments, the support is selected from the group consisting of: agarose beads, sepharose beads, and magnetic beads. In some embodiments, the method further comprises quantifying the biotinylation level of the purified polyprotein.

[0060] In some embodiments the method further comprises attaching the polypeptide to a particle, wherein the particle is a surface, a nanoparticle, a bead, or a polymer. In one embodiment, the polypeptide is attached via a linker to the particle. In one embodiment, the particle is a nanoparticle and the nanoparticle is a magnetic nanoparticle or a polystyrene nanoparticle. In one embodiment, the magnetic nanoparticle comprises magnetic iron oxide. In another embodiment, the bead is an agarose bead or a sepharose bead. In some embodiments, the polypeptide attached to the particle further comprises a fluorophore. In one embodiment, the fluorophore is attached, with or without a linker, to the particle.

[0061] In another embodiment, the method further comprises producing a library comprising greater than or equal to two distinct polyproteins attached to at least one particle, wherein the distinct polypeptide comprises (i) the first universal target peptide, (ii) the antigenic peptide, wherein the antigenic peptide is not the same for each of the greater than or equal to two polypeptide molecules, (iii) the second universal target peptide that is distinct from the first universal target peptide, (iv) the β 2M peptide, and (v) the MHC peptide. In one embodiment, the MHC peptide is not the same for each of the greater than or equal to two polypeptide molecules. In another embodiment, the library comprises 20 to 500 distinct single polyproteins. In another embodiment, the library comprises at least 66 distinct single polyproteins. In another embodiment, the MHC peptide is selected from the sequences shown in the group consisting of SEQ ID NO: 39-104.

[0062] In another aspect, disclosed herein is a method of manufacturing a polynucleotide molecule comprising the steps of: (a) obtaining a first polynucleotide sequence comprising, in a 5' to 3' orientation, (i) a first universal target sequence comprising a restriction site, (ii) a second universal target sequence comprising a restriction site that is not the same as the first universal target sequence, (iii) a β 2M sequence, and (iv) an MHC allele sequence; (b) obtaining a second polynucleotide comprising, in a 5' to 3' orientation, (i) the first universal target sequence, (ii) a sequence encoding an antigenic peptide, and (iii) the second universal target sequence; (c) performing a restriction digest on the first polynucleotide sequence with at least one restriction enzyme; (d) performing a restriction digest on the second polynucleotide sequence with at least one restriction enzyme; and (e) ligating the first and second polynucleotides together by mixing the digested first and second polynucleotides together with DNA ligation reaction reagents.

[0063] In another aspect, described herein is a method of manufacturing a polynucleotide molecule comprising the steps of: (a) obtaining a first polynucleotide sequence comprising a first universal target sequence; (b) obtaining a second polynucleotide sequence comprising, in a 5' to 3' orientation, (i) a second universal target sequence, (ii) a β 2M sequence, and (iii) an MHC allele sequence; (c) obtaining a third polynucleotide sequence comprising, in a 5' to 3' orientation, a sequence encoding an antigenic peptide and the second universal target sequence; (d) obtaining a fourth polynucleotide sequence comprising, in a 5' to 3' orientation, the reverse complement of the antigenic peptide and the reverse complement of the first universal target sequence; (e) combining the first, second, third, and fourth polynucleotides in a solution; (f) adding polymerase chain reaction (PCR) reagents; and (g) performing a PCR reaction wherein a complementary region of the second universal target sequence in the second and third polynucleotides anneal to each other and a complementary region of the first universal target sequence in the first and fourth polynucleotides anneal to each other; and wherein the annealed sequences provide priming sequences for a PCR extension and amplification reaction.

[0064] Also disclosed herein is a method for isolating an antigen specific T cell, the method comprising the steps of: (a) providing a polypeptide comprising, in an amino terminus to carboxyl terminus orientation, (i) a first universal target peptide, (ii) an antigenic peptide, (iii) a second universal target peptide that is distinct from the first universal target peptide, (iv) a β 2M peptide, and (v) an MHC peptide, wherein the polypeptide is linked to one particle; (b) providing a sample known or suspected to comprise one or more T cells; (c) contacting the polypeptide with the sample, wherein the contacting comprises providing conditions

sufficient for a single T cell to bind the polypeptide attached to the particle, and (d) isolating the single T cell associated with the particle.

BRIEF SUMMARY OF DRAWINGS

[0065] These and other features, aspects, and advantages of disclosed compositions and methods will become better understood with regard to the following description, and accompanying drawings, where:

[0066] **FIG 1** shows the design of an exemplary comPACT mini-gene. SS refers to the optional signal sequence; US1 refers to the first universal target site; neoantigen refers to the antigenic peptide sequence site; US2 refers to the second universal target site; L1 refers to the optional first linker sequence; β 2M refers to the β -2-microglobulin domain sequence; L2 refers to the optional second linker sequence; MHC heavy chain refers to the MHC heavy chain allele; L3 refers to the optional third linker sequence; and purification cluster refers to the optional purification cluster with a biotinylation sequence, a protease cleavage site, and an affinity tag sequence.

[0067] **FIG 2** shows a diagram of an exemplary modular off-the-shelf platform for rapidly assembling libraries of antigenic peptide ligands complexed with a chosen MHC allele.

[0068] **FIG 3** is a diagram of an exemplary restriction digest cloning reaction to replace the dummy insert in the MHC template with a chosen neoepitope sequence. The dummy insert (underlined, bold) contains four stop codons in different frames and a unique restriction site for destruction of uncut or re-ligated template. The restriction sites on either side of the insert are shown in boxes.

[0069] **FIG 4** is a diagram of an exemplary restriction digest cloning reaction to insert a chosen neoepitope sequence in the MHC template. The neoepitope sequence (underlined, bold) is synthesized as a primer flanked by two different restriction sites (boxed). A universal primer with the reverse complement sequence of the 3' restriction site is used in a PCR reaction to form a double stranded primer dimer of the neoepitope sequence. Restriction digest of both the neoepitope and the MHC template vector allow for a ligation reaction to insert the chosen neoepitope sequence into the MHC template sequence. Ligation reactions are transformed into *E. coli* and plasmids prepared from transformed *E. coli* are used in mammalian producer cell transfection reactions.

[0070] **FIG 5** is a diagram of an exemplary alternative form of a restriction digest cloning reaction to insert a chosen neoepitope sequence in the MHC template. Two complementary NeoE-encoding primers are synthesized with a portion of the 5' and 3' restriction sites. These primers are annealed and simulate the overhangs from restriction digestion. A precut vector

(which critically retains 5' phosphates on its overhang ends) is then ligated with the annealed NeoE insert and the ligation product is transformed into *E. coli* for plasmid production.

[0071] FIG 6 is a diagram of an exemplary PCR-based method to insert a chosen neoepitope sequence in the MHC template. Two complementary NeoE-encoding primers are synthesized, the forward primer with a 3' sequence for the second universal site in the MHC template; and the reverse primer with a 3' sequence for the complementary sequence of the first universal site in the MHC template. These primers are mixed with a 5' fragment of the MHC template with the first universal sequence site, and a second fragment of the MHC template with the second universal site and remainder of the comPACT mini-gene. The first PCR amplification cycle produces two nucleotide fragments, one fragment encoding the first universal site region with downstream neoepitope and the other encoding the neoepitope followed by the remainder of the comPACT gene. These two fragments, overlapping at the unique neoepitope sequence are then assembled and the full assemble amplified and cleaned up for transfection.

[0072] FIG 7 shows the total protein expression in 30 mL of mammalian cells transfected with a comPACT gene (Neo12) over a seven day time course and a Western Blot using an NTA-HRP reagent that detects the His-tag.

[0073] FIG 8 shows gels of Ni-NTA affinity chromatography purification of the Neo12 comPACT protein. Pre stands for crude lysate, FT stands for Flow-through, W stands for Wash, and E stands for Eluted.

[0074] FIG 9 shows the size exclusion chromatography spectra of the purified Neo12 protein. The major peak is the Neo12 protein, and the minor peak is ATP added during a biotinylation step.

[0075] FIG 10 shows a purification experiment similar to the one shown in FIG 8, using a 0.7 cell culture volume.

[0076] FIG 11 shows crude and purified protein of eight different NeoE comPACT proteins, each with a different antigenic sequence.

[0077] FIG 12 shows the size exclusion chromatography spectra of the eight NeoE comPACT protein of FIG 11.

[0078] FIG 13A shows a NeoE comPACT protein produced using the PCR assembly method described in FIG 6 (Linear amplicon) compared to a NeoE comPACT protein produced from a plasmid (plasmid). **FIG 13B** shows a DNA gel of linear amplicons produced by the PCR assembly method. Each lane contains a comPACT mini-gene with a different neoepitope sequence.

[0079] FIG 14 shows a streptavidin bead pulldown assay to test for complete biotinylation of the comPACT protein.

[0080] FIG 15 shows biotinylation of different comPACT proteins in crude cell lysate, visualized via a Western Blot using streptavidin-HRP.

[0081] FIG 16A-C show production and purification of BirA enzyme (16B and 16C) and TEV protease (16A) in *E. coli*.

[0082] FIG 17 show a biotinylation of a comPACT protein using BirA (lane 2) and cleavage of the His6 tag using TEV protease (lane 3). Untreated comPACT protein is shown in lane 1.

[0083] FIG 18 shows cell sorting of cells transduced with BirA and V5 based on V5 expression.

[0084] FIG 19 shows antigen-specific capture of T cells using multimerized comPACT protein.

[0085] FIG 20 shows comPACT NTamer production using an S88C β 2M comPACT protein.

[0086] FIG 21 shows coupling of Cy5 to S88C comPACT protein monomers.

[0087] FIG 22A shows an exemplary diagram of a cloning strategy to manufacture comPACT polynucleotides. FIG 22B provides sequence verification statistics obtained from 824 individual comPACT polynucleotides.

[0088] FIG 23 shows protein expression and purification for a representative selection of comPACT proteins.

[0089] FIG 24 provides an exemplary diagram of the work flow to manufacture comPACT polynucleotides and proteins.

[0090] FIG 25 show biotinylation of EBV and MART-1 comPACT protein using BirA.

[0091] FIG 26A-B show the percent of patients covered by top HLA I alleles in the United States relative to comPACT HLA repertoire size.

[0092] FIG 27A-C shows comPACT protein monodispersity, yield, and expression for a representative selection of comPACT proteins.

[0093] FIG 28A-C show that neo12 comPACT molecules modified with Cy5 and His tags bind neo12 TCR edited T cells.

[0094] FIG 29A-D show that Neo12 NTamers bind T cells in antigen specific manner and that the presence of imidazole prevents NTamer binding.

[0095] FIG 30A-D show that MART-1 NTamers bind T cells in antigen specific manner.

[0096] FIG. 31A-C show that F5 TCR edited T cells show rapid Cy5 signal decay after NTamer disruption by imidazole.

DETAILED DESCRIPTION

Definitions

[0097] Terms used in the claims and specification are defined as set forth below unless otherwise specified.

[0098] Embodiments of the compositions and methods disclosed herein include a recombinant antigen-MHC complex that is capable of pairing with cognate T cells. As used herein, “antigen-MHC,” “antigen-MHC complex,” “recombinant antigen-MHC complex,” “peptide MHC,” and “p/MHC,” are used interchangeably to refer to a major histocompatibility complex with a peptide in the antigen binding groove of MHC.

[0099] As used herein, “antigen” includes any antigen including patient-specific neoantigens. “Antigenic peptide” refers to a peptide or peptide fragment capable of binding an MHC molecule. “Neoantigen” refers to an antigen that has at least one alteration that makes the neoantigen or presentation of the neoantigen distinct from its corresponding wild-type antigen, *e.g.*, mutations in the polypeptide sequence, differences in post-translation modifications, or differences in expression level. “Tumor neoantigens” refer to neoantigens that are derived from a tumor or a cancer, *e.g.*, from the tumor of a patient.

[00100] As used herein, a “polynucleotide” may refer to ssDNA, dsDNA, ssRNA, dsRNA, or mRNA. One of skill in the art can understand which form is being referred to, *e.g.*, based on the context in which the polynucleotide is being used.

[00101] The term “in vivo” refers to processes that occur in a living organism, including a cell.

[00102] The term “mammal” as used herein includes both humans and non-humans and include but is not limited to humans, non-human primates, canines, felines, murines, bovines, equines, and porcines.

[00103] The term percent “identity,” in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (*e.g.*, BLASTP and BLASTN or other algorithms available to persons of skill) or by visual inspection. Depending on the application, the percent “identity” can exist over a region of the sequence being compared, *e.g.*, over a functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

[00104] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and

reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[00105] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., *infra*).

[00106] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/).

[00107] It must be noted that, as used in the specification and the appended claims, the singular forms “a,” “an” and “the” include plural referents unless the context clearly dictates otherwise.

Other interpretational conventions

[00108] Ranges recited herein are understood to be shorthand for all of the values within the range, inclusive of the recited endpoints. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting of 1, 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, and 50. For example, a sub-range of 1 to 50 can include 2-40, 5-25, and 10-20.

Nucleotide and Peptide Compositions

[00109] T-cell mediated immunity can be characterized by the activation of antigen-specific cytotoxic T cells that are able to induce death in cells that display antigen in a major histocompatibility complex (MHC) on their surface. These cells displaying an MHC complex loaded with antigen include virus-infected cells, cells with intracellular bacteria, cells that have internalized or phagocytosed extracellular sources of protein, and cancer cells displaying tumor antigens.

[00110] A natural class I MHC heavy chain comprises about 350 amino acids; a natural β 2-microglobulin comprises about 100 amino acids; and a class I antigen peptide typically has a length of from about 7 to about 15 amino acids. Class I heavy chains are encoded by genes of the major histocompatibility complex, designated HLA-A, -B and -C in humans, and H-2K, D, and L in mice. The class I heavy chains and β 2-microglobulin are separately encoded on different chromosomes. Antigen peptides are normally processed by cells from protein sources such as, for example, viruses, bacteria, or cancer cells. Diverse variants have been identified for the polypeptides encoded by the HLA-A, -B and -C MHC genes in humans, as well as the murine H-2K, D, and L MHC genes.

[00111] Embodiments of the method disclosed herein are directed to a method of manufacturing a single molecule in which a selected antigen (e.g., a neoantigen) is linked to an MHC complex comprising a β 2-microglobulin (β 2M) and an MHC heavy chain. Different MHC heavy chains can be linked to the β 2M molecule to form a varying number of MHC templates. The methods disclosed herein of inserting an antigen into an MHC template via restriction digest or PCR-based assembly by utilizing universal target sequences flanking the antigen insertion site results in the ability to construct a library of different antigen-MHC complexes in a high-throughput method that can be, e.g., personalized for a given patient. These exemplary complexes are included in the term “comPACTs,” and can then be, e.g., linked to a particle, barcoded particle, or surface for use in isolation and identification of patient-specific T cell populations targeted to patient-specific neoantigens. Methods of linking antigen-MHC complexes and use of such complexes are disclosed in PCT/US2018/021611, filed March 8, 2018, herein incorporated by reference in its entirety.

Peptide-MHC complex

[00112] Briefly, as used herein, “comPACTs” refer to a single polypeptide fusion including a universal target sequence, an antigen peptide, a second universal target sequence, a β 2-microglobulin, and a MHC class I heavy chain comprising, e.g., the α 1, α 2, and α 3 domains that forms an MHC display moiety. An MHC display moiety can include a recombinant MHC molecule. In certain embodiments, comPACTs can comprise disulfide traps, as described in US Publication No. 2009/0117153 and US Publication No. 2008/0219947; each of which is herein incorporated by reference. The antigen-MHC complex formed by a comPACT results in display of the antigens such that they are capable of recognition by a cognate TCR molecule. In some embodiments, the MHC complex can be an MHC Class I (MHC I) complex that pairs with CD8-positive (CD8+) T “killer” cells. In

some embodiments, the MHC complex can be an MHC Class II (MHC II) complex that pairs with CD4-positive (CD4+) T cells. The MHC allele encoded in each comPACT can be easily swapped out for other MHC I or II alleles, enabling antigenic interrogation of T cells from patients of any MHC haplotype.

[00113] In some embodiments, the MHC class I heavy chain sequence of a comPACT can include one or more amino acid substitutions, additions, and/or deletions, such as a substitution of tyr-84 with a non-aromatic amino acid other than proline. In these embodiments, the amino acid substitution can be an amino acid encoded by the standard genetic code such leucine, isoleucine, valine, serine, threonine, alanine, histidine, glutamine, asparagine, lysine, aspartic acid, glutamic acid, cysteine, arginine, serine or glycine, or can be a modified or unusual amino acid. In one embodiment, the MHC class I heavy chain sequence of a comPACT comprises a tyrosine-84 to alanine substitution. In another embodiment, the MHC class I heavy chain sequence of a comPACT comprises a tyrosine-84 to cysteine substitution.

[00114] In some embodiments, the MHC allele lacks a transmembrane domain. In some embodiments, the MHC allele lacks a cytoplasmic domain. In some embodiments, the MHC allele lacks a transmembrane and a cytoplasmic domain. In some embodiments, the HLA allele lacks a transmembrane domain. In some embodiments, the HLA allele lacks a cytoplasmic domain. In some embodiments, the HLA allele lacks a transmembrane and a cytoplasmic domain.

[00115] Any MHC or HLA allele may be used in the comPACT described herein. Exemplary HLA alleles include, but are not limited to, HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*11:01, HLA-A*23:01, HLA-A*30:01, HLA-A*33:03, HLA-A*25:01, HLA-A*26:01, HLA-A*29:02, HLA-A*68:02, HLA-B*07:02, HLA-B*14:02, HLA-B*18:01, HLA-B*27:02, HLA-B*39:01, HLA-B*40:01, HLA-B*44:02, HLA-B*46:01, HLA-B*50:01, HLA-B*57:01, HLA-B*58:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:03, HLA-B*35:01, HLA-B*40:02, HLA-B*42:01, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*13:02, HLA-B*15:07, HLA-B*27:05, HLA-B*35:03, HLA-B*37:01, HLA-B*38:01, HLA-B*41:02, HLA-B*44:05, HLA-B*49:01, HLA-B*52:01, HLA-B*55:01, HLA-C*02:02, HLA-C*03:04, HLA-C*05:01, HLA-C*07:01, HLA-C*01:02, HLA-C*04:01, HLA-C*06:02, HLA-C*07:02, HLA-C*16:01, HLA-C*03:03, HLA-C*07:04, HLA-C*08:01, HLA-C*08:02, HLA-C*12:02, HLA-C*12:03,

HLA-C*14:02, HLA-C*15:02, and HLA-C*17:01. Any other suitable HLA allele known in the art may be used in the comPACT described herein.

[00116] The β 2-microglobulin (β 2M) may include a recombinant β 2M molecule. In some embodiments, the β 2M sequence can include one or more amino acid substitutions, additions, and/or deletions as described above. In one embodiment, this substitution comprises a serine-88 to cysteine substitution.

[00117] Amino acid and nucleotide sequences for an exemplary comPACT protein with a hGH signal sequence, dummy neoantigen insert, HLA*A02:01 allele, AviTag peptide, TEV cleavage site, and concatenated His tag is shown below in **Table 1**.

Table 1		
SEQ ID NO		
201	comPACT HLA*A02:01 peptide	MATGSRTSLLLAFLGLCLPWLQEGSA*LEIKCE**GCGSGGGSGGGSGGGSIQRTPK IQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSF YLLYYTEFTPTTEKDEYACRVNHVTLSPKIVKWRDRMRKVKAHSQTHRVDLGLTRG CYNQSEAGSHTVQRMYGCDVGSWDRFLRGYHQYAYDGKDYIALKEDLSRWTADMA AQTTKHKEAAHVAEQLRAYLEGTCVEWLRRLYLENGKETLQRTDAPKTHMTHHAVS DHEATLRCWALSFYPAEITLTWQRDGEDQTQDTELVETRPAGDGTQKWAAVVPS GQEQRYTCHVQHEGLPKPLTLRWEPGSGSGSGGSAGGGLNDIFEAQKIEWHEGGGEN LYFQGGSHHHHHHGGGSGGGSGSHHHHHH
202	comPACT HLA*A02:01 nucleotide	atggcgacgggttcaagaacttccctacttcttgcatttggcctgctttgtttgccc gtggttacaggagggctcagcatgactcgagataaaatgtgaataatgaggatgcg gaggatccggcggaggcgggagcggagcggagggatccagcgtactccaaaga ttcagggtttactcacgtcatccagcagagaatggaaagtcaaatttcctgaattgc tatgtgtctgggtttcatccatccgacattgaagttgacttactgaagaatggaga gagaattgaaaaagtggagcattcagacttgtctttcagcaaggactggtctttct atctcttgactacactgaattcacccccactgaaaaagatgagtatgcctgccgt gtgaaccatgtgactttgtcacagcccaagatagttaagtgggacgagacatggg cgaggcgggagcggcggagcgggtccggcggagcgggtccggagggggaggca gcggtctctactccatgaggtatcttctcacatccgtgtcccgcccgccggcggg gagccccgcttcatcgcatgggtacgtggacgacacgcagttcgtgcggttcga cagcgacgccgcgagccagaggatggagccgcgggcgcggtggatagagcaggag gtcccaggtattgggacggggagacacggaaagtgaaggccactcacagactcac cgagtggacctggggaccctgcgcggctgctacaaccagagcggagccggttctca cacggtccagaggatgtatggctgcgacgtggggtcggactggcgcttccctccgcg ggtaccaccagtagcctacgacggcaaggattacatcgccctgaaagaggacctg cgctcttgaccgcggcgacatggcagctcagaccaccaagcacaagtgggagggc ggcccatgtggcggagcagttgagagcctacctggagggcagctgctggagtggc tccgcagatacctggagaacgggaaggagacgtgcagcgcacggacgccccaaa acgcatatgactcaccacgctgtctctgacctgaagccaccctgaggtgctgggc cctgagcttctaccctgcggagatcacactgacctggcagcgggatggggaggacc agaccaggacacggagctcgtggagaccaggcctgcaggggatggaaccttcag aagtggcggtgtgtgtgtgcttcttgacaggagcagagatacactgccatgt gcagcatgagggtttggcccaagccctcaccctgagatgggagccgggcagcgcg gcagcgggggctccgcggcgaggcctgaaacgacatcttcgaagcccagaagatc gagtggcacgagggcgggggagagaacctgtacttccagggcggcagccaccacca tcaccaccatggcgggcggaagcgggcggggtccggcagccaccatcaccatcacc at

[00118] In some embodiments, the polynucleotide molecule comprises, in a 5' to 3' orientation, (i) a promoter sequence, (ii) the first universal target sequence comprising the sequence shown in SEQ ID NO: 3, (iii) the nucleotide sequence encoding an antigenic peptide, wherein the antigenic peptide is a tumor neoantigen, (iv) the second universal target sequence comprising the sequence shown in SEQ ID NO: 4, (v) the β 2M sequence comprising the sequence shown in SEQ ID NO: 106, (vi) the MHC allele sequence comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOs: 109-174, (vii) a first affinity tag sequence comprising the sequence shown in SEQ ID NO: 29, (viii) a protease cleavage site sequence comprising the sequence shown in SEQ ID NO: 31, and (ix) a second affinity tag sequence comprising the sequence shown in SEQ ID NO: 35. In some embodiments, the polynucleotide molecule comprises, in a 5' to 3' orientation, (i) a promoter sequence, (ii) the first universal target sequence comprising the sequence shown in SEQ ID NO: 3, (iii) the nucleotide sequence encoding an antigenic peptide, wherein the antigenic peptide is a tumor neoantigen, (iv) the second universal target sequence comprising the sequence shown in SEQ ID NO: 4, (v) the β 2M sequence comprising the sequence shown in SEQ ID NO: 106, (vi) the MHC allele sequence comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOs: 109-174. In some embodiments, the polynucleotide molecule comprises, in a 5' to 3' orientation, (i) a promoter sequence, (ii) the first universal target sequence comprising the sequence shown in SEQ ID NO: 3, (iii) the nucleotide sequence encoding an antigenic peptide, wherein the antigenic peptide is a tumor neoantigen, (iv) the second universal target sequence comprising the sequence shown in SEQ ID NO: 4, (v) the β 2M sequence, (vi) the MHC allele sequence comprising a sequence selected from the group consisting of the sequences shown in SEQ ID NOs: 109-174, (vii) a first affinity tag sequence, (viii) a protease cleavage site sequence, and (ix) a second affinity tag.

[00119] In some embodiments, the polypeptide comprises in a 5' to 3' orientation, (i) a promoter peptide comprising the sequence shown in SEQ ID NO: 2, (ii) the first universal target peptide, (iii) the nucleotide sequence encoding an antigenic peptide, wherein the antigenic peptide is a tumor neoantigen, (iv) the second universal target peptide, (v) the β 2M peptide comprising the sequence shown in SEQ ID NO: 105, and (vi) the MHC peptide comprising a sequence shown in SEQ ID NOs: 38-103.

Universal Sequences

[00120] An antigenic peptide is generally flanked by universal target sequences or portions thereof. These sequences allow for rapid, high throughput methods for replacing or inserting the antigenic peptide encoding nucleotide in the polynucleotide MHC template. Universal sequences may comprise restriction sites for restriction digest-based cloning. Exemplary restriction sites include, but are not limited to, NotI, BamHI, BspEI, BstBI, XbaI, HindIII, EcoRI, ApaI, NotI, and any combination thereof. In certain aspect, one or more universal target sequences are not present in the genetic material being manipulated, e.g., to reduce or eliminate off-target effects and/or to increase specificity.

[00121] Universal target sequences may be 4-100, between 4-15, between 15-40, between 15-35, between 15-30, between 20-40, between 25-40, between 30-40, between 15-75, between 50-100, between 50-75, between 75-100, between 25-50, between 40-50, between 50-60, between 60-70, between 70-80, between 80-90, or between 90-100 nucleotides in length. Universal sequences may be at least 4, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 nucleotides in length. In some embodiments, the universal target sequence is 4-8 nucleotides in length, e.g., 4, 5, 6, 7, or 8. In other embodiments, the universal target sequence is between 25-35 nucleotides in length, e.g., 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or 35 nucleotides. In other embodiments, the universal target sequence is between 35-75 nucleotides in length, e.g., 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, or 75 nucleotides. In other embodiments, the universal target sequence is at least about 15 nucleotides in length. In some embodiments, the polynucleotide comprises at least two universal target sequences that are not the same. In some embodiments, the polynucleotide comprises at least two universal target sequences that are the same.

[00122] Exemplary universal target sequences are shown in Table 2:

Table 2. Universal Target Sequences		
SEQ ID NO.	Restriction Site	Sequence
3	BlpI	CGTGGTTACAGGAGGGCTCAGCA
4	BamHI	GGATGCGGAGGATCCGGCG
5	BamHI	GGAAGCGGAGGATCCGGCG
6	BamHI	GGAAGCGGAGGATCCACCAGC

[00123] In some embodiments, the universal target sequences comprise polymerase chain reaction (PCR) primer target sequences or primer binding sites. Universal primer sequences

known in the art may be used in the compositions and methods disclosed herein, or the sequences may be different than the previously described universal primer sequences.

[00124] In some embodiments, the first and/or second universal target sequences comprise restriction enzyme cleavage sites.

Linkers

[00125] In various embodiments, a comPACT can comprise a first flexible linker interposed between the antigenic peptide segment and the β 2-microglobulin segment. Such linkers can extend from and connect the carboxyl terminal of the antigenic peptide segment to the amino terminal of the β 2-microglobulin segment, or vice versa. Without being limited by theory, when a comPACT is expressed, the linked peptide ligand can fold into the binding groove resulting in a functional comPACT protein. In various embodiments, this linker can comprise at least about 4 amino acids, up to about 20 amino acids, e.g., 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids.

[00126] In various embodiments, a comPACT can comprise a second flexible linker interposed between the β 2-microglobulin segment and the MHC heavy chain segment. Such linkers can extend from and connect the carboxyl terminal of the β 2-microglobulin segment to the amino terminal of the heavy chain segment, or vice versa. Without being limited by theory, when a comPACT is expressed, the β 2-microglobulin and the heavy chain can fold into the binding groove resulting in a molecule which can function in promoting T cell expansion. In various embodiments, this linker can comprise at least about 4 amino acids, up to about 20 amino acids, e.g., 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids.

[00127] In various embodiments, a comPACT can comprise a third flexible linker interposed between the MHC heavy chain segment and the purification cluster. Such linkers can extend from and connect the carboxyl terminal of the heavy chain segment and the amino terminus of the purification cluster, or vice versa.

[00128] Any appropriate flexible linker sequence known in the art may be used. Such linker sequences include, but are not limited to, glycine-serine sequences comprising one or more repeating units of a GS, SG, GGGGS (G_4S), GGGS (G_3S), GSGGS, or GCGGS sequence motifs. Multiple consecutive units of the linker sequence motif may be used. In some embodiments, the flexible linker sequence comprises 1, 2, 3, 4, 5, 6, 7, 8, or 10 repeating units of an appropriate linker sequence.

[00129] Linkers can also comprise cysteine residues for disulfide bonds between the linker sequence and the MHC allele sequence, such that the cysteine residues form a disulfide trap. In some embodiments, the linker sequence or second universal target peptide comprises a cysteine residue. In some embodiments, the linker sequence or second universal target peptide comprises a cysteine residue that forms a disulfide bond with the MHC allele.

Signal Sequences

[00130] In various embodiments, the comPACT polynucleotide and polypeptide may comprise a signal sequence, e.g., encoding a signal peptide in the case of a polynucleotide. The signal sequence can be a secretion signal sequence. Secretion signal sequences direct translated proteins in mammalian cells through the secretory pathway, and ensure that the translated proteins are subject to cellular quality control. Inclusion of secretory signals can ensure that the comPACT proteins are secreted into the cellular media, such that they are homogeneously well-folded and more easily isolated from the media or clarified supernatant.

[00131] In one embodiment, the signal sequence is a signal sequence from Human Growth Hormone (hGH). In another embodiment, the signal sequence is a signal sequence from hIG1 Kappa light chain, Beta 2 microglobulin (β 2M), or IL2 signal sequence. Additional signal sequences may also be used, including a signal sequence from β 2M (e.g., human β 2M) or any other eukaryotic or prokaryotic signal sequence known in the art.

[00132] The signal sequence may be between 10-100, 10-20, 10-50, 10-40, 20-40, 20-60, 40-90, 40-60, 45-70, 50-80, 60-90, 55-70, 60-80, 70-80, 50-100, 60-100, 70-100, 80-100, or 90-100 nucleotides in length. The signal peptide may be between 3-33, 3-10, 10-30, 10-20, 15-30, or 20-30 amino acids in length.

[00133] Exemplary signal sequences are shown in Table 3:

Table 3. Signal Sequence		
SEQ ID NO.	Signal Protein	Sequence
1	Human Growth Hormone nucleotide	ATGGCGACGGGTTCAAGAACTTCCCTACTTCTTGCATTGGCCTGCTTTGTTTGCCGTGGTTACAGGAGGGCTCAGCA
2	Human Growth Hormone peptide	MATGSRTSLLLAFLGLCLPWLQEGSA
23	hIG1 Kappa light chain, signal sequence	MDMRVPAQLLGLLLWLSGARC
24	hIG1 Kappa light chain, signal sequence	atggatatgcgcgtgccggcgagctgctgggcctgctgctgctgtggc tgagcggcgcgctgc
25	B2M, signal sequence peptide	MSRSVALAVL ALLSLSGLEA
26	B2M, signal sequence	atgagccgcagcgtggcgctggcggtgctggcgctgctgagcctgagcgc gcctggaagcg

27	IL2, signal sequence	MYRMQLLSCL ALSIALVTNS
28	IL2, signal sequence	atgtatcgcatgcagctgctgagctgcattgcgctgagcctggcgctgg tgaccaacagc

[00134] In some embodiments, the signal sequence comprises a sequence comprising the sequence shown in SEQ ID NOs: 1, 24, 26, or 28. In some embodiments, the signal sequence comprises a sequence comprising the sequence shown in SEQ ID NOs: 2, 23, 25, or 27. In one embodiment, the signal sequence encodes the Human Growth Hormone (HGH) signal sequence. In another embodiment, the signal sequence comprises the sequence shown in SEQ ID NO: 1. In another embodiment, the signal sequence comprises the sequence shown in SEQ ID NO: 2.

Promoters

[00135] A comPACT polynucleotide composition may further comprise a promotor, e.g., for transcription of an mRNA transcript that can be translated by a host cell. Promoters may be prokaryotic or eukaryotic (e.g., mammalian) in origin. Any appropriate promoter for gene transcription in a cell may be used, such as EF1 α , cytomegalovirus (CMV), or SV40. In some embodiments, the comPACT polynucleotide comprises a CMV promoter.

[00136] In various embodiments, the 5' end of the polynucleotide sequence further comprises a promoter sequence linked to the 5' end of the first universal target. In some embodiments, the promoter sequence is selected from the group consisting of CMV, EF1 α , and SV40 promoters. In one embodiment, the promoter sequence is a CMV promoter sequence. In one embodiment, the promoter sequence comprises a sequence comprising the sequence shown in SEQ ID NO: 37.

Affinity tags

[00137] A comPACT polynucleotide composition may further comprise at least one sequence that encodes for an affinity tag or peptide. In some embodiments, the comPACT polynucleotide comprises at least two affinity tags or peptide sequences.

[00138] Any appropriate affinity tag or peptide may be used in a comPACT polynucleotide or polypeptide. Such peptides include, but are not limited to, AviTag, streptavidin-tag, polyhistidine (His6)-tag, FLAG-tag, HA-tag, and Myc-tag. The sequences in the polynucleotide comPACT gene are typically translated into peptides in the comPACT polypeptide. These epitopes may be used for affinity chromatography purification or quantification of the expressed comPACT polypeptide. For instance, the His6 tag may be used to purify the comPACT protein via Ni-NTA affinity chromatography. In some

embodiments, the His6 tag may be a concatenated His6 tag, comprising multiple His6 units, with an optional linker sequence. The first and second affinity tags can be the same, i.e. both His6 tags, or both HA tags, or they can be different tags, i.e. a streptavidin-tag and a His6 tag.

[00139] In addition, the AviTag encodes a biotinylation site that is recognized by BirA enzyme. Inclusion of this peptide sequence in a protein allows for biotinylation of the sequence via enzymatic modification by BirA. Thus, a comPACT polypeptide comprising an AviTag sequence and a His6 tag may be biotinylated, purified via Ni-NTA affinity chromatography via the His6 tag, and the purity or quantity of the purified protein assessed via biotin visualization with streptavidin or other avidin reagents.

[00140] In some embodiments, the comPACT polynucleotide comprises an AviTag sequence. In some embodiments, the comPACT polypeptide comprises an AviTag peptide. In some embodiments, the comPACT polynucleotide comprises a His6 sequence. In some embodiments, the comPACT polypeptide comprises a His6 peptide. In some embodiments, the comPACT polypeptide comprises a concatenated His6 peptide. In some embodiments, the comPACT polynucleotide comprises an AviTag sequence and a His6 sequence. In some embodiments, the comPACT polynucleotide comprises an AviTag sequence and a concatenated His6 sequence. In some embodiments, the comPACT polypeptide comprises an AviTag peptide and a His6 peptide. In some embodiments, the comPACT polypeptide comprises an AviTag peptide and a concatenated His6 peptide.

[00141] In some embodiments, the first and second affinity tags are selected from the group consisting of: AviTag, streptavidin-tag, polyhistidine (His6)-tag, FLAG-tag, HA-tag, and Myc-tag.

[00142] In some embodiments, the comPACT polynucleotide comprises a FLAG sequence. In some embodiments, the comPACT polypeptide comprises a FLAG peptide. In some embodiments, the comPACT polynucleotide comprises an HA sequence. In some embodiments, the comPACT polypeptide comprises an HA peptide. In some embodiments, the comPACT polynucleotide comprises a Myc sequence. In some embodiments, the comPACT polypeptide comprises a Myc peptide. In some embodiments, the comPACT polynucleotide comprises a streptavidin sequence. In some embodiments, the comPACT polypeptide comprises a streptavidin peptide. In some embodiments, the comPACT polynucleotide comprises an AviTag sequence and a FLAG sequence. In some embodiments, the comPACT polypeptide comprises an AviTag peptide and a FLAG peptide. In some embodiments, the comPACT polynucleotide comprises an AviTag sequence and an HA

sequence. In some embodiments, the comPACT polypeptide comprises an AviTag peptide and an HA peptide. In some embodiments, the comPACT polynucleotide comprises an AviTag sequence and a streptavidin sequence. In some embodiments, the comPACT polypeptide comprises an AviTag peptide and a streptavidin peptide.

[00143] In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for a TEV cleavage site, and the second affinity tag encodes for a FLAG peptide. In some embodiments, the purification cluster comprises an AviTag peptide, a TEV cleavage site, and a FLAG peptide. In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for a TEV cleavage site, and the second affinity tag encodes for an HA peptide. In some embodiments, the purification cluster comprises an AviTag peptide, a TEV cleavage site, and an HA epitope. In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for a TEV cleavage site, and the second affinity tag encodes for a Myc peptide. In some embodiments, the purification cluster comprises an AviTag peptide, a TEV cleavage site, and a Myc peptide. In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for a TEV cleavage site, and the second affinity tag encodes for a streptavidin peptide. In some embodiments, the purification cluster comprises an AviTag peptide, a TEV cleavage site, and a streptavidin peptide.

[00144] In some embodiments, the first affinity tag sequence encodes for a His peptide. In some embodiments, the first affinity tag sequence encodes for a FLAG peptide. In some embodiments, the first affinity tag sequence encodes for an HA peptide. In some embodiments, the first affinity tag sequence encodes for a Myc peptide. In some embodiments, the first affinity tag sequence encodes for a streptavidin peptide. In some embodiments, the purification cluster comprises a His peptide. In some embodiments, the purification cluster comprises a FLAG peptide. In some embodiments, the purification cluster comprises an HA peptide. In some embodiments, the purification cluster comprises a Myc peptide. In some embodiments, the purification cluster comprises a streptavidin peptide.

Protease cleavage sites

[00145] A comPACT polynucleotide composition may further comprises a sequence that encodes for a protease cleavage site, e.g., in the purification cluster. This cleavage site may be encoded between the first and second affinity tag sequences and allow for cleavage of the second affinity tag from the comPACT protein once the comPACT has been expressed and under gone a round of purification. Any appropriate protease cleavage site known in the art may be used, including, but not limited, cleavage sites that are recognized by TEV, thrombin, Factor Xa, enteropeptidases, and rhinovirus 3C protease, among others.

[00146] In one embodiment, the protease cleavage site sequence is a TEV cleavage site sequence, a thrombin cleavage site sequence, a Factor Xa cleavage site sequence, an enteropeptidase cleavage site sequence, and/or a rhinovirus 3C protease cleavage site sequence. In one embodiment, the protease cleavage site nucleotide sequence encodes for a TEV cleavage site. In another embodiment, the comPACT polypeptide comprises a TEV protease cleavage site, ENLYFQG. In one embodiment, the protease cleavage site nucleotide sequence encodes for a Factor Xa cleavage site. In one embodiment, the protease cleavage site nucleotide sequence encodes for a rhinovirus 3C cleavage site. In one embodiment, the protease cleavage site nucleotide sequence encodes for an enteropeptidase cleavage site. In one embodiment, the protease cleavage site nucleotide sequence encodes for a thrombin cleavage site.

[00147] In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for a TEV cleavage site, and the second affinity tag encodes for a His6 peptide. In some embodiments, the purification cluster comprises an AviTag epitope, a TEV cleavage site, and a His6 epitope. In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for a thrombin cleavage site, and the second affinity tag encodes for a His6 peptide. In some embodiments, the purification cluster comprises an AviTag epitope, a thrombin cleavage site, and a His6 epitope. In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for a Factor Xa cleavage site, and the second affinity tag encodes for a His6 peptide. In some embodiments, the purification cluster comprises an AviTag epitope, a Factor Xa cleavage site, and a His6 epitope. In some embodiments, the first affinity tag sequence encodes for an AviTag peptide, the protease cleavage site sequence encodes for an enteropeptidase cleavage site, and the second affinity tag encodes for a His6 peptide. In some embodiments, the purification cluster comprises an AviTag epitope, an enteropeptidase cleavage site, and a His6 epitope. In some embodiments, the first affinity tag sequence

encodes for an AviTag peptide, the protease cleavage site sequence encodes for a rhinovirus 3C cleavage site, and the second affinity tag encodes for a His6 peptide. In some embodiments, the purification cluster comprises an AviTag epitope, a rhinovirus 3C cleavage site, and a His6 epitope.

PolyA tail

[00148] A comPACT polynucleotide composition may further comprise a polyadenylation (polyA) tail. Mammalian, eukaryotic, or prokaryotic polyA sequence motifs may be used. This sequence may be included when the comPACT polynucleotide is assembled via PCR for direct transfection into a host cell (*e.g.*, not in the context of an expression construct or vector). Any appropriate polyA tail and sequence motif may be used in the comPACT polynucleotide, including, but not limited to, SV40, hGH, bGH, and rbGlob sequences. Such sequences include the RNA sequence motif: AAUAA. In one embodiment, the polyA sequence comprises a bGH polyA sequence. In one embodiment, the polyA sequence is selected from a bGH polyA sequence, an SV40 polyA sequence, an hGH polyA sequence, and an rbGlob polyA sequence. In one embodiment, the polyA sequence comprises an SV40 polyA sequence. In one embodiment, the polyA sequence comprises an hGH polyA sequence. In one embodiment, the polyA sequence comprises an rbGlob polyA sequence.

Antigenic sequences

[00149] Antigenic sequences may be between 5-100, between 5-10, between 10-20, between 10-30, between 10-40, between 10-50, between 10-60, between 10-70, between 10-80, between 10-90, between 10-100, between 20-100, between 30-100, between 40-100, between 50-100, between 60-100, between 70-100, between 80-100, between 90-100, between 20-40, between 20-50, between 20-60, between 20-70, between 20-80, between 20-19, between 20-100, between 20-30, between 25-35, between 20-45, between 30-45, between 30-50, between 30-60, between 30-70, between 30-80, between 30-90, between 30-100, between 40-50, between 40-60, between 45-60, between 40-70, between 40-80, between 40-90, between 40-100, between 50-60, between 50-70, between 50-80, between 50-90, between 50-100, between 60-70, between 60-80, between 60-90, between 60-100, between 70-80, between 80-90, between 80-100, or between 90-100 nucleotides in length. Antigenic sequences may be at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides in length. The antigenic peptide may be between 3-50, between 3-10, between 5-15, between 7-15, between 5-20, between 7-20, between 10-15, between 10-20, between 15-20, between 20-25, between 20-30, between 25-35, between 30-40, or between

40-50 amino acids in length. The antigenic peptide may be 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amino acids in length. The antigenic peptide may include a tumor antigen, a neoantigen, a tumor neoantigen, a viral antigen, bacterial antigen, phosphoantigen, or a microbial antigen. In one embodiment, the antigenic peptide is a neoantigen. The antigenic peptides may be selected from patient data to select antigens with one or more somatic mutations. The prediction of the antigenic peptide may include a predicative algorithm and predict binding of the antigenic peptide or neoantigen and an HMC allele. Prediction of the antigenic peptide is further discussed below.

[00150] In some embodiments, the nucleotide sequence encoding an antigenic peptide is between 20-60, between 20-30, between 25-35, between 20-45, between 30-45, between 40-60, or between 45-60 nucleotides in length. In other embodiments, the nucleotide sequence encoding an antigenic peptide is between 20-30 nucleotides in length. In some embodiments, the antigenic peptide is 7-15 amino acids, 7-10, 8-9, 7, 8, 9, 10, 11, 12, 13, 14, or 15 amino acids in length. In some embodiments, the antigenic peptide is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 20 amino acids in length.

Biotinylation

[00151] The compACT proteins described herein may further be biotinylated via any appropriate method. One such method utilizes the BirA Biotin-protein ligase and is commercially available. A specific amino acid sequence, known as the AviTag sequence (GLNDIFEAQKIEWHE), is encoded in the protein of interest. BirA ligase, d-biotin and ATP are added to a reaction mixture containing the protein of interest. BirA covalently ligates the biotin to the lysine in the AviTag sequence, thereby biotinylating the protein of interest. The newly biotinylated protein can then be purified and used in downstream applications. Other methods known in the art to biotinylate proteins may also be utilized.

[00152] In some embodiments, the compACT proteins are biotinylated after protein purification with a purified BirA protein. In some embodiments, the compACT proteins are biotinylated in cell lysate during protein purification with a purified BirA protein. Such methods using purified BirA and purified or partially purified compACT proteins are termed “in vitro” biotinylation. In some embodiments, the compACT proteins are biotinylated during protein production in the cell, by a cell-expressed BirA protein in the cell cytoplasm, on the cell surface, or secreted into the cell culture media. Such methods using cell-expressed BirA and unpurified compACT proteins are termed “in vivo” biotinylation.

Expression Constructs and Vectors

[00153] The comPACT polynucleotide molecules can be inserted into expression constructs or vectors, e.g., for plasmid and protein production. The expression construct or vector can be a plasmid or a viral vector. Any suitable expression construct or vector known in the art may be used, including bacterial expression plasmids, such as *Escherichia coli* or *Bacillus subtilis* plasmids; eukaryotic expression vectors, such as *Pichia pastoris* expression vectors; or viral vectors, such as lentiviral vectors, vaccinia vectors, or baculovirus vectors. Mammalian expression vectors for use in cultured mammalian cell lines such as Chinese hamster ovary (CHO), HEK293, Expi293, or any other suitable mammalian cell line are also contemplated. Additionally, the expression construct or vector may comprise a nucleotide barcode. The nucleotide barcode can be unique for each expression construct or vector. In some embodiments, the nucleotide sequences encoding for the signal sequence, beta-2-microglobulin, and MHC allele can be ligated into an expression construct or vector with a non-coding or dummy antigen insert. This non-coding antigen insert can then be removed by an appropriate cloning technique, such as restriction digest, and a desired antigen sequence inserted via ligation or any other appropriate cloning technique.

[00154] In some aspects, provided herein are a library comprising greater than or equal to two distinct vectors encoding different MHC alleles.

Host Cells

[00155] In another aspect, provided herein are host cells comprising the polynucleotide molecule or the expression construct as described herein. The host cell can be any suitable host cell known in the art, including, but not limited to bacterial cells such as *Escherichia coli* or *Bacillus subtilis*, or eukaryotic host cells such as Chinese hamster ovary (CHO), HEK293, Expi293, HeLa, insect cell lines such as Sf9 or Sf12, or yeast cells such as *Pichia pastoris*. The host cells may also stably express the biotinylation enzyme BirA. The host cell can be a primary cell or an immortalized cell.

[00156] In some embodiments, the polynucleotide is integrated into the cell genome. In some embodiments, the polynucleotide is extrachromosomal. In some embodiments, the host cell is a mammalian cell. In some embodiments, the cell is a human cell. In some embodiments, the cell is selected from the group consisting of: a stem cell, a tumor cell, an immortalized cell, and a fetal cell. In some embodiments, the host cell is a prokaryotic cell. In some embodiments, cell is an *Escherichia coli* cell. In some embodiments, the cell expresses a BirA protein or fragment thereof.

Libraries

[00157] Also considered are libraries comprising greater than or equal to two distinct compACT polynucleotide molecules, polypeptide molecules, or polypeptides molecules attached to particles. The library may comprise 2 to 1000 molecules. In some embodiments, the library comprises between 2-900, 2-800, 2-700, 2-600, 2-500, 2-480, 2-400, 2-300, 2-200, 2-100, 2-50, 2-66, 2-48, 2-30, 2-20, 2-19, 10-1000, 10-900, 10-800, 10-700, 10-600, 10-500, 10-480, 10-400, 10-300, 10-200, 10-100, 10-50, 10-66, 10-48, 10-30, 10-20, 20-1000, 20-900, 20-800, 20-700, 20-600, 20-500, 20-480, 20-400, 20-300, 20-200, 20-100, 20-50, 20-50, 20-66, 20-48, 20-30, 30-1000, 30-900, 30-800, 30-700, 30-600, 30-500, 30-480, 30-400, 30-300, 30-200, 30-100, 30-50, 30-50, 30-66, 30-48, 30-40, 40-1000, 40-900, 40-800, 40-700, 40-600, 40-500, 40-480, 40-400, 40-300, 40-200, 40-100, 40-60, 40-50, 40-66, 40-48, 50-1000, 50-900, 50-800, 50-700, 50-600, 50-500, 50-480, 50-400, 50-300, 50-200, 50-100, 50-60, 50-66, 60-1000, 60-900, 60-800, 60-700, 60-600, 60-500, 60-480, 60-400, 60-300, 60-200, 60-100, 70-1000, 70-900, 70-800, 70-700, 70-600, 70-500, 70-480, 70-400, 70-300, 70-200, 70-100, 70-80, 70-90, 80-1000, 80-900, 80-800, 80-700, 80-600, 80-500, 80-480, 80-400, 80-300, 80-200, 80-100 polynucleotide or polypeptide molecules. In some embodiments, the library comprises between 2-19, 48-480, between 48-66, between 66-480, between 220-240, between 40-60, between 48-66, between 50-70, or between 60-80 polynucleotide or polypeptide molecules. In some embodiments, the library comprises at least 2, 5, 10, 15, 20, 25, 30, 35, 40, 45, 48, 50, 55, 60, 65, 66, 70, 75, 80, 85, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 600, 562, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, or 1000 compACT polynucleotide or polypeptide molecules. In some embodiments, the library comprises 2, 10, 15, 20, 24, 48, 66, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 compACT polynucleotide or polypeptide molecules. The molecules may be polynucleotides, polypeptides, or polypeptides attached to particles. In some embodiments, the greater than or equal to two polynucleotide or polypeptide molecules have distinct antigenic peptide sequences. In some embodiments, the greater than or equal to two molecules have distinct antigenic peptide sequences and distinct MHC molecules.

[00158] In some embodiments, the library comprises greater than or equal to two distinct polynucleotide molecules, wherein each distinct polynucleotide molecule comprises (i) the first universal sequence, (ii) the nucleotide sequence encoding a antigenic peptide, wherein the nucleotide sequence is not the same for each of the greater than or equal to two polynucleotide molecules (iii) the second universal target sequence, (iv) the β 2M sequence,

and (v) the MHC allele sequence. In some embodiments, the MHC allele sequence is not the same for each of the greater than or equal to two polynucleotide molecules.

[00159] In one embodiment, the library comprises at least two or more of the HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*11:01, HLA-A*23:01, HLA-A*30:01, HLA-A*33:03, HLA-A*25:01, HLA-A*26:01, HLA-A*29:02, HLA-A*68:02, HLA-B*07:02, HLA-B*14:02, HLA-B*18:01, HLA-B*27:02, HLA-B*39:01, HLA-B*40:01, HLA-B*44:02, HLA-B*46:01, HLA-B*50:01, HLA-B*57:01, HLA-B*58:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:03, HLA-B*35:01, HLA-B*40:02, HLA-B*42:01, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*13:02, HLA-B*15:07, HLA-B*27:05, HLA-B*35:03, HLA-B*37:01, HLA-B*38:01, HLA-B*41:02, HLA-B*44:05, HLA-B*49:01, HLA-B*52:01, HLA-B*55:01, HLA-C*02:02, HLA-C*03:04, HLA-C*05:01, HLA-C*07:01, HLA-C*01:02, HLA-C*04:01, HLA-C*06:02, HLA-C*07:02, HLA-C*16:01, HLA-C*03:03, HLA-C*07:04, HLA-C*08:01, HLA-C*08:02, HLA-C*12:02, HLA-C*12:03, HLA-C*14:02, HLA-C*15:02, and HLA-C*17:01 alleles. In one embodiment, the library comprises at least HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*24:02, HLA-A*30:02, HLA-A*31:01, HLA-A*32:01, HLA-A*33:01, HLA-A*68:01, HLA-A*11:01, HLA-A*23:01, HLA-A*30:01, HLA-A*33:03, HLA-A*25:01, HLA-A*26:01, HLA-A*29:02, HLA-A*68:02, HLA-B*07:02, HLA-B*14:02, HLA-B*18:01, HLA-B*27:02, HLA-B*39:01, HLA-B*40:01, HLA-B*44:02, HLA-B*46:01, HLA-B*50:01, HLA-B*57:01, HLA-B*58:01, HLA-B*08:01, HLA-B*15:01, HLA-B*15:03, HLA-B*35:01, HLA-B*40:02, HLA-B*42:01, HLA-B*44:03, HLA-B*51:01, HLA-B*53:01, HLA-B*13:02, HLA-B*15:07, HLA-B*27:05, HLA-B*35:03, HLA-B*37:01, HLA-B*38:01, HLA-B*41:02, HLA-B*44:05, HLA-B*49:01, HLA-B*52:01, HLA-B*55:01, HLA-C*02:02, HLA-C*03:04, HLA-C*05:01, HLA-C*07:01, HLA-C*01:02, HLA-C*04:01, HLA-C*06:02, HLA-C*07:02, HLA-C*16:01, HLA-C*03:03, HLA-C*07:04, HLA-C*08:01, HLA-C*08:02, HLA-C*12:02, HLA-C*12:03, HLA-C*14:02, HLA-C*15:02, and HLA-C*17:01 alleles.

[00160] In some embodiments, the library comprises greater than or equal to two distinct polypeptide molecules, wherein the antigenic peptide is not the same for each of the greater than or equal to two polypeptide molecules, and wherein each distinct polypeptide is attached to a particle. In some embodiments, the library further comprises a unique defined barcode sequence operably associated with the identity of each distinct polypeptide.

[00161] Embodiments can include barcoded polynucleotides comprising a defined barcode sequence. The barcoded polynucleotides can be a polynucleotide that provides a unique antigen-specific sequence for identification after T cell isolation. Therefore, each unique compACT is attached to a particle with a unique defined barcode sequence. This allows an operative association between a given antigen and a given barcode that is unique to the pair.

[00162] The barcoded polynucleotides can be ssDNA or dsDNA. The polynucleotides comprising the barcodes can be modified at their 5' end to comprise an attachment moiety for attachment to a particle. For example, the polynucleotides comprising the barcode sequences are conjugated to a biotin molecule for binding to a streptavidin-core attached to a particle, such as dextran. However any suitable attachment moiety may be used for attachment of polynucleotides to a particle. As described herein and as understood by a person skilled in the art, suitable attachment moiety pairs are known in the art. Non-limiting examples of attachment moieties include thiol, maleimide, adamantane, cyclodextrin, amine, carboxy, azide, and alkyne.

Particles

[00163] As used herein, “nanoparticles” or alternatively “particles” refer to substrates capable of being specifically sorted or isolated, and to which other entities can be attached. In some embodiments, the nanoparticle is magnetic, e.g., for isolation using a magnet. In some embodiments, the magnetic nanoparticle comprises magnetic iron oxide. Examples of magnetic particles include, but are not limited, to Dynabeads™ (Thermo Fisher). In some embodiments, the nanoparticle is a polystyrene particle, e.g., for isolation by gravity. In other embodiments, the particle can be a surface, a bead, or a polymer. Examples of beads include, but are not limited to, agarose beads and sepharose beads. In particular embodiments, the particle or nanoparticle can be fluorescent or attached to a fluorophore directly or indirectly.

[00164] According to certain embodiments, the nanoparticle is modified with an attachment moiety for attaching additional molecules. Modification of the nanoparticle includes an attachment moiety that can pair with (e.g., covalently bind to) a corresponding cognate (e.g., complementary) attachment moiety attached to polynucleotides. Any suitable pair of attachment moieties may be used to modify the nanoparticle and the polynucleotide detection tag for attachment. Non-limiting examples of attachment moiety pairs include a streptavidin/biotin system, a thiol group (e.g., cysteine) and maleimide, adamantane and cyclodextrin, an amino group and a carboxy group, and an azido group and alkynyl group. In some embodiments, the attachment moiety can comprise a cleavage moiety. In other

embodiments, the attachment moiety bound to complementary cognate attachment moiety can be reversible, such as a reducible thiol group. In an exemplary embodiment, the modified nanoparticle is a streptavidin coated magnetic nanoparticle, such as 1 μ m nanoparticles (e.g., Dynabeads™ MyOne™ Streptavidin T1 beads from ThermoFisher Scientific), and the polynucleotides can be biotinylated for attachment to the modified nanoparticle.

[00165] The particle can be a dextran, such as a biotinylated dextran or streptavidin coated dextran. Modified dextrans are described in further detail in Bethune et al., *BioTechniques* 62:123-130 Mar. 2017 and US Publication No. 2015/0329617, herein incorporated by reference in its entirety. Biotinylated comPACTs can be attached to streptavidin coated dextran.

[00166] The comPACTs can also be assembled into tetramers, comprising 1, 2, 3, or 4 biotinylated comPACT proteins bound to a streptavidin core. The tetramer can also comprise a fluorophore, such as phycoerythrin (PE) or allophycocyanin (APC) bound to the streptavidin core. MHC class I and II tetramers are well known in the art. MHC class I tetramers are described in further detail in Burrow SR et al, *J Immunol* December 1, 2000, 165 (11) 6229-6234 and MHC class II tetramers are described in further detail in Nepom GT, *J Immunol* March 15, 2012, 188 (6) 2477-2482, both of which are herein incorporated by reference in their entirety.

[00167] ComPACT proteins can also be assembled into multimers. In some embodiments, the comPACT protein multimers can be a dimer, trimer, tetramer, pentamer, hexamer, or higher order multimer. In some embodiments, a multimer can comprise at least two or more comPACT proteins. In some embodiments, a multimer can comprise at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 comPACT proteins.

Methods of Producing

Antigen Prediction

[00168] To manufacture a comPACT, one of the initial steps can include identification of the patient's tumor-specific antigens (e.g., neoantigens). The compositions produced by this method can then be utilized in a T-cell mediated immunity process, e.g., for patient-specific cancer immunotherapy. For identification of a patient's putative neoantigens (tumor or pathogen), *in silico* predictive algorithmic programs can be utilized that analyze the tumor, viral, or bacterial sequencing data including whole genome, whole exome, or transcriptome sequencing data, to identify one or more mutations corresponding to putatively expressed neoantigens. Additionally, human leukocyte antigen (HLA) typing can be determined from a

tumor or blood sample of the patient, and this HLA information can be utilized together with the identified putative neoantigen peptide sequences in a predictive algorithm for MHC binding, as verified by Fritsch et al., 2014, *Cancer Immunol Res.*, 2:522-529, the entire contents of which are herein incorporated by reference. HLAs commonly found in the human population can also be included in neoantigen prediction algorithms, such as HLA-A*02, 24, 01; HLA-B*35, 44, 51; DRB1*11, 13, 07 in Caucasians, HLA-A*02, 03, 30; HLA-B*35, 15, 44; DRB1*13, 11, 03 in afro-Brazilians, and HLA-A*24, 02, 26; HLA-B*40, 51, 52; DRB1*04, 15, 09 in Asians. Specific pairing of HLA alleles can also be used. Common alleles found in the human population is further described in Bardi *et al.* (Rev Bras Hematol Hemoter. 2012; 34(1): 25–30.)

[00169] Additional examples of methods to identify neoantigens include combining sequencing with mass-spectrometry and MHC presentation prediction (*e.g.*, US Publication No. 2017/0199961), and combining sequencing with MHC binding affinity prediction (*e.g.*, issued US Patent 9,115,402). In addition, methods useful for identifying whether neoantigen specific T cells are present in a patient sample can be used in combination with the methods described here, *e.g.*, as described in US Publication No. 2017/0003288 and PCT/US17/59598, herein incorporated by reference in their entirety. These analyses result in a ranked list of the patient's candidate neoantigen peptides which can be readily synthesized using routine methods for screening of cognate antigen-specific T cells.

Primer Annealing and Restriction Digest Assembly

[00170] In general, preparation of a comPACT polynucleotide can be accomplished by procedures disclosed herein and by recognized recombinant DNA techniques, *e.g.*, preparation of plasmid DNA, cleavage of DNA with restriction enzymes, ligation of DNA, transformation or transfection of a host, culturing of the host, and isolation and purification of the expressed fusion complex. Such procedures are generally known and disclosed in standard references such as in Sambrook et al., *supra*.

[00171] In some aspects, DNA encoding an MHC class I heavy chain can be obtained from a suitable cell line such as, for example, human lymphoblastoid cells. In various configurations, a gene or cDNA encoding a class I heavy chain can be amplified by the polymerase chain reaction (PCR) or other means known in the art. In some aspects, a PCR product can also include sequences encoding linkers, and/or one or more restriction enzyme sites for ligation of such sequences.

[00172] In some embodiments, a vector encoding a comPACT polynucleotide can be prepared by ligation of sequences encoding the MHC heavy chain and the β 2-microglobulin to a sequence encoding an antigen peptide. DNA encoding the antigen peptide can be obtained by isolating DNA from natural sources or by known synthetic methods, e.g., the phosphate triester method. See, e.g., *Oligonucleotide Synthesis*, IRL Press (M. Gait, ed., 1984). Synthetic oligonucleotides can also be prepared using commercially available automated oligonucleotide synthesizers. A DNA sequence encoding a universal target sequence as discussed herein can be interposed between a sequence encoding a signal sequence and an sequence encoding an antigenic peptide and a second universal target sequence can be interposed between the sequence encoding an antigen peptide segment and a sequence encoding a β 2-microglobulin segment. In some embodiments, the segments can be joined using a ligase. In some embodiments, the sequence encoding an antigen peptide can be phosphorylated with a suitable polynucleotide kinase. In some embodiments the polynucleotide kinase is the T4 polynucleotide kinase. Any appropriate polynucleotide kinase known in the art may be used, including but not limited to T4 polynucleotide kinase, also known as T7 polynucleotide kinase.

PCR assembly

[00173] In some aspects, the comPACT may be assembled via polymerase chain reaction (PCR) amplification. Similar to the restriction digest method, DNA encoding the MHC heavy chain and the β 2-microglobulin may be obtained from a suitable source. A second DNA fragment encoding a chosen signal sequence may also be obtained from a suitable source. Both fragments of DNA may have different universal target sequences, such that primers for one universal sequence do not anneal to the second universal sequence. Two sequences encoding for a chosen antigenic peptide may be synthesized; one forward primer with the antigenic sequence at the 5' end and the complement of the universal primer sequence on the MHC DNA fragment at the 3' end; and one reverse primer with the reverse complement of the chosen antigenic sequence at the 5' end and the reverse complement of the universal primer from the signal sequence fragment at the 3' end. A PCR reaction with all four DNA fragments and primer for the 5' end of the signal sequence fragment and 3' end of the MHC allele fragment will result in amplification of two DNA fragments, one with the signal sequence at the 3' end and the antigenic sequence at the 5' end, and one with the antigenic sequence at the 3' end and the MHC allele at the 3' end. A further PCR amplification cycle will allow the overlapping antigenic peptide sequences to anneal and result in a single full-

length DNA fragment. In some embodiments, the signal peptide fragment further comprises a promoter sequence. In some embodiments, the MHC fragment further comprises a purification cluster and/or a polyA tail.

Transfection, transduction, and genetic modification of host cells

[00174] A comPACT polynucleotide may be inserted into the host cell via an appropriate method known, including, but not limited to, transfection, transduction, electroporation, lipofection, sonoporation, mechanical disruption, or viral vectors. Exemplary transfection reagents include, but are not limited to, Expifectamine, Lipofectamine, polyethyleneimine (PEI), or Fugene. In some examples, Expifectamine is used to transfect mammalian cells with the comPACT polynucleotide.

[00175] A comPACT polynucleotide may be transiently or stably expressed in the host cell. In some embodiments, the comPACT polynucleotide is integrated into the host genome. In other embodiments, the comPACT polynucleotide remains extra-chromosomal. Any appropriate genetic editing technique known in the art may also be employed to modify the host cell with the comPACT polynucleotide, including CRISPR/Cas9, zinc-finger nucleases, or TALEN nucleases.

Expression

[00176] A number of strategies can be employed to express a comPACT polypeptide. For example, the comPACT can be incorporated into a suitable vector by known methods such as by use of restriction enzymes and ligases (see, e.g., Sambrook et al., supra). A vector can be selected based on factors relating to the cloning protocol. For example, the vector can be compatible with, and have the proper replicon for the host that is being employed. Suitable host cells include eukaryotic and prokaryotic cells, and can be cells that can be easily transformed and exhibit rapid growth in culture medium. Examples of host cells include prokaryotes such as *E. coli* and *Bacillus subtilis*, and eukaryotes such as animal cells and yeasts, such as, for example, mammalian cells and human cells. Non-limiting examples of mammalian cells which can be used as hosts to express a comPACT include J558, NSO, SP2-O, 293T, Expi293, and CHO. Other examples of possible hosts include insect cells such as Sf9 or Sf12, which can be grown using conventional culturing conditions. See Sambrook, et al., supra. In various embodiments, cells expressing a comPACT polypeptide can be identified using known methods. For example, expression of a comPACT polypeptide can be determined by an ELISA or Western blot using an antibody probe directed against the MHC

heavy chain portion of the comPACT, or an antibody against an affinity tag, such as His6, or a streptavidin reagent if the comPACT has been biotinylated.

[00177] In some aspects, a comPACT is expressed in mammalian cells. The benefits of expressing protein in mammalian cells instead of in *E. coli* cells are multifold. Protein expressed in *E. coli* cells must be carefully purified away from lipopolysaccharide (LPS). Expression of proteins in mammalian cells results in no LPS contamination of the purified proteins. In addition, mammalian cells are more likely to properly fold mammalian proteins since mammalian cells produce proteins with correct post-translation modifications required for proper folding, including the proper formation of disulfide bonds. In addition, mammalian cells provide the correct chaperone proteins to assist with protein folding in the endoplasmic reticulum or Golgi apparatus. This results in increased purification of homogeneously well-folded proteins, as compared to proteins expressed in *E. coli* cells.

[00178] A comPACT can be substantially-free of LPS. A comPACT can be free of LPS, e.g., a comPACT can have no detectable LPS as measured using LPS-detection methods known in the art. A comPACT can be glycosylated. A comPACT can be modified via expression in a eukaryotic or mammalian cell, e.g., via one or more posttranslational modifications such as glycosylation. A comPACT can include one or more post-translational modifications. A comPACT can (1) be substantially free of LPS or free of LPS; and (2) be glycosylated.

Exemplary ComPACT workflow process

[00179] **FIG. 24** shows an exemplary schematic representation of the assembly and expression of a comPACT protein. Sense and antisense oligos that encode for the desired neoantigen peptide sequence are synthesized and annealed to form a double stranded oligo with overhangs at the 5' and 3' ends, which can then be ligated into a plasmid containing a β 2M gene and an MHC allele. The full comPACT oligo can be amplified into a double stranded amplicon and transfected into cells for protein expression and optional biotinylation. The comPACT protein can be assessed via SDS-PAGE. ComPACTs can then be chosen for scaled up plasmid production in *E. coli*. Protein producer cells are transfected with the selected plasmids and the comPACTs are purified from the producer cells for use in functional assays.

Purification (chromatography)

[00180] An expressed comPACT polypeptide can be isolated and purified by known methods. For example, a comPACT comprising a His6 affinity tag may be purified via

affinity chromatography on an Ni-NTA column by procedures that are generally known and disclosed. Additionally, a comPACT containing human HLA sequences can be purified by affinity chromatography on a monoclonal antibody-Sepharose column by procedures that are generally known and disclosed.

T cell Isolation

[00181] In another aspect, provided herein are methods of isolating an antigen specific T cell, the method comprising the steps of: (a) providing a polypeptide comprising, in an amino terminus to carboxyl terminus orientation, (i) a first universal target peptide, (ii) an antigenic peptide, (iii) a second universal target peptide that is distinct from the first universal target peptide, (iv) a β 2M peptide, and (v) an MHC peptide, wherein the polypeptide is linked to one particle; (b) providing a sample known or suspected to comprise one or more T cells; (c) contacting the polypeptide with the sample, wherein the contacting comprises providing conditions sufficient for a single T cell to bind the polypeptide attached to the particle, and (d) isolating the single T cell associated with the particle.

[00182] Isolation and identification of patient-derived and antigen-specific T cells using a comPACT as described herein can include incubating the comPACT protein with patient-derived T cells. In some embodiments, a library comprising at least two comPACTs can be incubated with patient-derived T cells. T cells can be prepared using standard methods that start from a tissue such as blood, a lymph node, or a tumor.

[00183] Patient-derived T cells can be isolated from the patient's peripheral blood mononuclear cells (PBMCs) or tumor infiltrating lymphocytes (TILs). For example, both CD4⁺ and CD8⁺ T cells can be labeled and sorted from PBMCs or TILs using anti-CD4 and anti-CD8 fluorescent antibodies, with live populations of CD4⁺ and CD8⁺ single-positive cells sorted using fluorescence-activated cell sorting (FACS), to isolate only CD4⁺ or CD8⁺ cells. In some embodiments, T cells that are positive for both CD4 and CD8 can be isolated using an anti-CD3 fluorescent antibody followed by FACS. A person skilled in the art is able to determine the type of T cells to isolate for the type or types of comPACT protein being used.

[00184] Incubation of the comPACT or comPACT library with the T cell suspension allows for a complete and thorough exposure of the particle-bound antigen to the various T-cell receptors. This method may include rocking or rotation of the cells. In some embodiments, the comPACT is associated with a particle.

[00185] Following incubation of the comPACT or comPACT library and the T cells, the bound comPACT-T cell complex is selectively separated or selectively collected. T cells will likely be bound to many identical copies of identical comPACT library elements, and can be separated based on these interactions. For example, if the comPACT comprises a fluorophore, or is attached to a particle with a fluorophore, fluorescent associated cell sorting (FACS), including single-cell sorting, can be used to selectively isolate the T cells. If the comPACT is attached to a magnetic particle, applying a magnet to the suspension can allow for separation of particles complexed with antigen-paired T cells and removal of unpaired T cells. Alternatively, if the particle is a polystyrene particle, the unpaired T cells may be separated by gravity (e.g., centrifugation). After removal of unpaired T cells, in some embodiments, the separated bound particles are washed at least once to remove any non-specifically associated T cells.

[00186] ComPACT-bound T cells can be also separated by FACS into individual collection containers, such as a multi-well plate. The individual collection container can be single-cell reaction vessels. For example, components used for downstream processing and analysis can be added to each single-cell reaction vessel. The comPACT-bound T cells can be separated by FACS into a bulk collection container (e.g., every T cell isolated is collected in the same container).

[00187] ComPACT-bound T cells can also be individually isolated in droplets using a droplet generating microfluidic device (*i.e.*, a “droplet generator”). Droplet generating devices used to encapsulate single cells are known to those skilled in the art, *e.g.*, as described in US Publication No. 2006/0079583, US Publication No. 2006/0079584, US Publication No. 2010/0021984, US Publication No. 2015/0376609, US Publication No. 2009/0235990, and US Publication No. 2004/0180346.

[00188] After isolation of comPACT-bound T cells into single-cell reaction vessels (*e.g.*, isolated in individual well or droplets), the nucleic acid of the comPACT-bound T cell can be further processed for downstream analysis. Specifically, the expressed TCR α and TCR β mRNA transcripts can be first converted to cDNA by reverse transcription and the cDNA amplified for next generation sequencing (NGS) methods known to those skilled in the art, including, but not limited to, sequencing by synthesis technologies (Illumina).

EXAMPLES

Example 1: Design and Cloning of comPACT mini-genes via restriction digest cloning

Structure of comPACT mini-genes for restriction digest:

[00189] The basic exemplary components of a comPACT mini-gene are the signal sequence that directs the secretion of the protein, universal target sequences such as restriction sites or primer binding sites, the antigenic peptide (or neoantigen, NeoE), a second universal target site, the invariant β 2M, the extracellular domain of an MHC allele, and a purification cluster enabling enzymatic modification (e.g. biotinylation) and purification of the comPACT via affinity tags. The cluster may also contain a protease cleavage site and linker sequences between the peptide components. The mini-gene may also contain cysteine mutations that act as a disulfide trap. A diagram of a comPACT mini-gene is shown in **Figure 1**. Additional restriction sites upstream and downstream of the MHC heavy chain sequence can be used to insert other MHC alleles to construct different MHC templates and build a library of MHC templates (**Figure 2**). The DNA fragments encoding the signal sequence, universal target sequences, invariant β 2M, and the extracellular domain of an MHC allele are the base MHC template.

[00190] For restriction digest cloning methods, each comPACT DNA construct is a base MHC template with a dummy antigenic sequence insert containing stop codons in three frames and a unique restriction site for destruction of uncut or re-ligated template (**Figure 3**) and can be used as part of an off-the-shelf platform for rapidly assembling libraries of antigenic peptides complexed with that MHC allele. The MHC alleles may also be modified or mutated (e.g., Y84A or Y84C) to improve folding or increase binding of the antigenic peptide with the MHC protein. In addition, the β 2M protein can also be mutated (e.g., S88C) to allow it to bind thiol dyes.

[00191] In this example, a comPACT mini-gene is shown with the following structure: a NotI restriction site at the 5' end; the signal sequence from human growth hormone, hGH, shown in **Table 3**; a restriction site BlnI upstream of the antigenic peptide region and a BamHI restriction site downstream of the antigenic peptide sequence, shown in **Table 2**; a linker sequences of predominantly glycine and serine residues (i.e. Gly-Ser linkers); the β 2M sequence; a second Gly-Ser linker sequence with a BspI restriction site; a MHC heavy chain; a third Gly-Ser linker sequence with a BstBI restriction site; and a purification cluster with an AviTag sequence, a TEV cleavage site, and a concatenated histidine tag.

Restriction digest cloning and assembly of comPACT mini-gene

[00192] Three different methods of inserting the neoantigen via restriction digest are described herein. In the first, shown as a diagram in **Figure 4**, the antigenic peptide (NeoE)-encoding primer spans the first restriction site (BlnI in this example) at the 5' end and the second restriction site (BamHI in this example) at the 3' end. This primer amplifies off a universal reverse primer encoding the second restriction site, yielding a primer dimer of ~70 bp.

[00193] In the second method, the antigenic peptide-encoding primer spans the second restriction site as the 5' end and is the reverse complement of the antigen. This primer primes in reverse orientation off the template DNA encoding the signal sequence. Paired with a forward primer spanning the first restriction site sequence, this reaction yields a 70 bp product, or a ~140 bp product if a forward primer spanning a restriction site farther upstream of the antigen site is used.

[00194] In both the first and second methods, the insert is cleaned up on a commercial column, digested with appropriate restriction enzymes, cleaned again on a commercial column, and then ligated with a pre-digested MHC template in a vector. Ligation reactions are transformed into *E. coli* and plasmids prepared from transformed *E. coli* are used in mammalian producer cell transfection reactions.

Example 2: Design and Cloning of comPACT mini-genes via primer annealing

[00195] In a third variation on MHC template vector ligation, PCR and restriction digestion was bypassed by annealing two reverse complementary neoantigen-encoding primers. These primers were designed to have 5' and 3' ends that begin and terminate in complementary sequences that simulate the overhangs from restriction digestion (**FIG. 5**). The sense and antisense primers were incubated with T4 polynucleotide kinase and ATP to phosphorylate the 5' ends (**FIG. 22A**). When these primers annealed to each other, they formed a double stranded oligonucleotide sequence that has overhang nucleotides as if it had been digested with a restriction enzyme. The phosphorylated neoantigen insert was ligated into a pre-cut MHC template in a vector. The comPACT minigene had the same structure as that described in Example 1. The ligation product was then used for PCR amplification of a linear comPACT amplicon using bookend universal primers to amplify the complete comPACT gene and sequenced. 824 comPACTs with unique neoantigen sequences were made using this method, with greater than 99% of the generated comPACTs having the correct neoantigen sequence (**FIG. 22B**).

[00196] Next, *E.coli* were transformed with the ligation product plasmids and plated onto selective agar plates containing ampicillin. Individual colonies were picked and grown overnight for plasmid purification and sequenced for full gene verification. After sequencing verification, plasmid lots were archived and propagated into larger quantities.

[00197] Alternatively, T4 kinase is not used if the precut MHC template vector retains 5' phosphates on its overhang ends. The annealed antigen insert can then be ligated with the cut MHC vector and the ligation product transformed into *E. coli* for plasmid production.

Example 3: Design and Cloning of comPACT mini-genes via PCR assembly

Structure of comPACT mini-genes for PCR assembly:

[00198] A fourth method of inserting the neoantigen may also be used. In this method, the neoantigen is inserted into the MHC template which is flanked by an upstream promoter and a downstream polyadenylation signal via polymerase chain reaction to form a 2.5 kb mini-gene. A diagram of the PCR assembly reaction is shown in **Figure 6**.

[00199] In this example, a comPACT mini-gene is shown with the following structure: a promoter at the 5' end; a signal sequence with a first universal target sequence; the antigenic peptide; a second universal target sequence with a linker sequence of predominantly glycine and serine residues (i.e. GlySer linkers); the β 2M sequence; a second Gly-Ser linker sequence; an MHC heavy chain allele; a third Gly-Ser linker sequence; a purification cluster; and a polyA sequence. The universal target sequences are not the same in this method.

PCR assembly of comPACT mini-genes:

[00200] In this method, two primers (<60 nt) with a chosen neoantigen sequence are synthesized. The first primer has the neoantigen sequence at the 5' end followed by a stretch of the second universal target sequence at the 3' end. The second primer has the reverse complement of the neoantigen sequence at the 5' end and the reverse complement of the first universal sequence at the 3' end. These primers are mixed with a DNA fragment encoding the promoter region, signal sequence and first universal target sequence, and another DNA fragment encoding the second universal target sequence, the β 2M sequence, MHC allele, purification cluster and a polyA sequence. Each antigenic peptide primer anneals to its complementary sequence and a PCR reaction is run that amplifies the neoantigen sequence onto either the promoter fragment or the MHC allele fragment. These two newly synthesized fragments now each have the neoantigen sequence. Further PCR reactions, along with primers for the 5' end of the promoter sequence and 3' end of the polyA sequence, allow the

neoantigen sequences to anneal to each other and prime the assembly of a full length linear comPACT amplicon.

[00201] The fully assembled linear comPACT polynucleotide is then cleaned up for direct transfection into mammalian producer cells, bypassing the steps using *E. coli* and plasmid production altogether.

Example 4: Expression and purification of comPACT proteins from plasmids

Expression of protein

[00202] Neoantigen12 (neo12) was ligated into an HLA-A2 template sequence and inserted into an expression plasmid (pPACT0010) via restriction digest of the NotI and BamHI restriction sites and ligation as previously described.

[00203] Expi293 mammalian producer cells in a 30 mL shake flask volume were transfected with pPACT0010 incubated with Expifectamine transfection reagent on day -1. Enhancers included in the Expifectamine transfection kit were added on day 0. Samples were collected from the cell supernatant on days 1 to day 7 and assessed for secreted protein via SDS-PAGE and total protein staining using Safestain (ThermoFisher). Levels of secreted comPACT protein increased until day 3, at which point the protein secretion leveled off (**Figure 7**). Secreted comPACT protein was initially identified by its apparent molecular weight (=53 kDa) and confirmed by a Western blot using NTA-HRP to detect the His6 affinity tag.

Purification of protein

[00204] The Neo12 comPACT protein collected on day 7 was purified by Ni-NTA affinity chromatography via binding of the His6 affinity tag. Samples were assayed for total protein via SDS-PAGE and Safestain. The lack of comPACT protein in the flow-through (FT) fraction of the affinity column confirmed that the His6 tag was not cleaved during expression and purification (**Figure 8**). The purified yield was >400 mg per L culture volume.

[00205] The Neo12 comPACT protein was biotinylated (discussed below in **Example 6**) and further purified by size-exclusion chromatography. A single major peak was observed, suggesting the protein was properly-folded and monomeric, with little aggregation (**Figure 9**). The second peak is ATP, which was added for the BirA-catalyzed biotinylation reaction.

Optimization of Production Volume and Parallel Production

[00206] The production of comPACTs was scaled down from a culture volume of 30 mL in a shake flask to 0.7 mL in a 96 deep-well shake block. Expi293 mammalian producer cells were transfected with plasmid DNA containing the pPACT0010 plasmid, and the secreted

Neo12 comPACT protein was purified as previously described. 437 mg/L of purified Neo12 comPACT protein was collected from a 0.7 mL well volume as compared to the previously described yield of >400 mg/L from the 30 mL purification experiment (**Figure 10**). The protein yield from the 0.7 mL experiment corresponds to >300 micrograms of protein, or ~1000-fold more than is typically needed for a typical flow cytometry experiment.

[00207] Next, parallel expression of multiple comPACT constructs was assessed. Eight different comPACT constructs with different neoantigens (neoantigens 10, 15, 64, 65, 66, 67, 80, and 83) were expressed in 30 mL shaker flasks as a mid-throughput assay (**figure 11**). Each comPACT construct was transfected into cells as previously described where the comPACT protein was expressed and secreted into the cell supernatant. The expressed protein was purified as previously described, concentrated, and normalized. Samples of crude supernatant and concentrated proteins were assayed for total protein as previously described. The comPACT proteins were purified via size exclusion chromatography (**Figure 12**). A single peak, containing 2-20 mg of protein, was seen for each protein, also suggesting that the comPACT proteins were properly-folded and monomeric.

Example 5: Expression and purification of comPACT proteins from linear amplicons

[00208] In the previous examples, comPACT proteins were expressed from plasmids transfected into mammalian producer cells. As an alternative approach, linear amplicons of the neo12 comPACT mini-gene (neoantigen 12 assembled into a mini-gene with the HLA-A2 template sequence) flanked by a promoter sequence and a polyA sequence were transfected into 0.7 mL of the producer cells in a 96-deep well plate. As a control, the pPACT0010 plasmid was also transfected into separate producer cells. Protein from both samples was expressed, purified and assayed for total protein as previously described. Similar levels of expressed proteins were produced by both the linear amplicon and the plasmid (**Figure 13A**), suggesting that the protein encoded by a comPACT mini-gene can be produced without the use of a plasmid intermediate. Multiple different comPACT mini-genes with different neo-epitope sequences have been produced (**Figure 13B**) for direct transfection of producer cells.

[00209] Additional comPACTs with different HLA alleles were made using the annealing and phosphorylation workflow described in Example 2. Linear amplicons were derived from the expression vector using bookend PCR and universal primers, and were transfected into Expi293F cells for comPACT protein production. **FIG. 23** shows a protein gel of representative comPACT proteins made using the annealing cloning workflow. ComPACT

proteins were run on an SDS-PAGE gel and stained with Safestain (ThermoFisher).

Example 6: Biotinylation of comPACT proteins

In vitro biotinylation of comPACTs with isolated BirA enzyme

[00210] The comPACT purification cluster includes a BirA recognition sequence (Avitag) for biotinylation. Purified comPACT proteins were unbiotinylated (No BirA treat) or biotinylated with commercial BirA protein according to the manufacturer's instructions (BirA-treated). Following overnight BirA enzymatic treatment, samples were bound to two different types of magnetic streptavidin beads (C1 and T1) and incubated to allow the biotinylated protein bind to the streptavidin beads. The supernatant (SN) and beads ("pellet," P) were separated via SDS-PAGE. Samples were assayed for total protein with Safestain and the presence of comPACT protein via Western Blot with NTA-HRP (**Figure 14**). In the untreated samples, the comPACT protein was mainly found in the SN fraction, confirming that it was unbiotinylated. In the biotinylated samples, the comPACT protein was found in the pellet samples of both C1 and T1 streptavidin beads, although the interaction between the biotinylated proteins and the C1 streptavidin beads was the most complete. Biotinylated comPACT protein was not detected via Western Blot in the C1 streptavidin bead-depleted supernatant, suggesting ~100% of comPACT protein was biotinylated.

[00211] ComPACT proteins may also be biotinylated in the clarified supernatant, prior to purification. Multiple comPACT proteins were expressed in producer cells as previously described. The cell culture supernatant was collected and clarified via centrifugation. The clarified supernatant was treated with commercial BirA protein according to the manufacturer's instructions and then purified via Ni-NTA affinity chromatography and biotinylation was assessed via Western Blot (**Figure 15**). All comPACT proteins tested were biotinylated using this method, indicating that biotinylation of comPACT proteins in clarified cell supernatants is effective.

[00212] To produce enough BirA for high-throughput biotinylation of comPACT proteins, a BirA protein with a His6 tag was expressed in *E. coli* cells. This His6 tagged BirA was purified via Ni-NTA affinity chromatography (**Figure 16B**) and can be used to biotinylated the comPACT proteins. A second version of BirA-His6 with a TEV-cleavable His6 tag was also expressed and purified via Ni-NTA affinity chromatography (**Figure 16C**). This BirA-TEV-His6 protein can be purified via Ni-NTA, the His6 tag removed via TEV cleavage, and the tagless BirA then used to biotinylated comPACT proteins. After

biotinylation of the comPACT proteins, the tagless BirA protein can then be purified away via Ni-NTA affinity chromatography. In addition, TEV protease was expressed heterologously in *E. coli* for use with the BirA-TEV-His6 (**Figure 16A**) for use in biotinylated comPACT protein production.

[00213] Cleavage of the His6 tag on comPACT proteins after biotinylation was also assessed and the results shown in **Figure 17**. ComPACT proteins were treated or untreated with BirA to biotinylate them as previously described (lanes 1 and 2 of **Figure 17**). A third sample of comPACT protein was treated with BirA and then with TEV to cleave the His6 tag present on the protein (lane 3). Samples were separated via SDS-PAGE, and total protein was assessed via Safestain. All three samples had equal amounts of comPACT protein.

Biotinylation of the comPACT proteins and cleavage of the His6 tag was assessed via Western blot using an SA-HRP reagent for the biotin signal and an NTA-HRP reagent for the His6 tag. The unbiotinylated sample did not show biotin signal, but did have a His6 signal and the biotinylated and uncleaved sample had both signals. The biotinylated and TEV cleaved sample only had the biotin signal, indicating that the His6 tag was successfully cleaved off the comPACT protein.

In vivo biotinylation of comPACTs in cells expressing BirA

[00214] A third approach for biotinylating the comPACTs is to express BirA in the Expi293 producer cells and biotinylate the comPACTs *in vivo* prior to purification. Expi293 cells were transduced with a lentiviral vector to co-express BirA flanked by V5 which acts as a cell surface transduction marker. Transduced cells sorted for V5+ also express BirA (**Figure 18**). These cells can be used to produce biotinylated comPACTs *in vivo* before comPACT protein purification. Two cell lines were generated, which secrete BirA into the media (Expi293v0223) or express BirA on the cell surface (Expi293v0263). A third cell line expressing BirA flanked by the endoplasmic reticulum (ER) retention sequence KDEL was also made.

[00215] To assess biotin labeling efficiency under the *in vitro* or *in vivo* biotinylation approaches, a streptavidin pull down assay was performed. ComPACT protein (EBV or MART-1) were biotinylated either by treating the clarified media with BirA enzyme and the proper reactions components (enzymatic, *in vitro*) or by co-expressing the comPACT with BirA (*in vivo*) and incubated with increasing concentration of Streptavidin coated beads. Magnetic beads coated with Streptavidin (Dynabeads, Thermo Fisher) were added to 20 µg of the comPACT protein samples and incubated 30 min at room temperature. Magnetic beads

were isolated at the bottom of the tube or well by magnetization and the protein content of the supernatant was assayed by SDS-PAGE. **FIG. 25** shows the SDS-PAGE gel results. Near complete biotinylation can be achieved by adding the BirA enzyme during purification (in vitro) or by co-expressing the BirA enzyme and comPACT protein (in vivo). Samples lacking streptavidin beads had significant amounts of biotinylated protein in the supernatant, while addition of increasing amounts of streptavidin beads resulted in almost all the protein being removed from the supernatant, indicating biotinylation of the comPACT protein.

Example 7: Antigen-specific T cell staining and affinity evaluation using comPACT proteins

[00216] To compare antigen-specific T cell staining using comPACTs and conventional peptide-MHCs, comPACT dextramers were prepared according to a published protocol (Bethune, M. T., et al. *BioTechniques* **62**, 123-130, doi:10.2144/000114525 (2017)). T cells were engineered to express an A2/neo12-specific TCR and stained with either HLA-A2/neo12 peptide-MHC dextramers or HLA-A2/neo12 peptide comPACT dextramers. Staining with the comPACT dextramers was at least as efficient as that for peptide-MHC dextramers (**Figure 19**). This data suggests that comPACT dextramers can be used to sort antigen-specific T cells for TCR sequencing.

Example 8: Functional T cell assays

NTAmer binding assay

Materials and Methods

[00217] Beyond antigen-specific capture of T cells, the modular design and ease-of-production of comPACTs facilitates their use in functional T cell assays. For example, incorporation of a mutated version (S88C) of β 2M enables comPACTs to be labeled with a maleimide-dye conjugate, assembled as NTAmers, and used to measure kinetic parameters of TCR-comPACT binding. For instance, NTAmers can be used to resolve monovalent TCR-MHC I binding events in live cells. CD8 binding and multiple TCR-MHC I interactions at immune synapse allow for extended contact between T cells and antigen presenting cells. Fluorescent dye conjugated-ComPACT NTAmers are incubated with cells to allow for binding of the comPACT with the TCR. The Ni-PE particle component of the NTAmer is dissociated from the comPACT via the addition of imidazole, and the release of the fluorescent-comPACT is measured over time.

Dye Conjugation

[00218] S88C mutant comPACT proteins were constructed and expressed at ~150 mg/L. These mutant comPACTs exhibit similar purity and elution profiles as un-mutated comPACTs (**Figure 20**). Other dyes, such as Cy5, can also be conjugated to S88C comPACTs. The comPACT88C-HisTag monomer was reduced with Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and coupled with maleimide-Cy5, resulting in a comPACT88c-Cy5 labeled monomer. **Figure 21** shows the A280, A260, and A650 quantification of comPACTs labeled with Cy5. The dye to comPACT ratio was 67%.

NTAmer assembly

[00219] The biotinylated comPACTs were bound to PE-Biotin-NTA bead by charging biotin-NTA with Nickel (Ni) (biotin-NiNTA) and then by assembling comPACT-Cy5 with biotin-NiNTA and PE-streptavidin (SA). Biotin-NiNTA was generated by mixing 5 mg biotin-NTA with 1mL Nickel charging solution (50 mM NiSO₄ in 100 mM HEPES pH7.5) to yield Ni²⁺+NTA-biotin (7nM). The solution was diluted with HBS to yield a 70 uM Ni²⁺+NTA-biotin solution. The Ni²⁺+NTA-biotin solution was combined with SA-PE (300kDa) in 5 additions, 5 min between each addition. 20uM Cy5-S88C comPACT was then added to NTAmer core, incubated for 10 min at room temperature in the dark and stored at 4C. Both Neo12 and MART1 comPACTs were made and assembled into NTAmers. The His tag on the comPACT protein binds to the NiNTA which itself binds to the PE-streptavidin. Each NTAmer comprises multiple copies of a comPACT monomer bound to a streptavidin core with a fluorophore. Assembly of NTAmers are further discussed in Schmidt et al, *J. Biol Chem*, December 2, 2011, 286 (48) 41723-41735, and Schmidt et al, *Front Immunol*, July 30, 2013 doi: 10.3389/fimmu.2013.00218, both of which are hereby incorporated by reference in their entirety.

[00220] Biotinylated Neo12 and MART1 comPACT proteins were also assembled into PE-labeled tetramers and dextramers. Peptide-bound refolded MHC I molecules with neo12 antigenic peptides were also assembled into PE-labeled dextramers as controls. Methods of assembling peptide-MHC proteins into tetramers and dextramers are generally known in the art.

T cell binding

[00221] Gene edited T cells that express TCRs that bind Neo12 or MART1 neogantigens were resuspended in 1x Stain Buffer (BD Bovine serum albumin Stain Buffer, BD554657) at a density between 1-2 x 10⁶ cells/mL. Cells were kept at 4°C for duration of assay. NTAmers were added at various concentrations from 1:50-1:400 (v:v dilution) to the T cells in stain

buffer and incubated for 15 min at 4°C in the dark to allow multimerized NTAmer to bind the T cells. The T cells were washed twice in 1x stain buffer to remove unbound NTAmer and resuspended in stain buffer. To disrupt multimerized NTAmer and monitor Cy5 signal decay, imidazole disruption buffer (1.452 g imidazole and 50 mL 1x Hank's Buffered Saline) was added, mixed, and the samples were assessed via flow cytometer to monitor Cy5 signal decay. Additional antibody staining for the detection of various cellular markers was performed by incubating T cells for 15 min at 4°C in dark, wash cells twice in 1x stain buffer. T cells were then fixed using IC Fixation Buffer (eBioscience IC Fixation Buffer, 00-8222-49).

Results

[00222] First, the ability of the biotinylated compACT-Cy5 NTAmers to bind their cognate TCR was assessed. Neo12 peptide-bound refolded MHC I molecules, neo12 compACT molecules, or neo12 NTAmers were assembled into dextramers and incubated with T cells expressing the neo12 TCR. Binding of each dextramer molecule to the cells was determined. **FIG. 28A** shows that gene edited T cells expressing neo12 TCRs demonstrated specific binding to neo12 peptide-bound refolded MHC I molecule dextramers (lower plot, x-axis is the dextramer PE signal) and this binding corresponded to the T cell population expressing CD3 at their surface (top plot, x-axis is the PE-Cy3 signal). **FIG. 28B** shows that gene edited T cells expressing neo12 TCR demonstrated specific binding to neo12 compACT molecule dextramers (lower plot, x-axis is the dextramer PE signal) and this binding corresponded to the T cell population expressing CD3 at their surface (top plot, x-axis is the PE-Cy3 signal). **FIG. 28C** shows that gene edited T cells expressing neo12 TCR demonstrated specific binding to neo12 compACT NTAmer dextramers (lower plot, x-axis is the dextramer PE signal) and this binding corresponded to the T cell population expressing CD3 at their surface (top plot, x-axis is the PE-Cy3 signal).

[00223] Next, the ability of the biotinylated compACT-Cy5 NTAmers to bind TCRs in an antigen specific manner was confirmed. Neo12 or F5 TCR T cells were incubated with NTAmers with neo12 antigen. F5 TCR T cells were used as a negative control. Neo12 TCR T cells were also incubated with a dextramer composed of neo12 compACT proteins. The bound T cells were then incubated with imidazole as described and binding of the NTAmers and dextramers was assessed as previously described. The addition of imidazole inhibited the ability of the neo12-ComPACT NTAmer to bind to neo12 T cells (**FIG. 29A-C**). Neo12 TCR T cells showed specific binding to neo12 NTAmer reagent (**FIG. 29A**, NTAmer Cy5 signal is the y-axis). The neo12-ComPACT NTAmer is specific to Neo12 T cells and did not bind to T cells expressing F5 TCRs (**FIG. 29B**). Neo12 compACT dextramers also bound to the neo12

TCR T cells (**FIG. 29D**). Neo12 TCR T cells showed no binding to neo12 NTamer reagent when incubated in the presence of imidazole (**FIG. 29C**) demonstrating the need for a multimerized reagent to observe Cy5 signal from S88C modified comPACT molecules.

[00224] MART-1 NTamers were also used to show antigen specific binding of T cells edited to express F5 TCR (**FIG. 30A-D**). **FIG. 30A** shows binding of MART-1 comPACT dextramer to F5 TCR T cells. **FIG. 30B** shows binding of MART-1 comPACT dextramer to M1W TCR T cells. **FIG. 30C** shows binding of MART-1 comPACT NTamers to F5 TCR T cells. **FIG. 30D** shows binding of MART-1 comPACT NTamers to M1W TCR T cells. The low levels of MART-1 comPACT NTamer binding to M1W T cells is likely due to the low affinity of the M1W TCR for the cognate antigen:MHC I protein.

[00225] The monomeric TCR:pMHC-Cy5 disassociation was assessed by measuring the Cy5 signal decay of two TCRs, MART-1 F5 and M1W, with differential TCR:MHC I binding affinities against the same peptide:MHC I complex (HLA:A02+MART-1). Binding of MART1 NTamers to T cells expressing F5 TCRs was stable over time in the absence of imidazole induced NTamer disruption (**FIG. 31A**). However, disruption of the multimerized MART-1 NTamers with imidazole resulted in a time dependent disassociation of monomeric MART-1 comPACT molecules to the F5 TCRs (**FIG. 31B**). **FIG. 31C** shows a zoomed in view of a subset of time data from **FIG. 31B**.

[00226] In conclusion, Neo12 and F5 T cells demonstrate similar binding levels between NTamers and equivalent comPACT dextramers and comPACT NTamers are able to bind neoantigen T cells in an antigen specific manner. Neo12 and F5 T cells demonstrate similar binding levels between NTamers and comPACT dextramers for both neo12 and MART-1 comPACTs. In addition, the NTamer complex can be disrupted by imidazole addition and monovalent TCR:MHC I off-rates can be measured by the decay of comPACT-Cy5 signal on live T cells. This decay takes place over a tens of seconds time scale.

Example 9: comPACT Library Production

[00227] HLA allele diversity across the US human populations was analyzed from the Allele Frequency Net Database (www.allelefrequencies.net) by bioinformatics to identify the optimal number of alleles to include in the HLA repertoire to effect high coverage of subject HLA frequencies. 9736 alleles were analyzed. **FIG. 26A** show the analysis of the percentage of patients in which one or both alleles from each of HLA A, B, and C loci are covered by a library of 66 HLA alleles. Solid lines indicate 1 allele is covered, while dashed lines indicate both alleles are covered. 66 alleles enables the coverage of at least 4 of 6 HLA alleles per patient in >95% of total population and 6/6 alleles in >80% of population (**FIG. 26B**). The

most frequent HLA-I allele is HLA-A02:01 with ~50% US prevalence. HLA libraries first shown herein therefore allow the most potential for broad implementation of personalized neoTCR-T cell therapies for a global and diverse population.

[00228] Next, a library of comPACT proteins with different neoepitopes and selected HLA alleles was made. Neoepitope candidates were chosen from the Immune Epitope Database (www.iedb.org). Full sequences for each of the 66 HLA-I alleles in the repertoire were obtained from the IMGT database and modified to include the Y84C mutation. All clones were sequence verified and banked in the database and reagent inventory. Ten neoepitope peptide were chosen from the IEDB database and inserted into a panel of 36 HLA alleles. ComPACTs of the selected neoepitopes and HLA alleles were expressed and purified via Size Exclusion Chromatography column (Agilent Sec Bio 300) connected to an Agilent Infinity II HPLC system (SEC-HPLC) according to the manufacturer's instructions. The results are shown in **FIG. 27A-C**. the comPACTs were purified as monodisperse polypeptides, as assessed via SEC-HPLC by measuring the area under the curve of the monomer peak divided by the area under the whole chromatogram (**FIG. 27A and 27B**) Most comPACTs were expressed at a high titer (**FIG. 27C**). At least one comPACT protein for each HLA allele described has been purified and characterized via HPLC, indicating that the comPACT platform is robust and amenable to many alleles.

[00229] While the invention has been particularly shown and described with reference to a preferred embodiment and various alternate embodiments, it will be understood by persons skilled in the relevant art that various changes in form and details can be made therein without departing from the spirit and scope of the invention.

[00230] All references, issued patents and patent applications cited within the body of the instant specification are hereby incorporated by reference in their entirety, for all purposes.