



(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/US2024/044819

(22) International Filing Date:
30 August 2024 (30.08.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/580,300 01 September 2023 (01.09.2023) US

(71) Applicant: **CHILDREN'S HOSPITAL MEDICAL CENTER** [US/US]; 3333 BURNET AVE., CINCINNATI, Ohio 45229 (US).

(72) Inventors: **SALOMONIS, Nathan**; c/o Children's Hospital Medical Center, 3333 Burnet Ave., Cincinnati, Ohio 45229 (US). **LI, Guangyuan**; c/o Children's Hospital Medical Center, 3333 Burnet Ave., Cincinnati, Ohio 45229 (US).

(74) Agent: **KUAN, Roger** et al.; Norton Rose Fulbright US LLP, 1301 McKinney, Ste. 5100, Houston, Texas 77010-3095 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ,

(54) Title: IDENTIFICATION OF TARGETS FOR IMMUNOTHERAPY IN MELANOMA USING SPLICING-DERIVED NEOANTIGENS

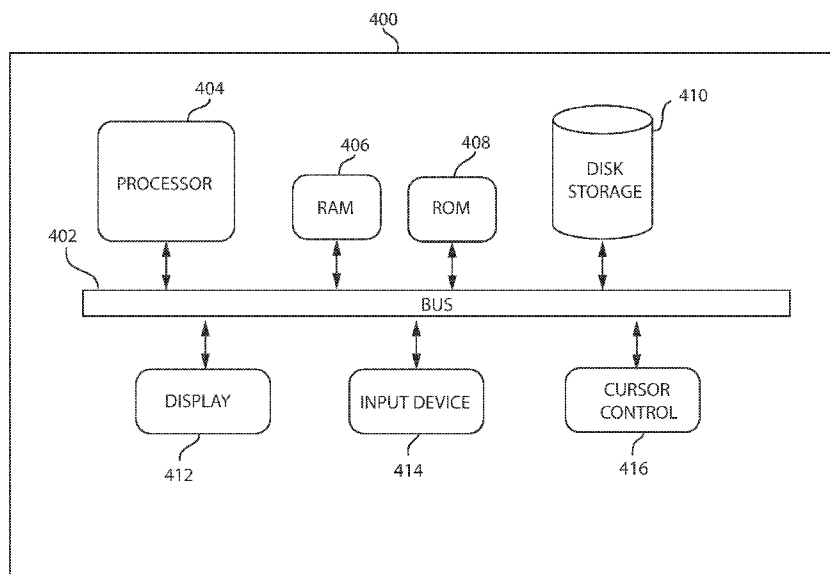


FIG. 1

(57) Abstract: Immunogenic targets associated with one or more types of cancer, wherein the immunogenic target includes a shared splicing amino acid-based neoantigen present in any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor, are described, as well as uses thereof, and methods of identifying such immunogenic targets. Particular embodiments relate to immunotherapies for treating a type of cancer, and methods of treating a type of cancer by administering such immunotherapies, wherein the immunotherapy includes an immunogenic target including a shared splicing amino acid-based neoantigen present in any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor.



DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT,
LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE,
SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
- *of inventorship (Rule 4.17(iv))*

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
- *with sequence listing part of description (Rule 5.2(a))*

IDENTIFICATION OF TARGETS FOR IMMUNOTHERAPY IN MELANOMA USING SPLICING-DERIVED NEOANTIGENS

STATEMENT REGARDING FEDERALLY-SPONSORED RESEARCH

This invention was made with government support under Grant No. R01 CA226802 awarded by the National Institutes of Health (NIH). The government has certain rights in the invention.

CROSS REFERENCE TO RELATED APPLICATION

[0001] The present application claims the benefit of priority under 35 U.S.C. § 119(e) to U.S. Provisional Application No. 63/580,300, IDENTIFICATION OF TARGETS FOR IMMUNOTHERAPY IN MELANOMA USING SPLICING-DERIVED NEOANTIGENS, filed on September 1, 2023, which is currently co-pending herewith and which is incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] The disclosure generally relates to methods of treating cancer, and in particular to computational methods for identifying molecular targets for use in cancer immunotherapy, as well as to the molecular targets per se and uses thereof.

BACKGROUND

[0003] Immunotherapy has emerged as a promising strategy to combat cancer by ‘reprogramming’ a patient’s own immune system. However, no current effective strategies exist to predict which cancer patients will respond to conventional immunotherapy (e.g., anti-PD-1). Effective targeted immunotherapies should accurately predict which cancer-specific neo-peptides (aka neoantigens) are most likely to elicit an immune response.

[0004] Numerous companies are currently developing personalized strategies for targeted immunotherapy focused on gene mutations specific to a cancer cell. However, RNA-based variants (i.e., alternative splicing events) unique to cancer cells have not been exploited in a similar way, clinically. There is a need for tools to efficiently and accurately identify cancer neoantigens for targeted therapies, such as immunotherapies.

SUMMARY OF THE INVENTION

[0005] Various embodiments of the disclosure relate to immunogenic targets associated with one or more types of cancer, wherein the immunogenic target includes a shared splicing amino acid-based neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor.

[0006] In some embodiments, the target includes a shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2241. In some embodiments, the target includes a shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-4. In some embodiments, the target includes a shared splicing neoantigen present in a protein having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2545. In some embodiments, the target includes a shared splicing neoantigen present in a protein having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2245. In some embodiments, the target includes a nucleotide coding for a shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2545. In some embodiments, the target includes a nucleotide coding for a shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-4 or SEQ ID NOs 2242-2245. In some embodiments, the immunogenic target includes at least 90% sequence homology to any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor. In some embodiments, the immunogenic target is *in vitro*.

[0007] In some embodiments, the target includes a shared splicing neoantigen selected from SEQ ID NOs: 1-2545. In some embodiments, the target includes a shared splicing neoantigen selected from SEQ ID NOs: 1-2241. In some embodiments, the target includes a shared splicing neoantigen selected from SEQ ID NOs: 1-4. In some embodiments, the target includes a shared splicing neoantigen selected from SEQ ID NOs: 2242-2545.

[0008] In some embodiments, the target is a shared splicing neoantigen selected from SEQ ID NOs: 1-2545. In some embodiments, the target is a shared splicing neoantigen selected from SEQ ID NOs: 1-2241. In some embodiments, the target is a shared splicing neoantigen selected from SEQ ID NOs: 1-4. In some embodiments, the target is a shared splicing neoantigen selected from SEQ ID NOs: 2242-2545.

[0009] Further embodiments of the disclosure include cells having any of the aforementioned immunogenic targets.

[0010] Additional embodiments of the disclosure include immunotherapies for treating one or more types of cancer, wherein the immunotherapy is specific for, or includes, one or more of the aforementioned immunogenic targets.

[0011] In some embodiments, the immunotherapy is specific for, or includes, one or more shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2241, or one or more nucleotide coding therefor, and the type of cancer includes skin cancer, breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS). In some embodiments, the immunotherapy is specific for, or includes, one or more shared splicing neoantigen selected from SEQ ID NOs: 1-2241, or one or more nucleotide coding therefor, and the type of cancer includes skin cancer, breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS). In some embodiments, the immunotherapy is specific for, or includes, one or more shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-4, or one or more nucleotide coding therefor, and wherein the type of cancer includes skin cancer. In some embodiments, the immunotherapy is specific for, or includes, one or more shared splicing neoantigen selected from SEQ ID NOs: 1-4, or one or more nucleotide coding therefor, and wherein the type of cancer includes skin cancer. In some embodiments, the immunotherapy is specific for, or includes, one or more shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2545, or one or more nucleotide coding therefor, and the type of cancer includes skin cancer, breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS). In some embodiments, the immunotherapy is specific for, or includes, one or more shared splicing neoantigen selected from SEQ ID NOs: 2242-2545, or one or more nucleotide coding therefor, and the type of cancer includes skin cancer, breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS). In some embodiments, the immunotherapy is specific for, or includes, one or more shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2245, or one or more nucleotide coding therefor, and the type of cancer includes skin cancer and/or head and neck cancer. In some embodiments, the immunotherapy is specific for, or includes, one or more shared splicing neoantigen selected from SEQ ID NOs: 2242-2245, or one or more nucleotide coding therefor, and the type of cancer includes skin cancer and/or head and neck cancer.

[0012] In some embodiments, the immunotherapy includes an RNA-, DNA-, or peptide-based vaccine, dendritic cell vaccine, chimeric antigen receptor T cell (CAR-T) therapy, T cell receptor-engineered T cell (TCR-T) therapy, antibody or antibody derivative, cell-based immunotherapy, immune checkpoint inhibitor, immunomodulator, topical immunotherapy, injection immunotherapy, adoptive cell transfer, oncolytic virus therapy, immunosuppressive drug, helminthic therapy, and/or other non-specific immunotherapy. In some embodiments, the immunotherapy includes an RNA-, DNA-, or peptide-based vaccine, dendritic cell vaccine, CAR-T therapy, and/or antibody or antibody derivative.

[0013] Further embodiments of the disclosure include compositions including any of the aforementioned immunogenic targets and/or immunotherapies. In some embodiments, the compositions can be formulated as a vaccine. In some embodiments, the compositions can further include one or more cytokines, growth factors, or adjuvants.

[0014] Further embodiments of the disclosure include *in vitro* isolated dendritic cells including any of the aforementioned immunogenic targets, and methods of making the same. For example, a dendritic cell (such as a dendritic cell vaccine) can be prepared by contacting a mature dendritic cell *in vitro* with any of the aforementioned immunogenic targets, and optionally screening the dendritic cell for one or more cellular properties. In some embodiments, the dendritic cell includes a mature dendritic cell. In some embodiments, the cell can be a cell with an HLA type selected from HLA-A, HLA-B, or HLA-C. In some embodiments, the immunogenic target is on the surface of the dendritic cell. In some embodiments, the immunogenic target is bound to a MHC molecule on the surface of the dendritic cell. In some embodiments, the dendritic cell includes a monocyte-derived dendritic cell. In some embodiments, the dendritic cell can be derived from a CD34⁺ hematopoietic stem or progenitor cell. In some embodiments, the dendritic cell can be derived from a peripheral blood monocyte (PBMC). In some embodiments, the dendritic cells or the cells in which the DCs are derived from can be isolated by leukaphereses.

[0015]

[0016] Further embodiments of the disclosure include engineered T-cell receptors (TCRs) or chimeric antigen receptors (CARs) that specifically recognize any of the aforementioned immunogenic targets, and methods of making the same. Additional

embodiments of the disclosure include cells including one or more of the aforementioned TCRs or CARs.

[0017] Further embodiments of the disclosure include antibodies, and antigen binding fragments thereof, that specifically recognize any of the aforementioned immunogenic targets, and methods of making the same.

[0018] Additional embodiments of the disclosure include methods of treating one or more type of cancer, the method including: administering, to a patient having a type of cancer, a treatment including therapeutically effective amount of a composition including any of the aforementioned immunotherapies. Further embodiments include methods of stimulating an immune response in a subject, the method including: administering, to a subject having a type of cancer, a treatment including therapeutically effective amount of a composition including one or more of the aforementioned immunotherapies.

[0019] In some embodiments, the immunotherapy can be administered in combination with one or more additional therapies. In some embodiments, the immunotherapy and the one or more additional therapies can be administered together in one administration or composition. In some embodiments, the immunotherapy and the one or more additional therapies can be administered separately in more than one administration or more than one composition. In some embodiments, the one or more additional therapies include surgical intervention, chemotherapy, radiation therapy, hormone therapies, immunotherapy, and/or adjuvant systematic therapies.

[0020] In some embodiments, the subject has a tumor resistant to chemotherapy and/or radiation therapy. In some embodiments, the subject has been determined to be resistant to the previous treatment. In some embodiments, the cancer includes stage I, II, III, or IV cancer. In some embodiments, the cancer includes metastatic and/or recurrent cancer. In some embodiments, the subject is enrolled in a clinical trial.

[0021] Further embodiments of the disclosure include methods of identifying one or more immunogenic targets for treating a type of cancer, wherein the method includes: receiving a sample dataset, wherein the sample dataset includes mRNA from a cancer cell line, from a tumor cohort, or from a tumor sample from an individual patient; analyzing the sample dataset by a machine learning model including: identifying and quantifying post transcriptional regulation to detect one or more neojunctions and/or alternative splicing and/or alternative

promoter regulatory events; determining one or more candidate neoantigen targets based on the neojunctions and/or alternative splicing and/or alternative promoter regulatory events; filtering the candidate neoantigen targets by: a) setting a minimum threshold for antigen abundance in the dataset; and/or b) assessing the specificity of exon-exon or exon-intron junctions for tumor samples versus a healthy reference cell line or cohort dataset based on junction counts and isoform structure association; and identifying a validated tumor- or cancer-specific immunogenic target when the candidate neoantigen target meets the filtering conditions.

[0022] In some embodiments, the methods further include identifying candidate neoantigen targets by one or more of the following: a. immunopeptidomics validation; b. long-read RNA isoform sequence validation; c. single-cell short-read and/or long-read RNA-Seq validation; d. identification of unique tumor-specific extracellular protein sequences; e. splicing event type filtering; f. BayesTS probability; g. survival association h. SashimiPlot visualization; i. BLAT searching; j. 3D topology evaluation; and/or i. filtering the candidate neoantigen targets based on one or more of the above steps and/or one or more additional filtering step.

[0023] In some embodiments, immunopeptidomics validation includes filtering candidate neoantigen targets using one or more immunopeptidomics datasets including LC/MS data, with or without matching bulk RNA-Seq data; long-read RNA isoform sequence validation includes confirming candidate neoantigen targets in long-read isoform sequencing data from the cancer cell line, tumor cohort, or tumor sample and/or identifying tumor-specific full-length transmembrane protein isoforms; single-cell short-read and/or long-read RNA-Seq validation includes identifying common candidate neoantigen targets which are detected in one or more patients in a tumor sample versus microenvironment in a matching cancer and/or identifying tumor-specific full-length transmembrane protein isoforms; identification of unique extracellular tumor-specific protein sequences includes sequence comparison and alignment with a prior database (e.g. Ensembl), determination of genomic correspondence, and prediction of secondary protein structures and protein features; splicing event type filtering includes detecting splice-junctions not present in a reference database (e.g. Ensembl); Bayes TS probability includes evaluating BayesTS for the healthy reference cell line or cohort dataset and/or the sample dataset; survival association includes using a univariate Cox regression analysis to identify candidate neoantigen targets with a parental PSI value that is positively or negatively associated with patient outcome; SashimiPlot visualization includes determining the specificity of exon-

exon or exon-intron junctions for the healthy reference cell line or cohort dataset vs the sample dataset via junction counts and isoform structure association relative to known isoforms in a reference database (e.g. Ensembl); BLAT searching includes evaluating full-length predicted mRNA sequences to confirm exon-region isoform specificity relative to prior annotated mRNA and ESTs and evaluating transposable element associations; and/or 3D topology evaluation includes predicting 3D conformational differences in protein folding and domain architecture each candidate neoantigen target relative to a reference isoform.

[0024] In some embodiments, filtering the candidate neoantigen targets can include one or more of the following: a. determining if the splicing event is previously unknown; b. setting a threshold percentage of the sample dataset having the candidate neoantigen target; c. predicting 2D and/or 3D folding; d. predicting extracellular domain impact; e. SashimiPlot validation; f. evaluating secondary evidence from long-read sequencing; g. exclusion of protein isoform sequence data from a reference database; h. setting a threshold BayesTS probability score; i. setting a threshold average control tissue count; j. setting a threshold MHCNetPan binding; and/or k. setting a threshold DeepImmuno score > 0.5 .

[0025] In some embodiments, a. the candidate neoantigen target is present in greater than 5%, 10%, 15%, 20%, 25%, 30%, or higher, of the sample dataset; b. the BayesTS probability score is less than 10, less than 5, less than 4, less than 3, less than 2, less than 1, or lower; c. the average control tissue count mean is less than 10, less than 5, less than 3, less than 2, less than 1, less than 0.5, or lower; and/or d. the MHCNetPan binding is less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.5%, less than 0.1%, or lower. In some embodiments, a. the candidate neoantigen target is present in greater than 15% of patients in the cohort; b. the BayesTS probability score is less than 3; c. the average control tissue count mean is less than 1; and/or d. the MHCNetPan binding is $< 2\%$ as a weak binder and $< 0.5\%$ as a strong binder.

[0026] In some embodiments, filtering the candidate neoantigen targets includes: a. long-read single-cell genomics data; b. patient survival association data; c. immunopeptidomics data; and/or d. patient survival association with antigen presence. In some embodiments, the methods further include validating candidate neoantigen targets by LC/MS, peptide-MHC binding, proteomics, and/or immunogenicity. In some embodiments, a. validation via LC/MS includes spectral matching; b. validation via peptide-MHC binding includes performing an MHC

stabilization assay; and/or c. validation via immunogenicity includes testing pMHCs on HLA genotyped healthy blood PMBC.

[0027] In some embodiments, a ranking is derived for the candidate neoantigen target. In some embodiments, the ranking further includes immunogenicity score, HLA-binding score, neoantigen tumor specificity scores relative to healthy control samples (maximum likelihood Estimation, Bayesian hierarchical model, control cohort mean counts), alternative splicing PSI value, mRNA junction occurrence in a reference transcriptome, patient cohort occurrence frequency, recurrence in independent cancers and cohorts, splice-event type, full-length peptide prediction length versus the reference, peptide sequence similarity to reference peptides, the presence of signal peptide sequence, secondary structure, 3D protein structure prediction, insertion versus deletion of amino acids, absolute junction read counts, topology predictions, and/or other programmatic outputs.

[0028] In some embodiments, alternative splicing and alternative promoter regulatory events includes cassette exon, 3'/5' splice site exon, intron retention, alternative terminal exon, trans-splicing, and/or other event which can produce unique exon-exon or exon-intron junctions for *in silico* translation. In some embodiments, wherein the method identifies intron retention associated antigens, and/or predicts immunogenicity of the candidate neoantigen target. In some embodiments, the neojunctions include gene-associated exon-exon and exon-intron junctions.

[0029] In some embodiments, the machine learning model includes a deep learning model, probabilistic algorithms (Maximum Likelihood Estimation (MLE), hierarchical Bayesian model), and/or long-read sequencing concordance analysis. In some embodiments, the machine learning model includes a Maximum Likelihood Estimation (MLE), hierarchical Bayesian model, and long-read sequencing concordance analysis.

[0030] In some embodiments, the immunogenic target is located on a transmembrane protein. In some embodiments, the sample dataset includes mRNA from a cancer cell line. In some embodiments, the immunogenic target is shared among a patient population. In some embodiments, the sample dataset includes mRNA from a tumor sample. In some embodiments, the immunogenic target includes a patient-specific neoantigen. In some embodiments, the method further includes one or more additional reference datasets and/or algorithms. In some embodiments, the method is automated.

[0031] In some embodiments, the immunogenic target includes a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor. In some embodiments, the immunogenic target includes a neoantigen present in any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor. In some embodiments, the immunogenic target includes a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2241. In some embodiments, the immunogenic target includes a neoantigen selected from SEQ ID NOs: 1-2241. In some embodiments, the immunogenic target includes a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-4. In some embodiments, the immunogenic target includes a neoantigen selected from SEQ ID NOs: 1-4. In some embodiments, the immunogenic target comprises a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2545. In some embodiments, the immunogenic target includes a neoantigen present in a protein selected from SEQ ID NOs: 2242-2545. In some embodiments, wherein the immunogenic target comprises a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID Nos: 2242-2245. In some embodiments, the immunogenic target includes a neoantigen present in a protein selected from SEQ ID Nos: 2242-2245. In some embodiments, the immunogenic target includes a nucleotide coding for a shared splicing neoantigen selected from SEQ ID NOs: 1-2545. In some embodiments, the immunogenic target includes a nucleotide coding for a neoantigen selected from SEQ ID NOs: 1-4 or SEQ ID NOs 2242-2245.

[0032] In some embodiments, the method further includes treating the type of cancer by administering, to a subject with the type of cancer, a therapeutically effective amount of a treatment, wherein the treatment targets the tumor or cancer-specific immunogenic target. In some embodiments, the sample is a tumor sample from the subject, and wherein the treatment targets the tumor-specific immunogenic target. In some embodiments, the type of cancer includes melanoma, ovarian, head and neck, breast, skin, lung, adult and pediatric acute myeloid leukemia, myelodysplastic syndrome (MDS), bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, and/or uterine cancer. In some embodiments, the type of cancer includes melanoma, ovarian, head and neck, breast, skin, and/or MDS.

[0033] In some embodiments, the treatment includes an immunotherapy. In some embodiments, the immunotherapy includes an RNA-, DNA-, or peptide-based vaccine, dendritic cell vaccine, chimeric antigen receptor T cell (CAR-T) therapy, T cell receptor-engineered T cell (TCR-T) therapy, antibody or antibody derivative, cell-based immunotherapy, immune checkpoint inhibitor, immunomodulator, topical immunotherapy, injection immunotherapy, adoptive cell transfer, oncolytic virus therapy, immunosuppressive drug, helminthic therapy, and/or other non-specific immunotherapy. In some embodiments, the immunotherapy includes an RNA vaccine, chimeric antigen receptor T cell (CAR-T) therapy, and/or antibody or antibody derivative. In some embodiments, the immunotherapy includes any of the aforementioned immunotherapy. In some embodiments, the type of cancer is responsive to targeting of the tumor- or cancer-specific immunogenic target.

[0034] Further embodiments of the disclosure relate to methods of treating a type of cancer in a subject in need thereof, the methods including administering to the subject a therapeutically effective amount of immunotherapy including an immunogenic target selected from the peptides and proteins listed in SEQ ID NOs: 1-2545, or a nucleotide coding therefor. Further embodiments of the disclosure include methods of treating a type of cancer in a subject in need thereof, the methods including administering to the subject a therapeutically effective amount of a compound or composition directed to an immunogenic target selected from the peptides and proteins listed in SEQ ID NOs: 1-2545, or a nucleotide coding therefor. In some embodiments, the compound binds to an immunogenic target selected from the peptides and proteins listed in SEQ ID NOs: 1-2545.

[0035] Additional embodiments of the disclosure include methods of activating or expanding peptide-specific T cells. For example, such methods include contacting a starting population of T cells from a mammalian subject ex vivo with any of the aforementioned immunogenic targets, thereby activating, stimulating proliferation, and/or expanding immunogenic target-specific T cells in the starting population. In some embodiments, contacting is further defined as co-culturing the starting population of T cells with antigen presenting cells (APCs), wherein the APCs can present any of the aforementioned immunogenic targets on their surface. In some embodiments, the APCs are dendritic cells. In some embodiments, the dendritic cells are autologous dendritic cells obtained from the mammalian subject. In some embodiments, contacting is further defined as co-culturing the starting population of T cells with artificial

antigen presenting cells (aAPCs). In some embodiments, the T cells are CD8⁺ T cells or CD4⁺ T cells. In some embodiments, the T cells are cytotoxic T lymphocytes (CTLs). In some embodiments, the starting population of cells includes or consists of peripheral blood mononuclear cells (PBMCs). In some embodiments, the method further includes isolating or purifying the T cells from the peripheral blood mononuclear cells (PBMCs). In some embodiments, the mammalian subject is a human. In some embodiments, the method further includes reinfusing or administering the activated or expanded peptide-specific T cells to the subject. Additional embodiments of the disclosure include peptide-specific T cell activated or expanded according to any of the aforementioned methods.

BRIEF DESCRIPTION OF THE DRAWINGS

[0036] Those of skill in the art will understand that the drawings, described below, are for illustrative purposes only. The drawings are not intended to limit the scope of the present teachings in any way.

[0037] Figure 1 depicts a block diagram illustrating a computer system 100 upon which embodiments of the present teachings may be implemented. In various embodiments of the present teachings, computer system 100 can include a bus 102 or other communication mechanism for communicating information and a processor 104 coupled with bus 102 for processing information. In various embodiments, computer system 100 can also include a memory, which can be a random-access memory (RAM) 106 or other dynamic storage device, coupled to bus 102 for determining instructions to be executed by processor 104. Memory can also be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 104. In various embodiments, computer system 100 can further include a read only memory (ROM) 108 or other static storage device coupled to bus 102 for storing static information and instructions for processor 104. A storage device 110, such as a magnetic disk or optical disk, can be provided and coupled to bus 102 for storing information and instructions.

[0038] Figure 2. Automated discovery of immunogenic and transmembrane splicing neoantigens with SNAF, with an exemplary outline of the two parallel workflows in the software SNAF to predict splicing neoantigens derived from proteolytic peptides or full-length proteins. SNAF begins with the identification and quantification of alternative splice junctions (exon-exon and exon-intron) from RNA-Seq BAM files and filters these against normal tissue reference

RNA-Seq profiles (BayesTS). Retained tumor-specific splicing junctions (neojunctions) are evaluated in parallel for detection of T-cell antigens (SNAF-T) and B-cell antigens (SNAF-B). SNAF-T performs *in silico* translation of each junction, MHC binding affinity prediction (netMHCpan or MHCflurry) and identifies high-confidence immunogenic neoantigens through deep learning (DeepImmuno). SNAF-B predicts full-length protein coding isoforms that produce cancer-specific extracellular neo-epitopes (ExNeoEpitopes), considering existing transcript annotations and full-length isoform sequencing for targeted antibodies.

[0039] Figure 3. Detection of known and novel spliced exons and introns in SNAF. Fig. 3A) Cartoon depicting the detection of exon-exon junctions from genome/transcriptome aligned BAM files, denoting the 5' and 3' splice sites for single-end and paired-end RNA-Seq. Fig. 3B) Detection of exon-intron junction (intron retention) for single-end versus paired-end RNA-Seq. Fig. 3C) Cartoon illustrating an example AltAnalyze gene model denoting different transcripts for the same gene, collapsed into a single combined exon and intron model (model used for exon-intron detection in Fig. 3B). Exons are annotated according to which exon block they are in (E1, E2, E3, etc.) and exonic region, defined from alternative splicing sites (E1.1, E1.2, E1.3, etc.).

[0040] Figure 4. MultiPath-PSI improves the detection of splicing associated with diverse mechanisms. Benchmarking of Multipath-PSI with existing tools (MAJIQ, LeafCutter, rMATS) using simulated deep bulk-RNA sequencing data. Fig. 4A) Overview of benchmark workflow. Fig. 4B) Precision-Recall curve of evaluated tools and their performance across different splicing types.

[0041] Figure 5. Inclusion of additional normal databases increases the specificity of retained junctions. Illustration of the number retained junctions for distinct SNAF tumor specific control database filters, including the addition of TCGA control tissue samples.

[0042] Figure 6. Immuno-peptidome validation of purported peptides. Fig. 6A) Validation workflow schematic for Ovarian, Melanoma immuno-peptidomics with either matched or unmatched RNA-Seq. MaxQuant is applied to find Peptide-Spectrum Match (PSM), followed by quantitative and expert MS2 spectra prioritization. HPLC-MS/MS confirmation is performed on synthesized nominated neoantigens. Fig. 6B) Number of SNAF-T predicted neoantigens and those confirmed by immuno-peptidomics. Fig. 6C) SashimiPlot visualization of HAAASFETL, derived from an exon-exon junction in the gene *FCRLA*, along with the junction/peptide

sequence, binding affinity and immunogenicity prediction. Fig. 6D) Raw read counts of the *FCRLA* neojunction between normal controls and TCGA melanoma cohort. Fig. 6E) Mirror plot of the immunopeptidomics and spike-in MS spectrum for HAAASFETL.

[0043] Figure 7. Immunopeptidome verification of ovarian and melanoma splicing neoantigens. MS2 spike-in confirmed SNAF-T splicing neoantigens. SashimiPlot comparing tumor and an example control tissue (left) and MS2 mirror plot (right), showing the immunopeptidomics (top) and spike-in (bottom) MS spectra. Fig. 7A) NQDEDPLEV occurring from *C6orf52* in ovarian cancer. Fig. 7B) KGPWYPLSL occurring from *C20orf204* in ovarian cancer. Fig. 7C) VAPGEAKNL occurring from *RAS43* in melanoma. Fig. 7D) YALANIKWI occurring from *DYNLT5* in melanoma. Fig. 7E) KEKLDQLVY occurring from *FBXO7* in melanoma. Fig. 7F) TELQRTLST occurring from *NGLY1* in melanoma.

[0044] Figure 8. Splicing-neoantigen burden predicts response to therapy in melanoma. Fig. 8A-B) Kaplan-Meier (KM) survival plots of Melanoma patient samples stratified into low and high neoantigen burden, considering overall survival for each sequential step in SNAF for two cohorts (Fig. 8A) TCGA (n=472), and (Fig. 8B) Van Allen (n=39). These steps are: 1) tumor-specific neojunctions (left column), MHC-bound neoantigens (middle column) and immunogenic neoantigens (right column). Van Allen cohort patients were subject to immune checkpoint inhibitors whereas TCGA was not. The number of neojunctions or Neoantigen peptides are shown at the top of each plot. Fig. 8C) Volcano plots of genes differentially expressed in patients with high versus low immunogenic splicing neoantigen burden in TCGA-SKCM, with a fold>1.5 and eBayes t-test $p < 0.05$ (FDR corrected). Red dots indicate genes that are up-regulated in the high burden group whereas blue dots represent down-regulated genes in the high burden group. Fig. 8D) Gene-set enrichment with GO-Elite of Gene Ontology and cell-type specific marker gene-sets (AltAnalyze) for genes down-regulated in high splicing neoantigen burden versus low patients (panel C). Fig. 8E) Immunogenic splicing neoantigen burden between TCGA Melanoma patients with or without mutations in *CAMKK2*. Mann Whitney two-sided test. Fig. 8F) Bubble-plot of survival associated splicing neoantigens from SNAF in TCGA-SKCM. Dot size corresponds to the number of patients the splicing neoantigen is detected in (10-470) and are colored according to their survival significance in the TCGA-SKCM and Van Allen cohorts (LRT $p < 0.05$ and $z \geq 1$). AS = alternative splicing.

[0045] Figure 9. Poor survival-associated splicing-neoantigens in melanoma. Fig. 9A) Gene-set enrichment plot of high-neoantigen associated genes in TCGA-SKCM (fold>1.5 and empirical Bayes t-test $p<0.05$, FDR corrected) in the software GO-Elite. Fig. 9B) SashimiPlot visualization of shared splicing neoantigen associated exon-exon junctions in an exemplar TCGA-SKCM and highly expressed control tissues (E-MTAB-2836). Exon IDs are derived from AltAnalyze. Curved lines indicated exon-exon junctions with associated read counts in a representative sample. The arrow denotes the splicing-neoantigen parental junction (unique to tumor samples).

[0046] Figure 10. Regulatory networks mediating splicing neoantigen burden in melanoma. Fig. 10A) Schematic overview of RBP prior network construction and benchmarking against 12 TF activity methods in HepG2 cell line RBP knockdowns. Fig. 10B) The correlation of inferred RBP activity with splicing neoantigen burden for all melanoma patients. Fig. 10C) Comparison of RBP activity-burden correlations with RBP differential gene expression, for high versus low burden. Fig. 10D) Type of splicing events observed with exon/intron inclusion or exclusion comparing high versus low burden. type distribution between low and high burden groups.

[0047] Figure 11. Validation and benchmarking of 12 TF activity methods in HepG2 cells. Fig. 11A) Label Ranking Average Precision (LRAP) of 12 evaluated methods for splicing factor activity inference. Fig. 11B) Normalized Discounted Cumulative Gain (NDCG) of the 12 evaluated methods. Fig. 11C) Comparison of inferred splicing factor activity in cancer implicated splicing factors for RBP knockdown and controls.

[0048] Figure 12. Shared splicing neoantigens are frequently detected by MS and are defined by their sequence composition. Fig. 12A) Identification of common (shared) and unique immunogenic splicing neoantigens in the TCGA Melanoma cohort, based on their frequency of occurrence among patients. Fig. 12B) Frequency of splicing-event types for shared and unique splicing neoantigen junctions. Fig. 12C) Gene-set enrichment with GO-Elite of the Gene Ontology and pathways of shared neoantigens (present in >15% of patients). Fig. 12D) MS recovery rate in an independent melanoma immunopeptidome dataset (Bassani-Sternberg et al.) between shared and unique neoantigens considered in the query database. Fig. 12E) Kernel density estimate plot comparing the observed occurrence in an independent immunopeptidomics MS experimental cohort, for all detected shared (>15% of patients) versus unique splicing-

neoantigens. Fig. 12F) Re-defined shared and unique neoantigens by normalizing the occurrence of their parental splicing junction, leveraging their respective observed amino acid bias (see Example 1). Fig. 12G) UMAP of splicing neoantigens based on their amino acid physiological properties, highlighting neoantigens that cluster based on shared amino acid physicochemical features. Fig. 12H) Distinct enriched amino acid motifs (MEME), comparing shared versus unique neoantigens.

[0049] Figure 13. Shared splicing-neoantigens evidenced by immuno-peptidomics. Fig. 13A) Empirical Cumulative Distribution Function (ECDF) plot of occurrence of shared (red) versus unique (blue) splice neoantigens predicted from TCGA-SKCM. Fig. 13B) Example SashimiPlots of shared splicing-neoantigens (present in >15% of patient samples). The specificity of each neoantigen to SKCM tumors is illustrated by a highly expressed non-tumor tissue sample. Fig. 13C) Distinct enriched amino acid motifs (MEME), comparing shared versus unique neoantigens in the Van Allen cohort.

[0050] Figure 14. Shared splicing neoantigens bind HLA and induce T-cell reactivity. Fig. 14A-B) Histograms and graph show HLA-A*02-PE staining on HLA-A*02 containing TAP deficient T2 cells without peptide (no pep), loaded with FLU and HCMV control peptides and RLLGTEFQT (RLL) and FQTTRRAMTL (FQT) peptide neoantigens. MFI = median fluorescence intensity. PE = Phycoerythrin conjugated antibodies. Fig. 14C-D) Dot plots and graph show the percentage of interferon gamma-positive (IFN γ +) CD8+ T-cells in response to 5 melanoma shared splicing antigens compared to negative (unstimulated, no pep) or positive (PMA/I, FLU, HCMV) controls. CD8+ T cells were primed using peptide loaded monocyte derived dendritic cells and thereafter tested against 721.221 cells selectively expressing the indicated HLA allele with and without peptide loading. Bars indicate median of 2-3.

[0051] Figure 15. T-cell immunogenicity analyses. Schematic depiction of the experimental design to a) generate dendritic cells (DCs) from healthy donor blood PBMCs; b) prime T cells using peptide loaded DCs; and c) evaluate immunogenicity by measurement of IFN γ production in response 221 cells selectively expressing the indicated HLA allele. CD8+ T-cell priming by IFN γ production was tested for 5 melanoma shared splicing antigens and compared to negative (unstimulated, no pep) or positive (PMA/I, FLU, HCMV) controls. Experimental design adapted from Rivero-Hinojosa et al. 2021 and Meriotti et al. 2023.

[0052] Figure 16. The cell of origin for splicing neoantigens depends on the mechanism of regulation. Fig. 16A-B) Venn diagrams comparing the number of parental neojunctions for TCGA SKCM splicing neoantigens unique to a single-patient (Fig. 16A) or shared in >15% of patients (Fig. 16B) to the specific cell-types they derive from in independent melanoma tumor biopsies by single-cell RNA-Seq analysis. Neojunctions are defined as tumor or immune if they are >2 fold enriched in either cell-population (absolute number of reads in all patients and cells for each lineage). Fig. 16C) Detection of example shared splicing neoantigens in individual melanoma patients (n=19) and associated single-cell populations.

[0053] Figure 17. Comparison of melanoma specific splicing junctions to normal proliferating tissues. Fig. 17A) Overview of collected RNA-Seq data from five independent study on proliferating tissues. Fig. 17B) Decreasing number of splicing junctions specific to TCGA melanoma when normal controls were applied. Fig. 17C) Number of shared splicing neojunction reads in TCGA and proliferating skin cells. Fig. 17D) Gene pathway enrichment analysis of the parental genes of overlapping splicing junctions.

[0054] Figure 18. SNAF-B finds full-length mRNAs and stable-proteoforms for targeted therapies. Fig. 18A) The SNAF-B prediction workflow. Fig. 18B) Comparison of a SNAF-B predicted full-length isoform in the transmembrane protein SIRPA to documented mRNA isoforms and those predicted from PacBio long-read IsoSeq of melanoma cell lines. Fig. 18C) SashimiPlot of alternative 3' splice site selection in Melanoma and Brain RNA-Seq for SIRPA. Fig. 18D) Specificity of the indicated SIRPA ExNeoEpitope to melanoma versus healthy controls tissue samples. Fig. 18E) Alphafold2 3D modeling of the reference isoform and the long-read verified ExNeoEpitope. Fig. 18F, Fig. 18G) Confocal microscopy detection co-localization if SIRPA reference (Fig. 18F) or melanoma specific splice isoform (Fig. 18G) with a cell surface stain (phalloidin). Arrow indicates the vector used to quantify fluorophore spatial coincidence.

[0055] Figure 19. SLC45A2 encodes for a novel transmembrane isoform localized to the plasma membrane. Fig. 19A) Melanoma long-read confirmed (Iso-Seq) SNAF-B predicted full-length isoform in the transmembrane protein SLC45A2, corresponding to the same observed shared splicing neoantigen predicted by SNAF T (Fig. 9B). Fig. 19B) This exon-exon junction is only observed in tumors and not in controls. Fig. 19C) The resultant protein translation, with an excluded 80AA polypeptide sequence, is predicted to disrupt a transmembrane domain and result

in the novel extracellular exposure of residues that are normally cytoplasmic. Fig. 19D) AlphaFold2 3D modeling of the reference isoform and the long-read verified ExNeoEpitope, with novel extracellular exposed cytoplasmic sequence (blue/green colored residues). Fig. 19E) Co-localization of the SLC45A2 reference or melanoma specific splice isoform with a cell surface stain (phalloidin) by confocal microscopy. Fluorescent intensity along the indicated red line crossing the cell membrane is quantified in the bottom line plots for each of the three indicated channels (color corresponding to the above image) for all reference and SNAF-B isoforms.

[0056] Figure 20. Long-read and topology prediction of novel insertion SNAF-B full-length transmembrane predicted ExNeoEpitopes. SNAF-B predicted full-length isoforms for three transcripts (*DCBLD2*, *IGSF11*, *ANO10* and *SEMA6A*) predicted to result in novel extracellular polypeptide sequences, produced through alternative exon inclusion. For each, the TCGA-SKCM splice junction was observed in the shown pan-cancer long-read (Iso-Seq) reference dataset. Fig. 20A, 20D, 20G, 20J) Genomic structure of each long-read associated ExNeoEpitopes from the UCSC genome browser (BLAT). The tumor-specific included cassette-exon is shown in black. Fig. 20B, 20E, 20H, 20K) Tumor specificity of each junction in TCGA-SKCM versus GTEx. Fig. 20C, 20F, 20I, 20L) AlphaFold2 3D modeling of the reference isoform and the long-read verified ExNeoEpitope contained the indicated inserted residues in the transmembrane protein extracellular domain.

[0057] Figure 21. ExNeoEpitopes localize to the cell surface. Fig. 21A) Protter 2D structural prediction of SIRPA reference and SNAF-B identified alternative isoform. Fig. 21B) Quantification of fluorophore spatial coincidence to assess co-localization of the reference or melanoma specific splice isoform with DAPI and/or phalloidin of SIRPA. Fig. 21C, 21D) Confocal microscopy co-localization of the MET and SEMA6A reference or melanoma-specific splice isoform with a cell surface stain (phalloidin). Fluorescent intensity along the indicated red line crossing the cell membrane is quantified in the bottom line plots for each of the three indicated channels (color corresponding to the above image) for all reference and SNAF-B isoforms.

[0058] Figure 22. Shared splicing antigens in MDS. Example of shared MDS splicing antigens (>10% of patients), predicted to be MHC presented and immunogenic (SNAF-T workflow). Fig. 22A) Splicing antigens derived from intron retention, in representative donor/patient samples. Fig. 22B) Splicing antigens derived from alternative splicing, in

representative donor/patient samples. The specificity of each splicing antigen to MDS versus 60 tissue regions (bulk RNA-Seq) is displayed in the dot plot at right in Figs. 22A and 22B.

[0059] Figure 23. MDS-specific extracellular domain epitopes. Example result from the SNAF-B workflow of a novel alternative N-terminus in the transmembrane protein UMODL1. Fig. 23A) The shorter MDS-isoform (MAS-Seq) occurs due to a novel N-terminal exon not found in the UCSC genome database. Fig. 23B) The novel junction is only detected 0.2% of healthy tissue samples and 18% of MDS patients in this cohort. Fig. 23C and Fig. 23 D) Protter 2D structure prediction of the novel (Fig. 23C) and the reference (Fig. 23D) isoforms

[0060] Figure 24. SNAF interactive exploration of putative splicing neoantigens, probabilities, and selective amino acid preferences. The SNAF virtual explorer provides advanced interactive analysis of SNAF results. Fig. 24A) The SNAF-T viewer allows users to explore predicted T antigens and enriched amino acid motifs in neoantigen UMAP clusters. Fig. 24B) The SNAF-B viewer allows for the interactive query of B-cell antigens, side-by-side comparison with annotated transcripts and Neoantigen peptides and neojunction expression in cancerous and healthy tissues.

DETAILED DESCRIPTION OF THE INVENTION

[0061] Unless otherwise noted, terms are to be understood according to conventional usage by those of ordinary skill in the relevant art.

[0062] This specification describes various exemplary embodiments of identifying one or more antigen targets for treating a type of cancer. The disclosure, however, is not limited to these exemplary embodiments and applications or to the manner in which the exemplary embodiments and applications operate or are described herein.

[0063] Unless otherwise defined, scientific and technical terms used in connection with the present teachings described herein shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

[0064] As used herein, the term “sample” encompasses a sample obtained from a subject or patient. The sample can be of any biological tissue or fluid. Such samples include, but are not limited to, sputum, saliva, buccal sample, oral sample, blood, serum, mucus, plasma, urine, blood cells (*e.g.*, white cells), circulating cells (*e.g.* stem cells or endothelial cells in the blood), tissue, core or fine needle biopsy samples, cell-containing body fluids, free floating

nucleic acids, urine, stool, peritoneal fluid, and pleural fluid, tear fluid, or cells therefrom. Samples can also include sections of tissues such as frozen or fixed sections taken for histological purposes or micro-dissected cells or extracellular parts thereof. A sample to be analyzed can be tissue material from a tissue biopsy obtained by aspiration or punch, excision or by any other surgical method leading to biopsy or resected cellular material. Such a sample can comprise cells obtained from a subject or patient. In some embodiments, the sample is a body fluid that include, for example, blood fluids, serum, mucus, plasma, lymph, ascitic fluids, gynecological fluids, or urine but not limited to these fluids. In some embodiments, the sample can be a non-invasive sample, such as, for example, a saline swish, a buccal scrape, a buccal swab, and the like.

[0065] As used herein, “blood” can include, for example, plasma, serum, whole blood, blood lysates, and the like.

[0066] As used herein, the term “assessing” includes any form of measurement, and includes determining if an element is present or not. The terms “determining,” “measuring,” “evaluating,” “assessing” and “assaying” can be used interchangeably and can include quantitative and/or qualitative determinations.

[0067] As used herein, the term “monitoring” with reference to a type of cancer refers to a method or process of determining the severity or degree of the type of cancer or stratifying the type of cancer based on risk and/or probability of mortality. In some embodiments, monitoring relates to a method or process of determining the therapeutic efficacy of a treatment being administered to a patient.

[0068] As used herein, “outcome” can refer to an outcome studied. In some embodiments, “outcome” can refer to survival / mortality over a given time horizon. For example, “outcome” can refer to survival / mortality over 1 month, 3 months, 6 months, 1 year, 5 years, or 10 years or longer. In some embodiments, “outcome” can refer to 28-day survival / mortality. In some embodiments, an increased risk for a poor outcome indicates that a therapy has had a poor efficacy, and a reduced risk for a poor outcome indicates that a therapy has had a good efficacy. In some embodiments, “outcome” can refer to resolution of organ failure after 14 days or 28 days or limb loss. Although mortality / survival is obviously an important outcome, survivors have clinically relevant short- and long-term morbidities that impact quality of life, which are not captured by the dichotomy of “alive” or “dead.”

[0069] As used herein, the terms “predicting outcome” and “outcome risk stratification” refers to a method or process of predicting and/or prognosticating a patient’s risk of a certain outcome. In some embodiments, predicting an outcome relates to monitoring the therapeutic efficacy of a treatment being administered to a patient. In some embodiments, predicting an outcome relates to determining a relative risk of an adverse outcome (e.g. complicated course) and/or mortality. In some embodiments, the predicted outcome is associated with administration of a particular treatment or treatment regimen. Such adverse outcome risk and/or mortality can be high risk, moderate risk, moderate-high risk, moderate-low risk, or low risk. Alternatively, such adverse outcome risk can be described simply as high risk or low risk, corresponding to high risk of adverse outcome (e.g. complicated course) and/or mortality probability, or high likelihood of therapeutic effectiveness, respectively.

[0070] As used herein, the term “high risk clinical trial” refers to one in which the test agent has “more than minimal risk” (as defined by the terminology used by institutional review boards, or IRBs). In some embodiments, a high risk clinical trial is a drug trial.

[0071] As used herein, the term “low risk clinical trial” refers to one in which the test agent has “minimal risk” (as defined by the terminology used by IRBs). In some embodiments, a low risk clinical trial is one that is not a drug trial. In some embodiments, a low risk clinical trial is one that involves the use of a monitor or clinical practice process. In some embodiments, a low risk clinical trial is an observational clinical trial.

[0072] As used herein, the terms “modulated” or “modulation,” or “regulated” or “regulation” and “differentially regulated” can refer to both up regulation (*i.e.*, activation or stimulation, *e.g.*, by agonizing or potentiating) and down regulation (*i.e.*, inhibition or suppression, *e.g.*, by antagonizing, decreasing or inhibiting), unless otherwise specified or clear from the context of a specific usage.

[0073] As used herein, the term “subject” refers to any member of the animal kingdom. In some embodiments, a subject is a human patient. In some embodiments, a subject is a pediatric patient. In some embodiments, a pediatric patient is a patient under 18 years of age, while an adult patient is 18 or older.

[0074] As used herein, the term “treating” (and its variations, such as “treatment,” “treating,” “treat,” and the like) is, unless stated otherwise, to be considered in its broadest context and refers to obtaining a desired pharmacologic and/or physiologic effect. In particular,

for example, the term “treating” may not necessarily imply or require that an animal is treated until total recovery. Accordingly, “treating” includes amelioration of the symptoms, relief from the symptoms or effects associated with a condition, decrease in severity of a condition, or preventing, preventively ameliorating symptoms, or otherwise reducing the risk of developing a particular condition. In some aspects, “treating” may not require or include prevention. In some aspects, “treatment” can refer to obtaining a desired pharmacologic and/or physiologic effect. The effect can be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or can be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. “Treatment,” as used herein, covers any treatment of a disease in a subject, particularly in a human, and includes: (a) preventing the disease from occurring in a subject which may be predisposed to the disease but has not yet been diagnosed as having it; (b) inhibiting the disease, *i.e.*, arresting its development; and (c) relieving the disease, *i.e.*, causing regression of the disease and/or relieving one or more disease symptoms. “Treatment” can also encompass delivery of an agent or administration of a therapy in order to provide for a pharmacologic effect, even in the absence of a disease or condition.

[0075] As used herein, reference to “treating” an animal includes but is not limited to prophylactic treatment and therapeutic treatment. The effect can be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or can be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. “Treatment,” as used herein, covers any treatment of a disease in a subject, preferably in a mammal (e.g., in a human), and may include one or more of: (a) preventing the disease from occurring in a subject which may be predisposed to the disease but has not yet been diagnosed as having it; (b) inhibiting the disease, *i.e.*, arresting its development; and (c) relieving the disease, *i.e.*, causing regression or elimination of the disease and/or relieving one or more disease symptoms.

[0076] “Treatment” can also encompass delivery of an agent or administration of a therapy in order to provide for a pharmacologic effect, even in the absence of a disease or condition. Any of the compositions (e.g., pharmaceutical compositions) described herein can be used to treat a suitable subject.

[0077] “Therapeutically effective amount” means an amount effective to achieve a desired and/or beneficial effect. An effective amount can be administered in one or more

administrations. In the methods, a therapeutically effective amount is an amount appropriate to treat an indication. By treating an indication is meant achieving any desirable effect, such as one or more of palliate, ameliorate, stabilize, reverse, slow, or delay disease progression, increase the quality of life, or to prolong life. Such achievement can be measured by any suitable method, such as measurement of tumor size or blood cell count, or any other suitable measurement.

[0078] As used herein, the term “marker” or “biomarker” refers to a biological molecule, such as, for example, a nucleic acid, peptide, protein, hormone, and the like, whose presence or concentration can be detected and correlated with a known condition, such as a disease state. It can also be used to refer to a differentially expressed gene whose expression pattern can be utilized as part of a predictive, prognostic or diagnostic process in healthy conditions or a disease state, or which, alternatively, can be used in methods for identifying a useful treatment or prevention therapy.

[0079] As used herein, an mRNA “isoform” is an alternative transcript for a specific mRNA or gene. This term includes pre-mRNA, immature mRNA, mature mRNA, cleaved or otherwise truncated, shortened, or aberrant mRNA, modified mRNA (e.g. containing any residue modifications, capping variants, polyadenylation variants, etc.), and the like.

[0080] “Antibody” or “antibody peptide(s)” refer to an intact antibody, or a binding fragment thereof that competes with the intact antibody for specific binding; this definition also encompasses monoclonal and polyclonal antibodies. Binding fragments are produced by recombinant DNA techniques, or by enzymatic or chemical cleavage of intact antibodies. Binding fragments include Fab, Fab', F(ab')₂, Fv, and single-chain antibodies. An antibody other than a “bispecific” or “bifunctional” antibody is understood to have each of its binding sites identical. An antibody, for example, substantially inhibits adhesion of a receptor to a counterreceptor when an excess of antibody reduces the quantity of receptor bound to counterreceptor by at least about 20%, 40%, 60% or 80%, and more usually greater than about 85% (as measured in an in vitro competitive binding assay).

[0081] As used herein, the term “expression levels” refers, for example, to a determined level of biomarker expression. The term “pattern of expression levels” refers to a determined level of biomarker expression compared either to a reference (e.g. a housekeeping gene or inversely regulated genes, or other reference biomarker) or to a computed average expression value (e.g. in RNA sequencing analysis, or DNA-chip analyses). A pattern is not

limited to the comparison of two biomarkers but is more related to multiple comparisons of biomarkers to reference biomarkers or samples. A certain “pattern of expression levels” can also result and be determined by comparison and measurement of several biomarkers as disclosed herein and display the relative abundance of these transcripts to each other.

[0082] As used herein, a “reference pattern of expression levels” refers to any pattern of expression levels that can be used for the comparison to another pattern of expression levels. In some embodiments of the disclosure, a reference pattern of expression levels is, for example, an average pattern of expression levels observed in a group of healthy or diseased individuals, serving as a reference group.

[0083] As used herein, a “neoantigen” is a tumor antigen, or a cancer antigen, i.e. an antigen found on or in cancer cells when certain mutations occur in tumor DNA. Similarly, a “splicing neoantigen” is a neoantigen that can be shared among patients with a type of cancer, or unique to a particular tumor. In some embodiments, a splicing neoantigen can be present in >15% of patients with a type of cancer. In some embodiments, a splicing neoantigen can be present in more than 8%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, or 90% of patients with a type of cancer. In some embodiments, a splicing neoantigen can predict survival. In some embodiments, a splicing neoantigen can be validated by immunopeptidomics (e.g., SNAF-T) or long-read isoform sequencing (e.g., SNAF-B). Splicing neoantigens can be produced through various alternative mRNA processing mechanisms, including, but not limited to, alternative splice-site selection, alternative exon inclusion, alternative intron inclusion, alternative promoter selection, alternative C-terminal exon selection, internal-hybrid splicing events, circular mRNA splicing, genomic rearrangements, alternative polyadenylation and exon bleeding.

[0084] Embodiments of the disclosure set forth herein include methods of identifying one or more targets for treating a disease or disorder, such as a type of cancer. Still other embodiments of the disclosure include methods of treating a disease or disorder, by administering a treatment which relates to an identified target. Some embodiments include methods of determining whether a patient is suitable for, or likely to respond favorably to, a particular treatment. Additional embodiments of the disclosure are also discussed herein.

Overview

[0085] The present disclosure relates to computational methods and molecular targets for cancer immunotherapy. These molecular targets derive, for example, from aberrant alternative splicing or alternative promoter mRNA products from tumors, that encode for neoantigens. Such neoantigens can be amino acid- or nucleotide-based and can be used in immunotherapy, including for the creation of RNA vaccines or synthetic chimeric antigen receptor T cell (CAR-T) therapy.

[0086] While immunotherapy is typically reserved for cancer patients with a high mutational burden, neoantigens produced from post-transcriptional regulation can provide an untapped reservoir of common immunogenic targets for new targeted cancer therapies. The identification and prioritization of shared neoantigens based on alternative splicing events unique to cancer cells within and across cancers can lead to new targeted immunotherapies. However, the development of existing targeted neoantigen immunotherapies is time-consuming and costly, as they must exploit specific MHC-presented mutations that are evidenced by proteomics and immunogenicity assays, in a precision manner.

[0087] As described herein, the present disclosure relates to a computational tool to more accurately predict T cell-directed immunogenicity of putative MHC-I bound peptides for immunotherapy (PMID: 33398286). By coupling such predictions with other bioinformatic algorithms previously developed by the inventors, such as, for example, algorithm MultiPath-PSI in the alternative splicing analysis software AltAnalyze (<https://www.nature.com/articles/s41467-021-26623-y>) and the peptide immunogenicity algorithm DeepImmuno (<https://academic.oup.com/bib/article/22/6/bbab160/6261914>), a new computational workflow, called SNAF (Splicing Neo Antigen Finder), was developed. This tool can find high-confidence common (shared) neoantigens from large cancer patient RNA-Sequencing data cohorts (T-cell and B-cell targets). The tool finds splicing-derived neoantigens that can be readily validated from existing experimental validation data (HLA targeted mass spectrometry). As such neoantigens are frequently shared among cancer patients, this prognostic has a broad and valuable use in treating cancer via immunotherapy, which is the most effective strategy for treatment (e.g., anti-PD-1, check-point inhibitor immunotherapy, anti-PD-L1, CAR-T, TCR-T therapies, neoantigen vaccines, and the like).

[0088] The computational systems developed as described herein are designed to systematically identify new classes of neoepitopes, e.g. splicing-neoantigens from tumor cohorts

or individual tumor RNA-Sequencing (RNA-Seq) samples, and prioritize and interpret distinct classes of splicing-neoantigens. Two new computational workflows focused on T-cell and B-cell based therapies were developed. The workflow incorporates advanced deep-learning and probabilistic algorithms to discover immunogenic (SNAF-T workflow), and full-length protein-coding transmembrane tumor-specific isoforms (SNAF-B workflow). These workflows enable the identification of splicing neoantigens that are frequently shared among patients (present in >25% of patients with the cancer), can predict survival, and can be validated by immunopeptidomics (SNAF-T) or long-read isoform sequencing (SNAF-B).

[0089] T-cell based therapies include, but are not limited to, cancer vaccines, which require that target antigens are processed, presented by MHC and immunogenic. B-cell based therapies, such as monoclonal antibodies, require the identification of transmembrane protein encoding neoantigens that will enable targeted approaches to selectively recognize cancer cells. SNAF was developed to recognize and prioritize both classes of neoantigens (SNAF-T and SNAF-B) in individual patient samples, while assessing the aggregate importance of each neoantigen at a population level (neoantigen burden) (Figure 2).

[0090] The disclosure describes the identification of numerous novel small molecule peptide and full length protein neoantigen targets to predict response to immunotherapy in cancer treatment. Such neoantigens are produced through a common splicing mechanism. In particular, a broad and clear amino acid preference has been found to be responsible for nearly all commonly produced melanoma neoantigens; this workflow can therefore be used to prioritize additional neoantigen candidates in the future (not just for melanoma). Many shared neopeptides (found in many individuals) harbor a proline at position 4 (each neoantigen is 9 residues), whereas patient-specific neoantigens harbor an arginine at position 9.

[0091] Current immunotherapy is not guided by the detection of splicing neoantigens. The presently identified targets are shared among patients (e.g. >10%, >15%, >25%, etc.), thus making them informative diagnostic indicators of immunotherapy success. In addition, the finding that shared neoantigen peptides (found in many individuals) harbor a proline at position 4 (each neoantigen is 9 residues), whereas patient-specific neoantigens harbor an arginine at position 9, indicates both a common alternative splicing mechanism with the same cancer type (shared exon-exon junction) and a sequence specific bias in the 5' versus 3' exon-intron splice-site unique to these cancer disruptions.

[0092] According to embodiments of the disclosure, after diagnosis of a type of cancer (for example, following a standard biopsy), the presence of the neoantigen associated exon-exon junctions can be determined (for example, via quantification by quantitative PCR (TaqMan assay) in the biopsy as total RNA as compared to a healthy reference control RNA). According to further embodiments of the present disclosure, chimeric antigen receptor (CAR) T-cell therapy for the neoantigen can be introduced in a patient's own T-cells, *ex vivo*.

[0093] According to further embodiments of the disclosure, splicing neoantigens can be used as shared (recurrent) targets for immunotherapy and overall splicing neoantigen burden as an independent prognostic indicator. As described herein, top-prioritized targets for RNA vaccine and CAR-T immunotherapy were identified from the analysis of multiple melanoma RNA-Seq cohorts and independently observed in multiple independent cancers, considering patient biopsies collected at the time of diagnosis with and without immunotherapy treatment following diagnosis. These targets are relevant in multiple cancer types, including melanoma, solid, and blood cancers, in which the underlying associated splicing junctions are recurrently detected.

[0094] Compared to independent filings, the predictions reported in this application are >98.1% unique, with only 9 prior filed splicing neoantigens partially or fully overlapping, out of 482 unique neoantigens reported. To assess the overlap of targets, each reported sequence was examined to identify splicing neoantigens with at least 5 continuous overlapping amino acids or overlapping exon-exon junctions (following liftover from hg38 to hg19). This difference in target overlap (considering thousands of prior reported candidates) reflects the precision of the presently described SNAF computational workflow over others and the type of cancer in which these targets were derived (melanoma) versus others (e.g., brain, prostate, breast). The predictive ability of the splicing neoantigen burden to independently predict patient prognosis and response to immunotherapy was performed and described for melanoma, but can be applicable in unrelated cancers.

Identification of Shared Splicing Neoantigens in Cancers

[0095] An overarching objective for the treatment of cancer is the discovery of standardized and accessible therapeutic strategies for shared targets that will be effective in a large percentage of patients. Tumor heterogeneity has been widely acknowledged as a hallmark of cancer, which poses challenges for developing new targeted therapies. Such heterogeneity is

further responsible for drug resistance, which leads to frequent cancer relapse. Since each tumor sample is unique, with distinct mutations, the search for tumor-specific neoantigens has been considered the “final common pathway” for the immune system to fight cancer.

[0096] Focused targeting of patients with selective mutations has produced promising results in cancers with a high mutational burden, such as melanoma, non-small cell lung cancer, and MSI-high (MSI-H) colorectal cancer. For example, 4-out-of-6 melanoma patients vaccinated with precision neoantigen vaccines show no evidence of relapse within 25 months post-therapy. Such promising clinical results have been attributed to the long-term persistence of neoantigen-specific memory T cells, illustrating the durability of neoantigen-based therapies. Other exciting examples include Moderna's mRNA-4157 combination with pembrolizumab, which achieved a 50% response rate in HPV-negative HNC compared to 14.6% for pembrolizumab monotherapy and adoptive T cell transfer, in which neoantigen-reactive T cells are cultured and reinfused into the same patient, resulting in a 55% objective response and 23% complete response rate in metastatic melanoma.

[0097] Although immune checkpoint blockade (ICB) has become the front-line clinical treatment in patients with high mutational burden, such therapies are not used in many cancers with low mutation burden, such as glioma and leukemia. While historically attributed to tumor-associated mutations, neoantigens can be produced from diverse post-transcriptionally regulatory mechanisms. Alternative splicing is one of the primary mechanisms used to achieve mRNA transcript and proteomic diversity in higher-order eukaryotes. In cancer, altered mRNA splicing can lead to aberrant protein products that promote oncogenic transformation, metastasis and confer chemotherapy resistance.

[0098] While historically attributed to tumor-associated mutations, neoantigens can be produced from diverse post-transcriptionally regulatory mechanisms. Following their initial identification using proteogenomics approaches, splicing neoantigens have become increasingly recognized as a potent source of neo-peptides that can elicit immune response and induce cancer cell death. Depending on the cancer, splicing neoantigens appear to often be the dominant source of tumor-specific peptides. Such splicing events include intron retention, which typically results in nonsense-mediated decay, but which produce MHC-presented neopeptides that can be detected by mass-spectrometry (MS). Such peptides require further experimental validation, as MHC presentation alone does not dictate binding to T-cell receptors (immunogenicity).

[0099] The prediction of such antigens, however, remains non-trivial, as splicing neoantigens must be degraded, bound, and presented by specific cognate HLA alleles, and interact with patient-specific T-cell receptors on CD8⁺ T cells to induce an immune response. As such, the precise relationship between splicing neoantigen expression and patient prognosis has remained largely unknown, and it is unclear whether overall splicing neoantigen burden impacts response to immunotherapy. Further, a concern for the use of splicing neoantigens as targets for therapy is that the occurrence of a splicing event is often non-binary (changes in percent exon/intron inclusion), relative to mutations (change in percent exon inclusion), making it difficult to know which splicing events are truly tumor-specific.

[00100] An alternative strategy to target tumor-specific splicing is to focus on events that specifically result in novel translated transmembrane proteins that can expose tumor-specific epitopes, bypassing the need to be presented by HLA. Such peptides can be recognized by new CAR-T therapies which use B-cell receptors to bind epitopes or selective monoclonal antibodies to mediate targeted tumor cell death. While attractive, identifying such novel isoforms requires an accurate prediction of full-length novel isoforms that do not undergo nonsense-mediated decay and result in properly folded protein structures that conserve the major domains of the reference protein. Given these challenges, no reusable and sufficiently comprehensive neoantigen prediction workflows exist, to unbiasedly and confidently identify splicing neoantigens that can be exploited by current immunotherapy strategies.

[00101] The disclosure relates to the development of the Splicing Neo-Antigen Finder (SNAF), an easy-to-use computational tool to identify, prioritize and interpret distinct classes of splicing-neoantigens, which can be used to perform a system analysis of splicing-neoantigens in cancer. The workflow incorporates advanced deep-learning and probabilistic algorithms to discover immunogenic (SNAF-T workflow), and full-length protein-coding transmembrane tumor specific isoforms (SNAF-B workflow) and regulators of altered splicing (RNA-SPRINT). The work described herein demonstrates that splicing neoantigens in melanoma are frequently shared among patients, can predict survival and can be validated *in vitro*, by immunopeptidomics (SNAF-T), single-cell genomics, and/or long-read isoform sequencing (SNAF-B). These analyses show that splicing-neoantigens represent an untapped reservoir of shared targets for targeted cancer immunotherapy.

[00102] This study demonstrates that aberrant splicing in melanoma frequently results in shared MHC-presented neoantigens, which can be confirmed in different patient cohorts and predict survival and response to immunotherapy. Patients with high splice neoantigen burden skew towards poor outcomes and associate with genes important to block immune-tumor recruitment. Further, the SNAF computational strategy for splicing neoantigen discovery is unique in its inclusion of probabilistic modeling to quantify immunogenicity and tumor specificity, interactive exploratory methods, quantification of splicing factor activities and interfaces for long-read and immunopeptidomics analysis. The high validation rate can be attributed to these rigorous approaches. These analyses establish the broad existence of highly shared splicing neoantigens in melanoma and nominate coordinated splicing failure as a broad mediator of mis-splicing. Intriguingly, shared versus patient specific splicing neoantigens were found to have distinct physicochemical characteristics and cells of origins, indicative of distinct mechanisms of regulation.

[00103] The presence of common splicing neoantigens indicates that, because cancer cells maintain a delicate balance between mutations in oncogenes and suppression of their presentation by MHC, a similar mechanism can exist for splicing neoantigens, which can further inform the type of therapy administered. Second, although current neoantigen predictions focus on HLA-I presentation and CD8 T-cell function, HLA-II and CD4 T-cells have also been reported to play an essential role in enhancing anti-tumor activities, together with other major immune cell types (i.e. neutrophils, dendritic cells). As current bioinformatics pipelines do not consider the activities of these other HLA mechanisms and T-cell subsets, future methods can be designed to incorporate additional T-cell and antigen presentation mechanisms. Such rich experimental data can further enable improved learning methods to identify neoantigens with the best likelihood of being either immunogenic or resulting in full-length transmembrane proteins with altered extracellular epitopes.

[00104] Targets identified via the methods described herein can be validated experimentally, by validating a range of predicted immunogenic and non-immunogenic peptides (MHC-presentation, immunogenicity) from mutations and splicing that are shared or unique and derived from different tools and statistical cutoffs. ExNeoEpitopes can be validated by new proteogenomics approaches that leverage targeted long sequencing isoform sequencing and proteomics along with antibodies that target specific conformational epitopes. Such antibodies

can represent powerful new molecular reagents for CAR-T or monoclonal antibody strategies, for shared and patient-specific neoantigens. While the current pipeline enables the identification of ExNeoEpitopes and deep visual interrogation of the impact and position of residues in novel cancer protein isoforms, improved automated bioinformatics methods can also be used to determine which introduced or removed residues can result specifically in novel extracellular or transmembrane sequences that retain the conformational integrity of the resulting protein isoforms.

[00105] Moreover, transposable elements (TE) have been reported to contribute to a subset of alternative splicing events and give rise to neoantigens, including endogenous retrovirus. Further systemic analysis can assess the convergence between these two types of neoantigens and their contribution, respectively.

[00106] A final important consideration is the tumor specificity of such neoantigens. As antigen assays do not provide information on tumor specificity, more comprehensive normal tissue references are beneficial. Most conventional RNA-Seq studies are on a limited set of adult human tissues, without considering rare cell-types or fetal developmental isoforms. An expanded atlas of normal tissues, with extremely high sequencing depths (>100 million reads) and longer reads (>100nt), which may be aided by newly reported cheaper and longer sequencing approaches and ideally single-cell resolution for hundreds of cell types, can provide an improved dataset. Given its flexibility, SNAF can be easily extended to new datasets, sequencing technologies, and neoantigen prediction libraries which can be deployed in a modular manner in custom bioinformatics pipelines. Applied broadly to new cancers and distinct forms of malignancy, SNAF can be used to identify splicing neoantigens that are unique and shared across malignancies and discover novel sequence motif preferences that expand the repertoire of targets for precision cancer therapy.

[00107] In summary, the present disclosure describes a reusable and comprehensive neoantigen identification and prediction workflow, to unbiasedly and confidently identify splicing neoantigens that can be targeted for treatment, including current and emerging immunotherapy strategies. The disclosure encompasses shared splicing neoantigens and neoantigen-derived targets, including polypeptides and full length proteins/antibodies, as well as methods of identifying shared splicing neoantigen targets, as well as methods of treatment

leveraging the identification of such shared splicing neoantigens, such as immunotherapy treatments, and the like.

SNAF Workflow and Secondary Target Prioritization and Validation

[00108] SNAF applies unique deep-learning, Bayesian, and long-read sequencing concordance analyses to identify shared high-confidence splicing neoantigens that can be experimentally confirmed using diverse orthogonal methods. SNAF-T and SNAF-B frequently identify tens of thousands of predictions. A small subset of these predictions are shared neoantigens (e.g. present in >8%, >10%, >15%, etc. of patients in a cohort). Splicing neoantigen validation, using the procedures described herein, currently necessitates reduction of targets to a small number of targets (dozens) for 9-11 amino acid antigens or full-length transmembrane isoforms. While the SNAF source code represents the foundation of splicing neoantigen target selection, a series of non-public domain computational and filtering strategies developed as described herein prioritizes optimal targets for validation. These filtering strategies enable target prioritization and entail the following steps for predicted immunogenic splicing neoantigens (SNAF-T) and extracellular transmembrane predicted neo-isoforms (SNAF-B).

SNAF-T Bioinformatics Workflow

[00109] Immunopeptidomics validation: Filter splicing neoantigen predictions using LC/MS immunopeptidomics datasets. These datasets comprise LC/MS with matching bulk RNA-Seq data on the same patient or unmatched datasets lacking patient bulk RNA-Seq. Splicing neoantigens predicted in the immunopeptidomics datasets are matched to common splicing neoantigens identified in one or more larger cancer RNA-Seq cohorts. These datasets are analyzed as described in the method using the SNAF-T and MaxQuant workflows. As LC/MS data is subject to false positive peptide-database matches report, to prioritize targets, a rigorous selection criterion have been developed for high-confidence matching hits. Specifically, high-confidence peptide-spectrum matched peptides (HC-PSMp) are performed based on the following criterion: minimum number of matched fragments (n=6), minimum ion density (<25% unassigned ions >10% relative abundance), presence of sequential fragment ions (e.g., y4-y5-y6), minimum proportion of explained and unexplained fragments, minimum known sequence-specific characteristic of MS2 spectrum. Splicing neoantigens with LC/MS evidence are prioritized for validation.

[00110] Long-read sequence validation: Long-read RNA isoform sequencing data are used to identify splicing neoantigens that can be confirmed in long-read isoform sequencing data that is selective to patient tumor samples or cell lines from any available malignancy.

[00111] Single-cell validation: Single-cell short-read and/or long-read RNA-Seq is employed to select common splicing neoantigens that can be detected in one or more patients restricted to presumed tumor versus microenvironment in the matching cancer.

[00112] Splicing-event type filtering: Splice-junctions not present in public RNA databases (Ensembl).

[00113] BayesTS probability: BayesTS is evaluated for both the entire default healthy tissue atlas control dataset (GTEx + TCGA non-tumor control samples), and an optional tumor-specific user-provided control dataset.

[00114] Survival association: To identify individual neojunctions or Neoantigens associated with survival, an univariate Cox Regression analysis is used to identify events/antigens with a parental PSI value that are positively or negatively associated with patient outcome. Here, the neojunction and its parental PSI value is ignored if the neoantigen was not predicted to be presented in that sample, resulting in different survival associations for different neoantigens produced from the same neojunction.

[00115] SashimiPlot visualization: The specificity of exon-exon or exon-intron junctions for tumor samples versus an independent (non-GTEx and non-TCGA) cohorts are assessed at the level junction counts and isoform structure association relative to known isoforms (Ensembl). Non-GTEx datasets comprise both internal bulk RNA-Seq data collections and prior published (e.g., EvoDevo).

[00116] SNAF-T event filtering: Using the above filtering methods, SNAF-T shared-splicing neoantigens are defined on the basis of:

- a. Optimized RNA-sequence alignment strategy to define challenging to detect exon-exon junctions;
- b. SNAF-T predicted antigen in a minimum threshold (e.g. >15%) of tumor patients; and
- c. SashimiPlot validation

[00117] Additional filtering steps which can be incorporated include:

- a. Novel-evidenced splicing event (Ensembl and GTEx);

- b. BayesTS score < 3 (applied to default and custom tissue databases);
- c. Average control tissue count mean < 1 (default and custom tissue databases);
- d. MHCNetPan binding < 2% of weak binder and < 0.5% as strong binder; and/or
- e. DeepImmuno score > 0.5.

[00118] The following optional filtering steps can improve results and can be incorporated as well:

- a. Evidence from long-read single-cell genomics;
- b. Evidence from patient survival associations;
- c. Evidence from immunopeptidomics; and/or
- d. Significant patient survival associations with antigen presence.

[00119] LC/MS validation: To confirm that splicing neoantigens can be detected as cell membrane bound peptide-MHCs, high confidence peptides defined by our HC-PSMp thresholds were synthesized, and spectral matching was defined by targeted LC/MS. Synthetic MS2 spectrum that match the cancer patient spectrum are considered validated (Figure 8). Applying this workflow to melanoma (n=22) and ovarian cancer (n=14) immunopeptidomics evidenced SNAF-T splicing neoantigens confirmed 7/35 peptides, in which sufficient 28 peaks were detectable from the synthetic peptide LC/MS analysis. These seven neoantigens derive from multiple mechanisms including known cassette exons, alternative 3' and 5' splice sites, intron retention and novel cassette exons (AltAnalyze defined). For example, the shared Melanoma splicing neoantigen HAAASFETL in the gene FCRLA occurs due to a known cassette exon-exon junction in an isoform that is weakly detected in blood and spleen (average read count = 0.51, BayesTS: 0.03) (Figure 6C). However, this junction is detected in >34% of TCGA melanoma patients (162/472, Average read count = 52.91) (Figure 6D). The resultant MS2 spectrum of this antigen has a high-confidence match (Andromeda score: 149.06, P-value: 0.0), with a synthetic spectrum Pearsonr similarity of 0.63 (p-value: 0.05, cosine similarity: 0.87). The other validated neoantigens derive from ubiquitin protein ligase complex (FBXO7), asparagine amidase (NGLY1), cytoskeletal motor protein (DYNLT5), negative regulator of RAS signaling (RASA3) and currently uncharacterized protein coding genes (C20orf204, C6orf52) (Figure 4). Lower spectrum similarity scores are observed in the ovarian cohort (Pearsonr: 0.55) compared to the melanoma cohort (Pearsonr: 0.84), which can be attributed to differences in the

MS technology applied (Ion trap MS versus Fourier transform MS) and fragmentation methods (CID versus HCD).

[00120] Peptide-MHC binding: To assess pMHC binding of splicing neoantigens, an MHC stabilization assay was performed in T2 cells, that contain the A allele 02:01 and are defective in TAP and thus cannot process endogenously derived peptides. Neoantigens were synthesized as custom 9AA peptides using a commercial solid-phase protocol (Abclonal). For these assays, these T2 cells were loaded with relevant HLA-A2 binding peptides (neoantigens), and stabilization of the pMHC and resurfaces was measured. This workflow was tested using IgG as a negative control, with flu vaccine and HCMV peptides used as positive controls. While pMHC binding was not observed with IgG, both positive controls and tested splicing neoantigens demonstrate evidence of binding (Figure 2).

[00121] Immunogenicity: To test immunogenicity of splicing neoantigens and detect T-cell responses, HLA-typed healthy blood PBMC was leveraged to test specific pMHCs predicted to induce immunogenicity by DeepImmuno. This assay has been deployed using common HLA genotypes that match those for shared splicing neoantigen predictions. This assay entails generation of Dendritic cells (DCs) through GM-CSF and IL-4 stimulation of monocytes at day 0 and day 3, following the beginning of cell culture from frozen healthy PBMCs. Reconstituted lyophilized neopeptides were co-cultured (pulsed) with these cells for 4-6 hours, prior inducing DC maturation (IL-4, GM-CSF, IL-1B, IL-6, TNF-alpha, LPS). T-cells (frozen PBMCs at day 7) were then stimulated via co-cultured with the antigen presenting DCs from the same donor co-cultured with to (IL-2, IL-6, IL-7, IL-12, IL-15), 2-to-3 times, prior to evaluation of immunogenicity in the expanded T-cells. Immunogenicity was evaluated at day 28 using flow cytometry to assess interferon gamma response in CD8+ T-cells. This assay confirmed CD8+ immunogenicity of both positive controls (PMA/I, FLU and/or HCMV) relative to unstimulated DCs, as well as 5 frequently shared SNAF-T melanoma splicing neoantigens (derived from three unique neojunctions - *SLC45A2*, *CDH19*, *PMEL*), with a preference for HLA A*02, C*04 or C*08 (Figure 2).

SNAF-B Bioinformatics Workflow

[00122] Long-read sequence evidence: In addition to using long-read isoform sequencing data to verify full-length isoform predictions derived from the short-read dataset, bulk and single-cell long-read isoform datasets were collected and aggregated to directly

evidence tumor-specific full-length transmembrane protein isoforms (ExNeoEpitopes) for validation. These dataset span diverse cancers and include primary patient specimens and cell lines, from prior published studies and in house generated datasets. Isoform evidence is evaluated at the level of isoform GTF files and matched to exon-exon or exon-intron junction genomic coordinate positions for shared SNAF-B splicing neoantigens evidenced from short-read RNA-Seq cancer cohort datasets.

[00123] Unique extracellular protein sequences: To confirm the specificity of extracellular domain encoding novel tumor-specific protein sequences in the ExNeoEpitope. A series of semi-automated analyses are performed. First, ExNeoEpitope sequence is compared to existing protein isoforms for all common gene annotated protein sequences from the Ensembl and UCSC Genome databases. Next, global (EMBOSS) alignment of the neo-isoform and the reference isoform are performed using EMBL-EBI pairwise sequence alignment API to confirm inserted or deleted sequence specificity. Next, genomic coordinate overlaps are computed for the neojunction and the annotated extracellular domain sequence corresponding to neojunction using genomic correspondences in the AltAnalyze database. Finally, secondary protein structure and protein feature predictions are compared for each ExNeoEpitope to its reference protein isoform to confirm the specificity of extracellular domain modifications, the introduction or removal of transmembrane domains and preservation of overall protein topology.

[00124] Splicing-event type filtering: Splice-junctions not present in public RNA databases (Ensembl).

[00125] BayesTS probability: BayesTS is evaluated for both the entire default healthy tissue atlas control dataset (GTEx + TCGA non-tumor control samples), and an optional tumor-specific user-provided control dataset.

[00126] Survival association: To identify individual neojunctions or Neoantigens associated with survival, an univariate Cox Regression analysis is used to identify events/antigens with a parental PSI value that are positively or negatively associated with patient outcome. Here, the neojunction and its parental PSI value is ignored if the neoantigen was not predicted to be presented in that sample, resulting in different survival associations for different neoantigens produced from the same neojunction.

[00127] SashimiPlot visualization: The specificity of exon-exon or exon-intron junctions for tumor samples versus an independent (non-GTEx and non-TCGA) cohorts are

assessed at the level junction counts and isoform structure association relative to known isoforms (Ensembl). Non-GTEX datasets comprise both internal bulk RNA-Seq data collections and prior published (e.g., EvoDevo).

[00128] Human BLAT search: Full-length predicted mRNA sequences corresponding to ExNeoEpitopes are evaluated sequentially in the UCSC Genome Browser via BLAT search (HG38) to confirm exon-region isoform specificity relative to prior annotated mRNA and ESTs and evaluate transposable element associations.

[00129] 3D topology evaluation: External to SNAF, we predict 3D conformational differences in protein folding and domain architecture using the software AlphaFold2 for each ExNeoEpitope relative to the reference isoform.

[00130] SNAF-B event filtering: Using this above filtering methods, SNAF-B shared-splicing neoantigens are defined on the basis of:

- a. Optimized RNA-sequence alignment strategy to define challenging to detect exon-exon junctions;
- b. SNAF-B predicted antigen in >15% of tumor patients;
- c. 2D and/or 3D folding prediction (Protter and/or AlphaFold);
- d. Extracellular domain impact predictions (custom code);
- e. SashimiPlot validation;
- f. Secondary evidence from long-read; and
- g. Protein isoform sequence is excluded from Ensembl and UniProt but may be present in TREMBL

[00131] Additional filtering steps which can be incorporated include:

- a. Novel-evidenced splicing event (Ensembl and GTEX);
- b. BayesTS $p < 3$ (default and custom tissue databases);
- c. Average control tissue count mean < 1 (default and custom tissue databases); and/or
- d. DeepImmuno score > 0.5 .

[00132] The following optional filtering steps can improve results and can be incorporated as well:

- a. Evidence from single-cell genomics;
- b. Significant patient survival associations with antigen presence; and/or

c. Evidence from proteomics.

[00133] The SNAF-B experimental workflow includes protein expression and transmembrane localization. To confirm the expression and cell membrane localization of SNAF-B novel isoforms, the long-read sequencing evidenced alternative isoforms and their annotated reference isoforms that are C-terminal tagged with either mNeon-Green or eGFP (VectorBuilder) are synthesized. These constructs (CMV promoter) are transfected with TransIT-LT1, separately into HEK-293T cells with 1 µg of plasmid (99% pUc19 dummy plasmid+1% engineered plasmid). Expression and co-localization are assessed with a phalloidin membrane stain after 24 hours, by confocal spinning disk microscopy (Yokogawa SoRa W1 dual camera system) using the Nikon Elements software.

[00134] The SNAF-T and SNAF-B workflows described above have a number of points of differentiation and tremendous advantages over current and previously described process. First, they are productionized and fully automated. Further, they provide the first validated transmembrane cancer isoforms that localize to the plasma membrane. SNAF-T and SNAF-B also represent the first probabilistic model of tumor specificity using a novel hierarchical Bayesian framework (BayesTS). While other workflows which have been described previously use arbitrary thresholds, this workflow considers tunable tissue weights for different tissues or cell types using a vastly improved background including normal tissue control datasets. These workflows have the additional advantage of using more accurate re-alignment of control cohorts with customized precision control cohorts for each study included to augment the default controls, to reduce batch effects. Notably, they also provide the highest possible validation rate for T-cell reactivity (100% of tested targets).

[00135] The SNAF-T and SNAF-B workflows described herein also provide the first identification of amino acid sequence bias for shared versus unique neoantigens and the first integrated long-read sequencing evidence of transmembrane isoforms. In addition, these workflows also allow for the evaluation of the clonal sources of splicing neoantigens from immune cell types and proliferative gene programs, and use the first splicing factor regulatory prediction (RNA-SPRINT) to assess the mechanistic origin of splicing neoantigens..

Antigenic Peptides

[00136] Various embodiments are directed to development of and use of antigenic peptides, such as the shared splicing neoantigens identified as described herein. In many

embodiments, antigenic peptides can be produced by chemical synthesis or by molecular expression in a host cell. Peptides can be purified and utilized in a variety of applications including (but not limited to) assays to determine peptide immunogenicity, assays to determine recognition by T cells, peptide vaccines for treatment of cancer, development of modified TCRs of T cells, development of antibodies, and development of CAR-T cells to recognize extracellular peptides.

[00137] Peptides can be synthesized chemically by a number of methods. One common method is to use solid-phase peptide synthesis (SPPS). Generally, SPPS is performed by repeating cycles of alternate N-terminal deprotection and coupling reactions, building peptides from the c-terminus to the n-terminus. The c-terminus of the first amino acid is coupled to the resin, wherein then the amine is deprotected and then coupled with the free acid of the second amino acid. This cycle repeats until the peptide is synthesized.

[00138] Peptides can also be synthesized utilizing molecular tools and a host cell. Nucleic acid sequences corresponding with antigenic peptides can be synthesized. In some embodiments, synthetic nucleic acids synthesized in in vitro synthesizers (e.g., phosphoramidite synthesizer), bacterial recombination system, or other suitable methods. Furthermore, synthesized nucleic acids can be purified and lyophilized, or kept stored in a biological system (e.g., bacteria, yeast). For use in a biological system, synthetic nucleic acid molecules can be inserted into a plasmid vector, or similar. A plasmid vector can also be an expression vector, wherein a suitable promoter and a suitable 3'-polyA tail is combined with the transcript sequence.

[00139] Embodiments include expression vectors and expression systems that produce antigenic peptides or proteins. These expression systems can incorporate an expression vector to express transcripts and proteins in a suitable expression system. Typical expression systems include bacterial (e.g., *E. coli*), insect (e.g., SF9), yeast (e.g., *S. cerevisiae*), animal (e.g., CHO), or human (e.g., HEK 293) cell lines. RNA and/or protein molecules can be purified from these systems using standard biotechnology production procedures.

[00140] Assays to determine immunogenicity and/or TCR binding can be performed. One such as is the dextramer flow cytometry assay. Generally, custom-made HLA-matched MHC Class I dextramer:peptide (pMHC) complexes are developed or purchased (Immudex, Copenhagen, Denmark). T cells from peripheral blood mononuclear cells (PBMCs) or tumor-

infiltrating lymphocytes (TILs) are incubated the pMHC complexes and stained, which are then run through a flow cytometer to determine if the peptide is capable of binding a TCR of a T cell.

Engineered T Cell Receptors

[00141] T-cell receptors comprise two different polypeptide chains, termed the T-cell receptor α (TCR α) and β (TCR β) chains, linked by a disulfide bond. These α : β heterodimers are very similar in structure to the Fab fragment of an immunoglobulin molecule, and they account for antigen recognition by most T cells. A minority of T cells bear an alternative, but structurally similar, receptor made up of a different pair of polypeptide chains designated γ and δ . Both types of T cell receptor differ from the membrane-bound immunoglobulin that serves as the B-cell receptor: a T cell receptor has only one antigen-binding site, whereas a B-cell receptor has two, and T-cell receptors are never secreted, whereas immunoglobulin can be secreted as antibody.

[00142] Both chains of the T-cell receptor have an amino-terminal variable (V) region with homology to an immunoglobulin V domain, a constant (C) region with homology to an immunoglobulin C domain, and a short hinge region containing a cysteine residue that forms the interchain disulfide bond. Each chain spans the lipid bilayer by a hydrophobic transmembrane domain, and ends in a short cytoplasmic tail.

[00143] The three-dimensional structure of the T-cell receptor has been determined. The structure is indeed similar to that of an antibody Fab fragment, as was suspected from earlier studies on the genes that encoded it. The T-cell receptor chains fold in much the same way as those of a Fab fragment, although the final structure appears a little shorter and wider. There are, however, some distinct differences between T-cell receptors and Fab fragments. The most striking difference is in the C α domain, where the fold is unlike that of any other immunoglobulin-like domain. The half of the domain that is juxtaposed with the C β domain forms a β sheet similar to that found in other immunoglobulin-like domains, but the other half of the domain is formed of loosely packed strands and a short segment of α helix. The intramolecular disulfide bond, which in immunoglobulin-like domains normally joins two β strands, in a C α domain joins a β strand to this segment of α helix.

[00144] There are also differences in the way in which the domains interact. The interface between the V and C domains of both T-cell receptor chains is more extensive than in antibodies, which may make the hinge joint between the domains less flexible. And the

interaction between the C α and C β domains is distinctive in being assisted by carbohydrate, with a sugar group from the C α domain making a number of hydrogen bonds to the C β domain. Finally, a comparison of the variable binding sites shows that, although the complementarity-determining region (CDR) loops align fairly closely with those of antibody molecules, there is some displacement relative to those of the antibody molecule. This displacement is particularly marked in the V α CDR2 loop, which is oriented at roughly right angles to the equivalent loop in antibody V domains, as a result of a shift in the β strand that anchors one end of the loop from one face of the domain to the other. A strand displacement also causes a change in the orientation of the V β CDR2 loop in two of the seven V β domains whose structures are known. As yet, the crystallographic structures of seven T cell receptors have been solved to this level of resolution.

[00145] Embodiments of the disclosure relate to engineered T cell receptors. The term “engineered” refers to T cell receptors that have TCR variable regions grafted onto TCR constant regions to make a chimeric polypeptide that binds to peptides and antigens of the disclosure. In certain embodiments, the TCR comprises intervening sequences that are used for cloning, enhanced expression, detection, or for therapeutic control of the construct, but are not present in endogenous TCRs, such as multiple cloning sites, linker, hinge sequences, modified hinge sequences, modified transmembrane sequences, a detection polypeptide or molecule, or therapeutic controls that may allow for selection or screening of cells comprising the TCR.

[00146] In some embodiments, the TCR comprises non-TCR sequences. Accordingly, certain embodiments relate to TCRs with sequences that are not from a TCR gene. In some embodiments, the TCR is chimeric, in that it contains sequences normally found in a TCR gene, but contains sequences from at least two TCR genes that are not necessarily found together in nature.

Antibodies

[00147] Various aspects of the disclosure include antibodies that target the peptides of the disclosure, or fragments thereof. The term “antibody” refers to an intact immunoglobulin of any isotype, or a fragment thereof that can compete with the intact antibody for specific binding to the target antigen, and includes chimeric, humanized, fully human, and bispecific antibodies. As used herein, the terms “antibody” or “immunoglobulin” are used interchangeably and refer to any of several classes of structurally related proteins that function as part of the immune

response of an animal, including IgG, IgD, IgE, IgA, IgM, and related proteins, as well as polypeptides comprising antibody CDR domains that retain antigen-binding activity.

[00148] The term “antigen” refers to a molecule or a portion of a molecule capable of being bound by a selective binding agent, such as an antibody. An antigen may possess one or more epitopes that are capable of interacting with different antibodies.

[00149] The term “epitope” includes any region or portion of molecule capable eliciting an immune response by binding to an immunoglobulin or to a T-cell receptor. Epitope determinants may include chemically active surface groups such as amino acids, sugar side chains, phosphoryl or sulfonyl groups, and may have specific three-dimensional structural characteristics and/or specific charge characteristics. Generally, antibodies specific for a particular target antigen will preferentially recognize an epitope on the target antigen within a complex mixture.

[00150] The epitope regions of a given polypeptide can be identified using many different epitope mapping techniques are well known in the art, including: x-ray crystallography, nuclear magnetic resonance spectroscopy, site-directed mutagenesis mapping, protein display arrays, see, e.g., *Epitope Mapping Protocols*, (Johan Rockberg and Johan Nilvebrant, Ed., 2018) Humana Press, New York, N.Y. Such techniques are known in the art and described in, e.g., U.S. Pat. No. 4,708,871; Geysen et al. *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); Geysen et al. *Proc. Natl. Acad. Sci. USA* 82:178-182 (1985); Geysen et al. *Molec. Immunol.* 23:709-715 (1986). Additionally, antigenic regions of proteins can also be predicted and identified using standard antigenicity and hydropathy plots.

[00151] The term “immunogenic sequence” means a molecule that includes an amino acid sequence of at least one epitope such that the molecule is capable of stimulating the production of antibodies in an appropriate host. The term “immunogenic composition” means a composition that comprises at least one immunogenic molecule (e.g., an antigen or carbohydrate).

[00152] An intact antibody is generally composed of two full-length heavy chains and two full length light chains, but in some instances may include fewer chains, such as antibodies naturally occurring in camelids that may comprise only heavy chains. Antibodies as disclosed herein may be derived solely from a single source or may be “chimeric,” that is, different portions of the antibody may be derived from two different antibodies. For example, the variable

or CDR regions may be derived from a rat or murine source, while the constant region is derived from a different animal source, such as a human. The antibodies or binding fragments may be produced in hybridomas, by recombinant DNA techniques, or by enzymatic or chemical cleavage of intact antibodies. Unless otherwise indicated, the term “antibody” includes derivatives, variants, fragments, and muteins thereof, examples of which are described below (Sela-Culang et al., Front Immunol. 2013; 4: 302; 2013).

[00153] The term “light chain” includes a full-length light chain and fragments thereof having sufficient variable region sequence to confer binding specificity. A full-length light chain has a molecular weight of around 25,000 Daltons and includes a variable region domain (abbreviated herein as VL), and a constant region domain (abbreviated herein as CL). There are two classifications of light chains, identified as kappa (κ) and lambda (λ). The term “VL fragment” means a fragment of the light chain of a monoclonal antibody that includes all or part of the light chain variable region, including CDRs. A VL fragment can further include light chain constant region sequences. The variable region domain of the light chain is at the amino-terminus of the polypeptide.

[00154] The term “heavy chain” includes a full-length heavy chain and fragments thereof having sufficient variable region sequence to confer binding specificity. A full-length heavy chain has a molecular weight of around 50,000 Daltons and includes a variable region domain (abbreviated herein as VH), and three constant region domains (abbreviated herein as CH1, CH2, and CH3). The term “VH fragment” means a fragment of the heavy chain of a monoclonal antibody that includes all or part of the heavy chain variable region, including CDRs. A VH fragment can further include heavy chain constant region sequences. The number of heavy chain constant region domains will depend on the isotype. The VH domain is at the amino-terminus of the polypeptide, and the CH domains are at the carboxy-terminus, with the CH3 being closest to the —COOH end. The isotype of an antibody can be IgM, IgD, IgG, IgA, or IgE and is defined by the heavy chains present of which there are five classifications: mu (μ), delta (δ), gamma (γ), alpha (α), or epsilon (ϵ) chains, respectively. IgG has several subtypes, including, but not limited to, IgG1, IgG2, IgG3, and IgG4. IgM subtypes include IgM1 and IgM2. IgA subtypes include IgA1 and IgA2.

1. Types of Antibodies

[00155] Antibodies can be whole immunoglobulins of any isotype or classification, chimeric antibodies, or hybrid antibodies with specificity to two or more antigens. They may also be fragments (e.g., F(ab')₂, Fab', Fab, Fv, and the like), including hybrid fragments. An immunoglobulin also includes natural, synthetic, or genetically engineered proteins that act like an antibody by binding to specific antigens to form a complex. The term antibody includes genetically engineered or otherwise modified forms of immunoglobulins.

[00156] The term “monomer” means an antibody containing only one Ig unit. Monomers are the basic functional units of antibodies. The term “dimer” means an antibody containing two Ig units attached to one another via constant domains of the antibody heavy chains (the Fc, or fragment crystallizable, region). The complex may be stabilized by a joining (J) chain protein. The term “multimer” means an antibody containing more than two Ig units attached to one another via constant domains of the antibody heavy chains (the Fc region). The complex may be stabilized by a joining (J) chain protein.

[00157] The term “bivalent antibody” means an antibody that comprises two antigen-binding sites. The two binding sites may have the same antigen specificities or they may be bi-specific, meaning the two antigen-binding sites have different antigen specificities.

[00158] Bispecific antibodies are a class of antibodies that have two paratopes with different binding sites for two or more distinct epitopes. In some embodiments, bispecific antibodies can be biparatopic, wherein a bispecific antibody may specifically recognize a different epitope from the same antigen. In some embodiments, bispecific antibodies can be constructed from a pair of different single domain antibodies termed “nanobodies”. Single domain antibodies are sourced and modified from cartilaginous fish and camelids. Nanobodies can be joined together by a linker using techniques typical to a person skilled in the art; such methods for selection and joining of nanobodies are described in PCT Publication No. WO2015044386A1, No. WO2010037838A2, and Bever et al., *Anal Chem.* 86:7875–7882 (2014), each of which are specifically incorporated herein by reference in their entirety.

[00159] Bispecific antibodies can be constructed as: a whole IgG, Fab'₂, Fab'PEG, a diabody, or alternatively as scFv. Diabodies and scFvs can be constructed without an Fc region, using only variable domains, potentially reducing the effects of anti-idiotypic reaction. Bispecific antibodies may be produced by a variety of methods including, but not limited to, fusion of hybridomas or linking of Fab' fragments. See, e.g., Songsivilai and Lachmann, *Clin. Exp.*

Immunol. 79:315-321 (1990); Kostelny et al., J. Immunol. 148:1547-1553 (1992), each of which are specifically incorporated by reference in their entirety.

[00160] In certain aspects, the antigen-binding domain may be multispecific or heterospecific by multimerizing with VH and VL region pairs that bind a different antigen. For example, the antibody may bind to, or interact with, (a) a cell surface antigen, (b) an Fc receptor on the surface of an effector cell, or (c) at least one other component. Accordingly, aspects may include, but are not limited to, bispecific, trispecific, tetraspecific, and other multispecific antibodies or antigen binding fragments thereof that are directed to epitopes and to other targets, such as Fc receptors on effector cells.

[00161] In some embodiments, multispecific antibodies can be used and directly linked via a short flexible polypeptide chain, using routine methods known in the art. One such example is diabodies that are bivalent, bispecific antibodies in which the VH and VL domains are expressed on a single polypeptide chain, and utilize a linker that is too short to allow for pairing between domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain creating two antigen binding sites. The linker functionality is applicable for embodiments of triabodies, tetrabodies, and higher order antibody multimers. (see, e.g., Hollinger et al., Proc Natl. Acad. Sci. USA 90:6444-6448 (1993); Polijak et al., Structure 2:1121-1123 (1994); Todorovska et al., J. Immunol. Methods 248:47-66 (2001)).

[00162] Bispecific diabodies, as opposed to bispecific whole antibodies, may also be advantageous because they can be readily constructed and expressed in *E. coli*. Diabodies (and other polypeptides such as antibody fragments) of appropriate binding specificities can be readily selected using phage display (WO94/13804) from libraries. If one arm of the diabody is kept constant, for instance, with a specificity directed against a protein, then a library can be made where the other arm is varied and an antibody of appropriate specificity selected. Bispecific whole antibodies may be made by alternative engineering methods as described in Ridgeway et al., (Protein Eng., 9:616-621, 1996) and Krah et al., (N Biotechnol. 39:167-173, 2017), each of which is hereby incorporated by reference in their entirety.

[00163] Heteroconjugate antibodies are composed of two covalently linked monoclonal antibodies with different specificities. See, e.g., U.S. Patent No. 6,010,902, incorporated herein by reference in its entirety.

[00164] The part of the Fv fragment of an antibody molecule that binds with high specificity to the epitope of the antigen is referred to herein as the “paratope.” The paratope consists of the amino acid residues that make contact with the epitope of an antigen to facilitate antigen recognition. Each of the two Fv fragments of an antibody is composed of the two variable domains, VH and VL, in dimerized configuration. The primary structure of each of the variable domains includes three hypervariable loops separated by, and flanked by, Framework Regions (FR). The hypervariable loops are the regions of highest primary sequences variability among the antibody molecules from any mammal. The term hypervariable loop is sometimes used interchangeably with the term “Complementarity Determining Region (CDR).” The length of the hypervariable loops (or CDRs) varies between antibody molecules. The framework regions of all antibody molecules from a given mammal have high primary sequence similarity/consensus. The consensus of framework regions can be used by one skilled in the art to identify both the framework regions and the hypervariable loops (or CDRs) which are interspersed among the framework regions. The hypervariable loops are given identifying names which distinguish their position within the polypeptide, and on which domain they occur. CDRs in the VL domain are identified as L1, L2, and L3, with L1 occurring at the most distal end and L3 occurring closest to the CL domain. The CDRs may also be given the names CDR-1, CDR-2, and CDR-3. The L3 (CDR-3) is generally the region of highest variability among all antibody molecules produced by a given organism. The CDRs are regions of the polypeptide chain arranged linearly in the primary structure, and separated from each other by Framework Regions. The amino terminal (N-terminal) end of the VL chain is named FR1. The region identified as FR2 occurs between L1 and L2 hypervariable loops. FR3 occurs between L2 and L3 hypervariable loops, and the FR4 region is closest to the CL domain. This structure and nomenclature is repeated for the VH chain, which includes three CDRs identified as H1, H2 and H3. The majority of amino acid residues in the variable domains, or Fv fragments (VH and VL), are part of the framework regions (approximately 85%). The three dimensional, or tertiary, structure of an antibody molecule is such that the framework regions are more internal to the molecule and provide the majority of the structure, with the CDRs on the external surface of the molecule.

[00165] Several methods have been developed and can be used by one skilled in the art to identify the exact amino acids that constitute each of these regions. This can be done using

any of a number of multiple sequence alignment methods and algorithms, which identify the conserved amino acid residues that make up the framework regions, therefore identifying the CDRs that may vary in length but are located between framework regions. Three commonly used methods have been developed for identification of the CDRs of antibodies: Kabat (as described in T. T. Wu and E. A. Kabat, "AN ANALYSIS OF THE SEQUENCES OF THE VARIABLE REGIONS OF BENGE JONES PROTEINS AND MYELOMA LIGHT CHAINS AND THEIR IMPLICATIONS FOR ANTIBODY COMPLEMENTARITY," J Exp Med, vol. 132, no. 2, pp. 211–250, Aug. 1970); Chothia (as described in C. Chothia et al., "Conformations of immunoglobulin hypervariable regions," Nature, vol. 342, no. 6252, pp. 877–883, Dec. 1989); and IMGT (as described in M.-P. Lefranc et al., "IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains," Developmental & Comparative Immunology, vol. 27, no. 1, pp. 55–77, Jan. 2003). These methods each include unique numbering systems for the identification of the amino acid residues that constitute the variable regions. In most antibody molecules, the amino acid residues that actually contact the epitope of the antigen occur in the CDRs, although in some cases, residues within the framework regions contribute to antigen binding.

[00166] One skilled in the art can use any of several methods to determine the paratope of an antibody. These methods include: 1) Computational predictions of the tertiary structure of the antibody/epitope binding interactions based on the chemical nature of the amino acid sequence of the antibody variable region and composition of the epitope. 2) Hydrogen-deuterium exchange and mass spectroscopy 3) Polypeptide fragmentation and peptide mapping approaches in which one generates multiple overlapping peptide fragments from the full length of the polypeptide and evaluates the binding affinity of these peptides for the epitope. 4) Antibody Phage Display Library analysis in which the antibody Fab fragment encoding genes of the mammal are expressed by bacteriophage in such a way as to be incorporated into the coat of the phage. This population of Fab expressing phage are then allowed to interact with the antigen which has been immobilized or may be expressed in by a different exogenous expression system. Non-binding Fab fragments are washed away, thereby leaving only the specific binding Fab fragments attached to the antigen. The binding Fab fragments can be readily isolated and the genes which encode them determined. This approach can also be used for smaller regions of the Fab fragment including Fv fragments or specific VH and VL domains as appropriate.

[00167] In certain aspects, affinity matured antibodies are enhanced with one or more modifications in one or more CDRs thereof that result in an improvement in the affinity of the antibody for a target antigen as compared to a parent antibody that does not possess those alteration(s). Certain affinity matured antibodies will have nanomolar or picomolar affinities for the target antigen. Affinity matured antibodies are produced by procedures known in the art, e.g., Marks et al., *Bio/Technology* 10:779 (1992) describes affinity maturation by VH and VL domain shuffling, random mutagenesis of CDR and/or framework residues employed in phage display is described by Rajpal et al., *PNAS*. 24: 8466-8471 (2005) and Thie et al., *Methods Mol Biol.* 525:309-22 (2009) in conjugation with computation methods as demonstrated in Tiller et al., *Front. Immunol.* 8:986 (2017).

[00168] Chimeric immunoglobulins are the products of fused genes derived from different species; “humanized” chimeras generally have the framework region (FR) from human immunoglobulins and one or more CDRs are from a non-human source.

[00169] In certain aspects, portions of the heavy and/or light chain are identical or homologous to corresponding sequences from another particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired biological activity. U.S. Pat. No. 4,816,567; and Morrison et al., *Proc. Natl. Acad. Sci. USA* 81:6851 (1984). For methods relating to chimeric antibodies, see, e.g., U.S. Pat. No. 4,816,567; and Morrison et al., *Proc. Natl. Acad. Sci. USA* 81:6851-6855 (1985), each of which are specifically incorporated herein by reference in their entirety. CDR grafting is described, for example, in U.S. Pat. Nos. 6,180,370, 5,693,762, 5,693,761, 5,585,089, and 5,530,101, which are all hereby incorporated by reference for all purposes.

[00170] In some embodiments, minimizing the antibody polypeptide sequence from the non-human species optimizes chimeric antibody function and reduces immunogenicity. Specific amino acid residues from non-antigen recognizing regions of the non-human antibody are modified to be homologous to corresponding residues in a human antibody or isotype. One example is the “CDR grafted” antibody, in which an antibody comprises one or more CDRs from a particular species or belonging to a specific antibody class or subclass, while the remainder of the antibody chain(s) is identical or homologous to a corresponding sequence in antibodies

derived from another species or belonging to another antibody class or subclass. For use in humans, the V region composed of CDR1, CDR2, and partial CDR3 for both the light and heavy chain variance region from a non-human immunoglobulin, are grafted with a human antibody framework region, replacing the naturally occurring antigen receptors of the human antibody with the non-human CDRs. In some instances, corresponding non-human residues replace framework region residues of the human immunoglobulin. Furthermore, humanized antibodies may comprise residues that are not found in the recipient antibody or in the donor antibody to further refine performance. The humanized antibody may also comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. See, e.g., Jones et al., *Nature* 321:522 (1986); Riechmann et al., *Nature* 332:323 (1988); Presta, *Curr. Op. Struct. Biol.* 2:593 (1992); Vaswani and Hamilton, *Ann. Allergy, Asthma and Immunol.* 1:105 (1998); Harris, *Biochem. Soc. Transactions* 23; 1035 (1995); Hurlle and Gross, *Curr. Op. Biotech.* 5:428 (1994); Verhoeyen et al., *Science* 239:1534-36 (1988).

[00171] Intrabodies are intracellularly localized immunoglobulins that bind to intracellular antigens as opposed to secreted antibodies, which bind antigens in the extracellular space.

[00172] Polyclonal antibody preparations typically include different antibodies against different determinants (epitopes). In order to produce polyclonal antibodies, a host, such as a rabbit or goat, is immunized with the antigen or antigen fragment, generally with an adjuvant and, if necessary, coupled to a carrier. Antibodies to the antigen are subsequently collected from the sera of the host. The polyclonal antibody can be affinity purified against the antigen rendering it monospecific.

[00173] Monoclonal antibodies or “mAb” refer to an antibody obtained from a population of homogeneous antibodies from an exclusive parental cell, e.g., the population is identical except for naturally occurring mutations that may be present in minor amounts. Each monoclonal antibody is directed against a single antigenic determinant.

2. Functional Antibody Fragments and Antigen-Binding Fragments

[00174] Certain aspects relate to antibody fragments, such as antibody fragments that bind to a peptide of the disclosure. The term functional antibody fragment includes antigen-binding fragments of an antibody that retain the ability to specifically bind to an antigen. These fragments are constituted of various arrangements of the variable region heavy chain (VH)

and/or light chain (VL); and in some embodiments, include constant region heavy chain 1 (CH1) and light chain (CL). In some embodiments, they lack the Fc region constituted of heavy chain 2 (CH2) and 3 (CH3) domains. Embodiments of antigen binding fragments and the modifications thereof may include: (i) the Fab fragment type constituted with the VL, VH, CL, and CH1 domains; (ii) the Fd fragment type constituted with the VH and CH1 domains; (iii) the Fv fragment type constituted with the VH and VL domains; (iv) the single domain fragment type, dAb, (Ward, 1989; McCafferty et al., 1990; Holt et al., 2003) constituted with a single VH or VL domain; (v) isolated complementarity determining region (CDR) regions. Such terms are described, for example, in Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, NY (1989); Molec. Biology and Biotechnology: A Comprehensive Desk Reference (Myers, R. A. (ed.), New York: VCH Publisher, Inc.); Huston et al., *Cell Biophysics*, 22:189-224 (1993); Pluckthun and Skerra, *Meth. Enzymol.*, 178:497-515 (1989) and in Day, E. D., *Advanced Immunochemistry*, 2d ed., Wiley Liss, Inc. New York, N.Y. (1990); *Antibodies*, 4:259-277 (2015), each of which are incorporated by reference.

[00175] Antigen-binding fragments also include fragments of an antibody that retain exactly, at least, or at most 1, 2, or 3 complementarity determining regions (CDRs) from a light chain variable region. Fusions of CDR-containing sequences to an Fc region (or a CH2 or CH3 region thereof) are included within the scope of this definition including, for example, scFv fused, directly or indirectly, to an Fc region are included herein.

[00176] The term Fab fragment means a monovalent antigen-binding fragment of an antibody containing the VL, VH, CL and CH1 domains. The term Fab' fragment means a monovalent antigen-binding fragment of a monoclonal antibody that is larger than a Fab fragment. For example, a Fab' fragment includes the VL, VH, CL and CH1 domains and all or part of the hinge region. The term F(ab')₂ fragment means a bivalent antigen-binding fragment of a monoclonal antibody comprising two Fab' fragments linked by a disulfide bridge at the hinge region. An F(ab')₂ fragment includes, for example, all or part of the two VH and VL domains, and can further include all or part of the two CL and CH1 domains.

[00177] The term Fd fragment means a fragment of the heavy chain of a monoclonal antibody, which includes all or part of the VH, including the CDRs. An Fd fragment can further include CH1 region sequences.

[00178] The term Fv fragment means a monovalent antigen-binding fragment of a monoclonal antibody, including all or part of the VL and VH, and absent of the CL and CH1 domains. The VL and VH include, for example, the CDRs. Single-chain antibodies (sFv or scFv) are Fv molecules in which the VL and VH regions have been connected by a flexible linker to form a single polypeptide chain, which forms an antigen-binding fragment. Single chain antibodies are discussed in detail in International Patent Application Publication No. WO 88/01649 and U.S. Pat. Nos. 4,946,778 and 5,260,203, the disclosures of which are herein incorporated by reference. The term (scFv)₂ means bivalent or bispecific sFv polypeptide chains that include oligomerization domains at their C-termini, separated from the sFv by a hinge region (Pack et al. 1992). The oligomerization domain comprises self-associating α -helices, e.g., leucine zippers, which can be further stabilized by additional disulfide bonds. (scFv)₂ fragments are also known as “miniantibodies” or “minibodies.”

[00179] A single domain antibody is an antigen-binding fragment containing only a VH or the VL domain. In some instances, two or more VH regions are covalently joined with a peptide linker to create a bivalent domain antibody. The two VH regions of a bivalent domain antibody may target the same or different antigens.

[00180] An Fc region (Fragment Crystallizable Region) contains two heavy chain fragments comprising the CH2 and CH3 domains of an antibody. The two heavy chain fragments are held together by two or more disulfide bonds and by hydrophobic interactions of the CH3 domains. The term “Fc polypeptide” as used herein includes native and mutein forms of polypeptides derived from the Fc region of an antibody. Truncated forms of such polypeptides containing the hinge region that promotes dimerization are included.

[00181] Polypeptides can be combined or integrated with antibody complementarity-determining regions (CDRs) and scaffolding domains that display the CDRs. Antigen-binding peptide scaffolds, such as CDRs, are used to generate protein-binding molecules in accordance with the embodiments. Generally, a person skilled in the art can determine the type of protein scaffold on which to graft at least one of the CDRs. It is known that scaffolds, optimally, must meet a number of criteria such as: good phylogenetic conservation; known three-dimensional structure; small size; few or no post transcriptional modifications; and/or be easy to produce, express, and purify. Skerra, J Mol Recognit, 13:167-87 (2000).

[00182] The protein scaffolds can be sourced from, but not limited to: fibronectin type III FN3 domain (known as “monobodies”), fibronectin type III domain 10, lipocalin, anticalin, Z- domain of protein A of *Staphylococcus aureus*, thioredoxin A or proteins with a repeated motif such as the “ankyrin repeat”, the “armadillo repeat”, the “leucine-rich repeat” and the “tetratricopeptide repeat”. Such proteins are described in US Patent Publication Nos. 2010/0285564, 2006/0058510, 2006/0088908, 2005/0106660, and PCT Publication No. WO2006/056464, each of which are specifically incorporated herein by reference in their entirety. Scaffolds derived from toxins from scorpions, insects, plants, mollusks, etc., and the protein inhibitors of neuronal NO synthase (PIN) may also be used.

3. Antibody Binding

[00183] The term “selective binding agent” refers to a molecule that binds to an antigen. Non limiting examples include antibodies, antigen-binding fragments, scFv, Fab, Fab', F(ab')₂, single chain antibodies, peptides, peptide fragments and proteins.

[00184] The term “binding” refers to a direct association between two molecules, due to, for example, covalent, electrostatic, hydrophobic, and ionic and/or hydrogen-bond interactions, including interactions such as salt bridges and water bridges. “Immunologically reactive” means that the selective binding agent or antibody of interest will bind with antigens present in a biological sample. The term “immune complex” refers the combination formed when an antibody or selective binding agent binds to an epitope on an antigen.

4. Affinity/Avidity

[00185] The term “affinity” refers the strength with which an antibody or selective binding agent binds an epitope. In antibody binding reactions, this is expressed as the affinity constant (K_a or k_a sometimes referred to as the association constant) for any given antibody or selective binding agent. Affinity is measured as a comparison of the binding strength of the antibody to its antigen relative to the binding strength of the antibody to an unrelated amino acid sequence. Affinity can be expressed as, for example, 20- fold greater binding ability of the antibody to its antigen then to an unrelated amino acid sequence. As used herein, the term “avidity” refers to the resistance of a complex of two or more agents to dissociation after dilution. The terms “immunoreactive” and “preferentially binds” are used interchangeably herein with respect to antibodies and/or selective binding agent.

[00186] There are several experimental methods that can be used by one skilled in the art to evaluate the binding affinity of any given antibody or selective binding agent for its antigen. This is generally done by measuring the equilibrium dissociation constant (K_D or K_d), using the equation $K_D = k_{off} / k_{on} = [A][B]/[AB]$. The term k_{off} is the rate of dissociation between the antibody and antigen per unit time, and is related to the concentration of antibody and antigen present in solution in the unbound form at equilibrium. The term k_{on} is the rate of antibody and antigen association per unit time, and is related to the concentration of the bound antigen-antibody complex at equilibrium. The units used for measuring the K_D are mol/L (molarity, or M), or concentration. The K_a of an antibody is the opposite of the K_D , and is determined by the equation $K_a = 1/K_D$. Examples of some experimental methods that can be used to determine the K_D value are: enzyme-linked immunosorbent assays (ELISA), isothermal titration calorimetry (ITC), fluorescence anisotropy, surface plasmon resonance (SPR), and affinity capillary electrophoresis (ACE). The affinity constant (K_a) of an antibody is the opposite of the K_D , and is determined by the equation $K_a = 1/ K_D$.

[00187] Antibodies deemed useful in certain embodiments may have an affinity constant (K_a) of about, at least about, or at most about 10^6 , 10^7 , 10^8 , 10^9 , or 10^{10} M or any range derivable therein. Similarly, in some embodiments, antibodies may have a dissociation constant of about, at least about or at most about 10^{-6} , 10^{-7} , 10^{-8} , 10^{-9} , 10^{-10} M, or any range derivable therein. These values are reported for antibodies discussed herein and the same assay may be used to evaluate the binding properties of such antibodies. An antibody of the invention is said to “specifically bind” its target antigen when the dissociation constant (K_D) is $\leq 10^{-8}$ M. The antibody specifically binds antigen with “high affinity” when the K_D is $\leq 5 \times 10^{-9}$ M, and with “very high affinity” when the K_D is $\leq 5 \times 10^{-10}$ M.

5. Epitope Specificity

[00188] The epitope of an antigen is the specific region of the antigen for which an antibody has binding affinity. In the case of protein or polypeptide antigens, the epitope is the specific residues (or specified amino acids or protein segment) that the antibody binds with high affinity. An antibody does not necessarily contact every residue within the protein. Nor does every single amino acid substitution or deletion within a protein necessarily affect binding affinity. For purposes of this specification and the accompanying claims, the terms “epitope” and “antigenic determinant” are used interchangeably to refer to the site on an antigen to which B

and/or T cells respond or recognize. Polypeptide epitopes can be formed from both contiguous amino acids and noncontiguous amino acids juxtaposed by tertiary folding of a polypeptide. An epitope typically includes at least 3, and typically 5-10 amino acids in a unique spatial conformation.

[00189] Epitope specificity of an antibody can be determined in a variety of ways. One approach, for example, involves testing a collection of overlapping peptides of about 15 amino acids spanning the full sequence of the protein and differing in increments of a small number of amino acids (e.g., 3 to 30 amino acids). The peptides are immobilized in separate wells of a microtiter dish. Immobilization can be accomplished, for example, by biotinylating one terminus of the peptides. This process may affect the antibody affinity for the epitope, therefore different samples of the same peptide can be biotinylated at the N and C terminus and immobilized in separate wells for the purposes of comparison. This is useful for identifying end-specific antibodies. Optionally, additional peptides can be included terminating at a particular amino acid of interest. This approach is useful for identifying end-specific antibodies to internal fragments. An antibody or antigen binding fragment is screened for binding to each of the various peptides. The epitope is defined as a segment of amino acids that is common to all peptides to which the antibody shows high affinity binding.

6. Modification of Antibody Antigen-Binding Domains

[00190] It is understood that the antibodies of the present invention may be modified, such that they are substantially identical to the antibody polypeptide sequences, or fragments thereof, and still bind the epitopes of the present invention. Polypeptide sequences are “substantially identical” when optimally aligned using such programs as Clustal Omega, IGBLAST, GAP or BESTFIT using default gap weights, they share at least 80% sequence identity, at least 90% sequence identity, at least 95% sequence identity, at least 96% sequence identity, at least 97% sequence identity, at least 98% sequence identity, or at least 99% sequence identity or any range therein.

[00191] As discussed herein, minor variations in the amino acid sequences of antibodies or antigen-binding regions thereof are contemplated as being encompassed by the present invention, providing that the variations in the amino acid sequence maintain at least 75%, more preferably at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%

and most preferably at least 99% sequence identity. In particular, conservative amino acid replacements are contemplated.

[00192] Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are generally divided into families based on the chemical nature of the side chain; e.g., acidic (aspartate, glutamate), basic (lysine, arginine, histidine), nonpolar (alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), and uncharged polar (glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine). For example, it is reasonable to expect that an isolated replacement of a leucine moiety with an isoleucine or valine moiety, or a similar replacement of an amino acid with a structurally related amino acid in the same family, will not have a major effect on the binding or properties of the resulting molecule, especially if the replacement does not involve an amino acid within a framework site. Whether an amino acid change results in a functional peptide can readily be determined by assaying the specific activity of the polypeptide derivative. Standard ELISA, Surface Plasmon Resonance (SPR), or other antibody binding assays can be performed by one skilled in the art to make a quantitative comparison of antigen binding affinity between the unmodified antibody and any polypeptide derivatives with conservative substitutions generated through any of several methods available to one skilled in the art.

[00193] Fragments or analogs of antibodies or immunoglobulin molecules can be readily prepared by those skilled in the art. Preferred amino- and carboxy-termini of fragments or analogs occur near boundaries of functional domains. Structural and functional domains can be identified by comparison of the nucleotide and/or amino acid sequence data to public or proprietary sequence databases. Preferably, computerized comparison methods are used to identify sequence motifs or predicted protein conformation domains that occur in other proteins of known structure and/or function. Standard methods to identify protein sequences that fold into a known three-dimensional structure are available to those skilled in the art; Dill and McCallum., Science 338:1042-1046 (2012). Several algorithms for predicting protein structures and the gene sequences that encode these have been developed, and many of these algorithms can be found at the National Center for Biotechnology Information (on the World Wide Web at ncbi.nlm.nih.gov/guide/proteins/) and at the Bioinformatics Resource Portal (on the World Wide Web at expasy.org/proteomics). Thus, the foregoing examples demonstrate that those of skill in

the art can recognize sequence motifs and structural conformations that may be used to define structural and functional domains in accordance with the invention.

[00194] Framework modifications can be made to antibodies to decrease immunogenicity, for example, by “backmutating” one or more framework residues to a corresponding germline sequence.

[00195] It is also contemplated that the antigen-binding domain may be multi-specific or multivalent by multimerizing the antigen-binding domain with VH and VL region pairs that bind either the same antigen (multi-valent) or a different antigen (multi-specific).

Proteinaceous Compositions

[00196] As used herein, a “protein” “peptide” or “polypeptide” refers to a molecule comprising at least five amino acid residues. As used herein, the term “wild-type” refers to the endogenous version of a molecule that occurs naturally in an organism. In some embodiments, wild-type versions of a protein or polypeptide are employed, however, in many embodiments of the disclosure, a modified protein or polypeptide is employed to generate an immune response. The terms described above may be used interchangeably. A “modified protein” or “modified polypeptide” or a “variant” refers to a protein or polypeptide whose chemical structure, particularly its amino acid sequence, is altered with respect to the wild-type protein or polypeptide. In some embodiments, a modified/variant protein or polypeptide has at least one modified activity or function (recognizing that proteins or polypeptides may have multiple activities or functions). It is specifically contemplated that a modified/variant protein or polypeptide may be altered with respect to one activity or function yet retain a wild-type activity or function in other respects, such as immunogenicity.

[00197] Where a protein is specifically mentioned herein, it is in general a reference to a native (wild-type) or recombinant (modified) protein or, optionally, a protein in which any signal sequence has been removed. The protein may be isolated directly from the organism of which it is native, produced by recombinant DNA/exogenous expression methods, or produced by solid phase peptide synthesis (SPPS) or other in vitro methods. In particular embodiments, there are isolated nucleic acid segments and recombinant vectors incorporating nucleic acid sequences that encode a polypeptide (e.g., an antibody or fragment thereof). The term “recombinant” may be used in conjunction with a polypeptide or the name of a specific

polypeptide, and this generally refers to a polypeptide produced from a nucleic acid molecule that has been manipulated in vitro or that is a replication product of such a molecule.

[00198] In certain embodiments the size of a protein or polypeptide (wild-type or modified) may comprise, but is not limited to, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2250, 2500 amino acid residues or greater, and any range derivable therein, or derivative of a corresponding amino sequence described or referenced herein. It is contemplated that polypeptides may be mutated by truncation, rendering them shorter than their corresponding wild-type form, also, they might be altered by fusing or conjugating a heterologous protein or polypeptide sequence with a particular function (e.g., for targeting or localization, for enhanced immunogenicity, for purification purposes, etc.).

[00199] The polypeptides, proteins, or polynucleotides encoding such polypeptides or proteins of the disclosure may include 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, or 50 (or any derivable range therein) or more variant amino acids or nucleic acid substitutions or be at least 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% (or any derivable range therein) similar, identical, or homologous with, with at least, or with at most 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163,

164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 300, 400, 500, 550, 1000 or more contiguous amino acids or nucleic acids, or any range derivable therein, of SEQ ID Nos:1-1403. In specific embodiments, the peptide or polypeptide is or is based on a human sequence. In certain embodiments, the peptide or polypeptide is not naturally occurring and/or is in a combination of peptides or polypeptides.

[00200] In some embodiments, a peptide or polypeptide described herein comprises, comprises at least, or comprises at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 substitutions (or any derivable range therein) at amino acid position 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, or 130 (or any range derivable therein) of SEQ ID NOS:1-1403. In some embodiments, the amino acid at position 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, or 130 of a peptide or polypeptide of SEQ ID NO:1-1403 is substituted with an alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, or valine.

[00201] In some embodiments, the protein or polypeptide may comprise amino acids 1 to 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81,

82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675,

676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, or 1000, (or any derivable range therein) of SEQ ID NOs:1-1403.

[00202] In some embodiments, the protein, polypeptide, or nucleic acid may comprise 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257,

258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846,

847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, or 1000, (or any derivable range therein) contiguous amino acids of SEQ ID NOs: 1-1403.

[00203] In some embodiments, the polypeptide, protein, or nucleic acid may comprise, comprise at least, comprises at most, or comprise about 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417,

418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, or 1000 (or any derivable range

therein) contiguous amino acids of SEQ ID Nos:1-1403 that are, are at least, are at most, are exactly, or are about 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% (or any derivable range therein) similar, identical, or homologous with one of SEQ ID NOS:1-1403.

[00204] In some aspects there is a nucleic acid molecule or polypeptide starting at position 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, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501,

502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, or 1000 of any of SEQ ID NOS:1-1403 and comprising, comprising at least, comprising at most, or comprising about 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96,

97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686,

687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, or 1000 (or any derivable range therein) contiguous amino acids or nucleotides of any of SEQ ID NOS:1-1403.

[00205] The nucleotide as well as the protein, polypeptide, and peptide sequences for various genes have been previously disclosed, and may be found in the recognized computerized databases. Two commonly used databases are the National Center for Biotechnology Information's Genbank and GenPept databases (on the World Wide Web at ncbi.nlm.nih.gov/) and The Universal Protein Resource (UniProt; on the World Wide Web at uniprot.org). The coding regions for these genes may be amplified and/or expressed using the techniques disclosed herein or as would be known to those of ordinary skill in the art.

[00206] It is contemplated that in compositions of the disclosure, there is between about 0.001 mg and about 10 mg of total polypeptide, peptide, and/or protein per ml. The concentration of protein in a composition can be about, at least about or at most about 0.001, 0.010, 0.050, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0 mg/ml or more (or any range derivable therein).

[00207] The following is a discussion of changing the amino acid subunits of a protein to create an equivalent, or even improved, second-generation variant polypeptide or peptide. For example, certain amino acids may be substituted for other amino acids in a protein or polypeptide sequence with or without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's functional activity, certain amino acid substitutions can be made in a protein sequence and in its corresponding DNA coding sequence, and nevertheless produce a protein with similar or desirable properties. It is thus contemplated by the inventors that various changes may be made in the DNA sequences of genes which encode proteins without appreciable loss of their biological utility or activity.

[00208] The term "functionally equivalent codon" is used herein to refer to codons that encode the same amino acid, such as the six different codons for arginine. Also considered are "neutral substitutions" or "neutral mutations" which refers to a change in the codon or codons that encode biologically equivalent amino acids.

[00209] Amino acid sequence variants of the disclosure can be substitutional, insertional, or deletion variants. A variation in a polypeptide of the disclosure may affect 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, 50, or more non-contiguous or contiguous amino acids of the protein or polypeptide, as compared to wild-type. A variant can comprise an amino acid sequence that is at least 50%, 60%, 70%, 80%, or 90%, including all values and ranges there between, identical to any sequence provided or referenced herein. A variant can include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more substitute amino acids.

[00210] It also will be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids, or 5' or 3' sequences, respectively, and yet still be essentially identical as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region.

[00211] Deletion variants typically lack one or more residues of the native or wild type protein. Individual residues can be deleted or a number of contiguous amino acids can be deleted. A stop codon may be introduced (by substitution or insertion) into an encoding nucleic acid sequence to generate a truncated protein.

[00212] Insertional mutants typically involve the addition of amino acid residues at a non-terminal point in the polypeptide. This may include the insertion of one or more amino acid residues. Terminal additions may also be generated and can include fusion proteins which are multimers or concatemers of one or more peptides or polypeptides described or referenced herein.

[00213] Substitutional variants typically contain the exchange of one amino acid for another at one or more sites within the protein or polypeptide, and may be designed to modulate one or more properties of the polypeptide, with or without the loss of other functions or properties. Substitutions may be conservative, that is, one amino acid is replaced with one of similar chemical properties. "Conservative amino acid substitutions" may involve exchange of a member of one amino acid class with another member of the same class. Conservative substitutions are well known in the art and include, for example, the changes of: alanine to serine; arginine to lysine; asparagine to glutamine or histidine; aspartate to glutamate; cysteine to serine; glutamine to asparagine; glutamate to aspartate; glycine to proline; histidine to asparagine or glutamine; isoleucine to leucine or valine; leucine to valine or isoleucine; lysine to arginine; methionine to leucine or isoleucine; phenylalanine to tyrosine, leucine or methionine; serine to threonine; threonine to serine; tryptophan to tyrosine; tyrosine to tryptophan or phenylalanine; and valine to isoleucine or leucine. Conservative amino acid substitutions may encompass non-naturally occurring amino acid residues, which are typically incorporated by chemical peptide synthesis rather than by synthesis in biological systems. These include peptidomimetics or other reversed or inverted forms of amino acid moieties.

[00214] Alternatively, substitutions may be "non-conservative", such that a function or activity of the polypeptide is affected. Non-conservative changes typically involve substituting an amino acid residue with one that is chemically dissimilar, such as a polar or charged amino acid for a nonpolar or uncharged amino acid, and vice versa. Non-conservative substitutions may involve the exchange of a member of one of the amino acid classes for a member from another class.

[00215] One skilled in the art can determine suitable variants of polypeptides as set forth herein using well-known techniques. One skilled in the art may identify suitable areas of the molecule that may be changed without destroying activity by targeting regions not believed to be important for activity. The skilled artisan will also be able to identify amino acid residues and portions of the molecules that are conserved among similar proteins or polypeptides. In further embodiments, areas that may be important for biological activity or for structure may be subject to conservative amino acid substitutions without significantly altering the biological activity or without adversely affecting the protein or polypeptide structure.

[00216] In making such changes, the hydropathy index of amino acids may be considered. The hydropathy profile of a protein is calculated by assigning each amino acid a numerical value ("hydropathy index") and then repetitively averaging these values along the peptide chain. Each amino acid has been assigned a value based on its hydrophobicity and charge characteristics. They are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cysteine (+2.5); methionine (+1.9); alanine (+1.8); glycine (−0.4); threonine (−0.7); serine (−0.8); tryptophan (−0.9); tyrosine (−1.3); proline (1.6); histidine (−3.2); glutamate (−3.5); glutamine (−3.5); aspartate (−3.5); asparagine (−3.5); lysine (−3.9); and arginine (−4.5). The importance of the hydropathy amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte et al., J. Mol. Biol. 157:105-131 (1982)). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein or polypeptide, which in turn defines the interaction of the protein or polypeptide with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and others. It is also known that certain amino acids may be substituted for other amino acids having a similar hydropathy index or score, and still retain a similar biological activity. In making changes based upon the hydropathy index, in certain embodiments, the substitution of amino acids whose hydropathy indices are within ± 2 is included. In some aspects of the invention, those that are within ± 1 are included, and in other aspects of the invention, those within ± 0.5 are included.

[00217] It also is understood in the art that the substitution of like amino acids can be effectively made based on hydrophilicity. U.S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein. In

certain embodiments, the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigen binding, that is, as a biological property of the protein. The following hydrophilicity values have been assigned to these amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 $\pm$ 1); glutamate (+3.0 $\pm$ 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (−0.4); proline (−0.5 $\pm$ 1); alanine (−0.5); histidine (−0.5); cysteine (−1.0); methionine (−1.3); valine (−1.5); leucine (−1.8); isoleucine (−1.8); tyrosine (−2.3); phenylalanine (−2.5); and tryptophan (−3.4). In making changes based upon similar hydrophilicity values, in certain embodiments, the substitution of amino acids whose hydrophilicity values are within ± 2 are included, in other embodiments, those which are within ± 1 are included, and in still other embodiments, those within ± 0.5 are included. In some instances, one may also identify epitopes from primary amino acid sequences based on hydrophilicity. These regions are also referred to as “epitopic core regions.” It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still produce a biologically equivalent and immunologically equivalent protein.

[00218] Additionally, one skilled in the art can review structure-function studies identifying residues in similar polypeptides or proteins that are important for activity or structure. In view of such a comparison, one can predict the importance of amino acid residues in a protein that correspond to amino acid residues important for activity or structure in similar proteins. One skilled in the art may opt for chemically similar amino acid substitutions for such predicted important amino acid residues.

[00219] One skilled in the art can also analyze the three-dimensional structure and amino acid sequence in relation to that structure in similar proteins or polypeptides. In view of such information, one skilled in the art may predict the alignment of amino acid residues of an antibody with respect to its three-dimensional structure. One skilled in the art may choose not to make changes to amino acid residues predicted to be on the surface of the protein, since such residues may be involved in important interactions with other molecules. Moreover, one skilled in the art may generate test variants containing a single amino acid substitution at each desired amino acid residue. These variants can then be screened using standard assays for binding and/or activity, thus yielding information gathered from such routine experiments, which may allow one skilled in the art to determine the amino acid positions where further substitutions should be

avoided either alone or in combination with other mutations. Various tools available to determine secondary structure can be found on the world wide web at expasy.org/proteomics/protein_structure.

[00220] In some embodiments of the invention, amino acid substitutions are made that: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter ligand or antigen binding affinities, and/or (5) confer or modify other physicochemical or functional properties on such polypeptides. For example, single or multiple amino acid substitutions (in certain embodiments, conservative amino acid substitutions) may be made in the naturally occurring sequence. Substitutions can be made in that portion of the antibody that lies outside the domain(s) forming intermolecular contacts. In such embodiments, conservative amino acid substitutions can be used that do not substantially change the structural characteristics of the protein or polypeptide (e.g., one or more replacement amino acids that do not disrupt the secondary structure that characterizes the native antibody).

Nucleic Acids

[00221] In certain embodiments, nucleic acid sequences can exist in a variety of instances such as: isolated segments and recombinant vectors of incorporated sequences or recombinant polynucleotides encoding one or both chains of an antibody, or a fragment, derivative, mutein, or variant thereof, polynucleotides sufficient for use as hybridization probes, PCR primers or sequencing primers for identifying, analyzing, mutating or amplifying a polynucleotide encoding a polypeptide, anti-sense nucleic acids for inhibiting expression of a polynucleotide, and complementary sequences of the foregoing described herein. Nucleic acids that encode the epitope to which certain of the antibodies provided herein are also provided. Nucleic acids encoding fusion proteins that include these peptides are also provided. The nucleic acids can be single-stranded or double-stranded and can comprise RNA and/or DNA nucleotides and artificial variants thereof (e.g., peptide nucleic acids).

[00222] The term “polynucleotide” refers to a nucleic acid molecule that either is recombinant or has been isolated from total genomic nucleic acid. Included within the term “polynucleotide” are oligonucleotides (nucleic acids 100 residues or less in length), recombinant vectors, including, for example, plasmids, cosmids, phage, viruses, and the like. Polynucleotides include, in certain aspects, regulatory sequences, isolated substantially away from their naturally

occurring genes or protein encoding sequences. Polynucleotides may be single- stranded (coding or antisense) or double- stranded, and may be RNA, DNA (genomic, cDNA or synthetic), analogs thereof, or a combination thereof. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide.

[00223] In this respect, the term “gene,” “polynucleotide,” or “nucleic acid” is used to refer to a nucleic acid that encodes a protein, polypeptide, or peptide (including any sequences required for proper transcription, post-translational modification, or localization). As will be understood by those in the art, this term encompasses genomic sequences, expression cassettes, cDNA sequences, and smaller engineered nucleic acid segments that express, or may be adapted to express, proteins, polypeptides, domains, peptides, fusion proteins, and mutants. A nucleic acid encoding all or part of a polypeptide may contain a contiguous nucleic acid sequence encoding all or a portion of such a polypeptide. It also is contemplated that a particular polypeptide may be encoded by nucleic acids containing variations having slightly different nucleic acid sequences but, nonetheless, encode the same or substantially similar protein.

[00224] In certain embodiments, there are polynucleotide variants having substantial sequence identity to the sequences disclosed herein; those comprising at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% or higher sequence identity, including all values and ranges there between, compared to a polynucleotide sequence provided herein using the methods described herein (e.g., BLAST analysis using standard parameters). In certain aspects, the isolated polynucleotide will comprise a nucleotide sequence encoding a polypeptide that has at least 90%, preferably 95% and above, identity to an amino acid sequence described herein, over the entire length of the sequence; or a nucleotide sequence complementary to said isolated polynucleotide.

[00225] The nucleic acid segments, regardless of the length of the coding sequence itself, may be combined with other nucleic acid sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. The nucleic acids can be any length. They can be, for example, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 125, 175, 200, 250, 300, 350, 400, 450, 500, 750, 1000, 1500, 3000, 5000 or more nucleotides in length, and/or can comprise one or more additional sequences, for example, regulatory sequences, and/or be a part of a larger nucleic acid, for example, a vector. It is therefore contemplated that a nucleic acid

fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant nucleic acid protocol. In some cases, a nucleic acid sequence may encode a polypeptide sequence with additional heterologous coding sequences, for example to allow for purification of the polypeptide, transport, secretion, post-translational modification, or for therapeutic benefits such as targeting or efficacy. As discussed above, a tag or other heterologous polypeptide may be added to the modified polypeptide-encoding sequence, wherein “heterologous” refers to a polypeptide that is not the same as the modified polypeptide.

1. Hybridization

[00226] The nucleic acids that hybridize to other nucleic acids under particular hybridization conditions. Methods for hybridizing nucleic acids are well known in the art. See, e.g., Current Protocols in Molecular Biology, John Wiley and Sons, N.Y. (1989), 6.3.1-6.3.6. As defined herein, a moderately stringent hybridization condition uses a prewashing solution containing 5× sodium chloride/sodium citrate (SSC), 0.5% SDS, 1.0 mM EDTA (pH 8.0), hybridization buffer of about 50% formamide, 6×SSC, and a hybridization temperature of 55° C. (or other similar hybridization solutions, such as one containing about 50% formamide, with a hybridization temperature of 42° C), and washing conditions of 60° C. in 0.5×SSC, 0.1% SDS. A stringent hybridization condition hybridizes in 6×SSC at 45° C., followed by one or more washes in 0.1×SSC, 0.2% SDS at 68° C. Furthermore, one of skill in the art can manipulate the hybridization and/or washing conditions to increase or decrease the stringency of hybridization such that nucleic acids comprising nucleotide sequence that are at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% identical to each other typically remain hybridized to each other.

[00227] The parameters affecting the choice of hybridization conditions and guidance for devising suitable conditions are set forth by, for example, Sambrook, Fritsch, and Maniatis (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., chapters 9 and 11 (1989); Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley and Sons, Inc., sections 2.10 and 6.3-6.4 (1995), both of which are herein incorporated by reference in their entirety for all purposes) and can be readily determined by those having ordinary skill in the art based on, for example, the length and/or base composition of the DNA.

2. Mutation

[00228] Changes can be introduced by mutation into a nucleic acid, thereby leading to changes in the amino acid sequence of a polypeptide (e.g., an antibody or antibody derivative) that it encodes. Mutations can be introduced using any technique known in the art. In one embodiment, one or more particular amino acid residues are changed using, for example, a site-directed mutagenesis protocol. In another embodiment, one or more randomly selected residues are changed using, for example, a random mutagenesis protocol. However it is made, a mutant polypeptide can be expressed and screened for a desired property.

[00229] Mutations can be introduced into a nucleic acid without significantly altering the biological activity of a polypeptide that it encodes. For example, one can make nucleotide substitutions leading to amino acid substitutions at non-essential amino acid residues. Alternatively, one or more mutations can be introduced into a nucleic acid that selectively changes the biological activity of a polypeptide that it encodes. See, eg., Romain Studer et al., *Biochem. J.* 449:581-594 (2013). For example, the mutation can quantitatively or qualitatively change the biological activity. Examples of quantitative changes include increasing, reducing or eliminating the activity. Examples of qualitative changes include altering the antigen specificity of an antibody.

3. Probes

[00230] In another aspect, nucleic acid molecules are suitable for use as primers or hybridization probes for the detection of nucleic acid sequences. A nucleic acid molecule can comprise only a portion of a nucleic acid sequence encoding a full-length polypeptide, for example, a fragment that can be used as a probe or primer or a fragment encoding an active portion of a given polypeptide.

[00231] In another embodiment, the nucleic acid molecules may be used as probes or PCR primers for specific antibody sequences. For instance, a nucleic acid molecule probe may be used in diagnostic methods or a nucleic acid molecule PCR primer may be used to amplify regions of DNA that could be used, inter alia, to isolate nucleic acid sequences for use in producing variable domains of antibodies. See, eg., Gaily Kivi et al., *BMC Biotechnol.* 16:2 (2016). In a preferred embodiment, the nucleic acid molecules are oligonucleotides. In a more preferred embodiment, the oligonucleotides are from highly variable regions of the heavy and

light chains of the antibody of interest. In an even more preferred embodiment, the oligonucleotides encode all or part of one or more of the CDRs.

[00232] Probes based on the desired sequence of a nucleic acid can be used to detect the nucleic acid or similar nucleic acids, for example, transcripts encoding a polypeptide of interest. The probe can comprise a label group, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used to identify a cell that expresses the polypeptide.

Antibody Production

[00233] Methods for preparing and characterizing antibodies for use in diagnostic and detection assays, for purification, and for use as therapeutics are well known in the art as disclosed in, for example, U.S. Pat. Nos. 4,011,308; 4,722,890; 4,016,043; 3,876,504; 3,770,380; and 4,372,745 (see, e.g., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988; incorporated herein by reference). These antibodies may be polyclonal or monoclonal antibody preparations, monospecific antisera, human antibodies, hybrid or chimeric antibodies, such as humanized antibodies, altered antibodies, F(ab')₂ fragments, Fab fragments, Fv fragments, single-domain antibodies, dimeric or trimeric antibody fragment constructs, minibodies, or functional fragments thereof which bind to the antigen in question. In certain aspects, polypeptides, peptides, and proteins and immunogenic fragments thereof for use in various embodiments can also be synthesized in solution or on a solid support in accordance with conventional techniques. See, for example, Stewart and Young, (1984); Tarn et al, (1983); Merrifield, (1986); and Barany and Merrifield (1979), each incorporated herein by reference.

[00234] Briefly, a polyclonal antibody is prepared by immunizing an animal with an antigen or a portion thereof and collecting antisera from that immunized animal. The antigen may be altered compared to an antigen sequence found in nature. In some embodiments, a variant or altered antigenic peptide or polypeptide is employed to generate antibodies. Inocula are typically prepared by dispersing the antigenic composition in a physiologically tolerable diluent to form an aqueous composition. Antisera is subsequently collected by methods known in the arts, and the serum may be used as-is for various applications or else the desired antibody fraction may be purified by well known methods, such as affinity chromatography (Harlow and Lane, *Antibodies: A Laboratory Manual* 1988).

[00235] Methods of making monoclonal antibodies are also well known in the art (Kohler and Milstein, 1975; Harlow and Lane, 1988, U.S. Patent 4,196,265, herein incorporated by reference in its entirety for all purposes). Typically, this technique involves immunizing a suitable animal with a selected immunogenic composition, e.g., a purified or partially purified protein, polypeptide, peptide or domain. Resulting antibody-producing B-cells from the immunized animal, or all dissociated splenocytes, are then induced to fuse with cells from an immortalized cell line to form hybridomas. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing and have high fusion efficiency and enzyme deficiencies that render them incapable of growing in certain selective media that support the growth of only the desired fused cells (hybridomas). Typically, the fusion partner includes a property that allows selection of the resulting hybridomas using specific media. For example, fusion partners can be hypoxanthine/aminopterin/thymidine (HAT)-sensitive. Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Next, selection of hybridomas can be performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. Fusion procedures for making hybridomas, immunization protocols, and techniques for isolation of immunized splenocytes for fusion are known in the art.

[00236] Other techniques for producing monoclonal antibodies include the viral or oncogenic transformation of B-lymphocytes, a molecular cloning approach may be used to generate a nucleic acid or polypeptide, the selected lymphocyte antibody method (SLAM) (see, e.g., Babcook et al., Proc. Natl. Acad. Sci. USA 93:7843-7848 (1996), the preparation of combinatorial immunoglobulin phagemid libraries from RNA isolated from the spleen of the immunized animal and selection of phagemids expressing appropriate antibodies, or producing a cell expressing an antibody from a genomic sequence of the cell comprising a modified immunoglobulin locus using Cre-mediated site-specific recombination (see, e.g., U.S. 6,091,001).

[00237] Monoclonal antibodies may be further purified using filtration, centrifugation, and various chromatographic methods such as HPLC or affinity chromatography. Monoclonal antibodies may be further screened or optimized for properties relating to specificity, avidity,

half-life, immunogenicity, binding association, binding disassociation, or overall functional properties relative to being a treatment for infection. Thus, monoclonal antibodies may have alterations in the amino acid sequence of CDRs, including insertions, deletions, or substitutions with a conserved or non-conserved amino acid.

[00238] The immunogenicity of a particular immunogen composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Adjuvants that may be used in accordance with embodiments include, but are not limited to, IL-1, IL-2, IL-4, IL-7, IL 12, -interferon, GMCSP, BCG, aluminum hydroxide, MDP compounds, such as thur-MDP and nor MDP, CGP (MTP-PE), lipid A, and monophosphoryl lipid A (MPL). Exemplary adjuvants may include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis*), incomplete Freund's adjuvants, and/or aluminum hydroxide adjuvant. In addition to adjuvants, it may be desirable to co-administer biologic response modifiers (BRM), such as but not limited to, Cimetidine (CIM; 1200 mg/d) (Smith/Kline, PA); low-dose Cyclophosphamide (CYP; 300 mg/m²) (Johnson/ Mead, NJ), cytokines such as β interferon, IL-2, or IL-12, or genes encoding proteins involved in immune helper functions, such as B-7.A phage-display system can be used to expand antibody molecule populations in vitro. Saiki, et al., *Nature* 324:163 (1986); Scharf et al., *Science* 233:1076 (1986); U.S. Pat. Nos. 4,683,195 and 4,683,202; Yang et al., *J Mol Biol.* 254:392 (1995); Barbas, III et al., *Methods: Comp. Meth Enzymol.* (1995) 8:94; Barbas, III et al., *Proc Natl Acad Sci USA* 88:7978 (1991).

Fully Human Antibody Production

[00239] Methods are available for making fully human antibodies. Using fully human antibodies can minimize the immunogenic and allergic responses that may be caused by administering non-human monoclonal antibodies to humans as therapeutic agents. In one embodiment, human antibodies may be produced in a non-human transgenic animal, e.g., a transgenic mouse capable of producing multiple isotypes of human antibodies to protein (e.g., IgG, IgA, and/or IgE) by undergoing V-D-J recombination and isotype switching. Accordingly, this aspect applies to antibodies, antibody fragments, and pharmaceutical compositions thereof, but also non-human transgenic animals, B-cells, host cells, and hybridomas that produce monoclonal antibodies. Applications of humanized antibodies include, but are not limited to, detect a cell expressing an anticipated protein, either in vivo or in vitro, pharmaceutical

preparations containing the antibodies of the present invention, and methods of treating disorders by administering the antibodies.

[00240] Fully human antibodies can be produced by immunizing transgenic animals (usually mice) that are capable of producing a repertoire of human antibodies in the absence of endogenous immunoglobulin production. Antigens for this purpose typically have six or more contiguous amino acids, and optionally are conjugated to a carrier, such as a hapten. See, for example, Jakobovits et al., *Proc. Natl. Acad. Sci. USA* 90:2551-2555 (1993); Jakobovits et al., *Nature* 362:255-258 (1993); Bruggermann et al., *Year in Immunol.* 7:33 (1993). In one example, transgenic animals are produced by incapacitating the endogenous mouse immunoglobulin loci encoding the mouse heavy and light immunoglobulin chains therein, and inserting into the mouse genome large fragments of human genome DNA containing loci that encode human heavy and light chain proteins. Partially modified animals, which have less than the full complement of human immunoglobulin loci, are then crossbred to obtain an animal having all of the desired immune system modifications. When administered an immunogen, these transgenic animals produce antibodies that are immunospecific for the immunogen but have human rather than murine amino acid sequences, including the variable regions. For further details of such methods, see, for example, International Patent Application Publication Nos. WO 96/33735 and WO 94/02602, which are hereby incorporated by reference in their entirety. Additional methods relating to transgenic mice for making human antibodies are described in U.S. Pat. Nos. 5,545,807; 6,713,610; 6,673,986; 6,162,963; 6,300,129; 6,255,458; 5,877,397; 5,874,299 and 5,545,806; in International Patent Application Publication Nos. WO 91/10741 and WO 90/04036; and in European Patent Nos. EP 546073B1 and EP 546073A1, all of which are hereby incorporated by reference in their entirety for all purposes.

[00241] The transgenic mice described above, referred to herein as “HuMAb” mice, contain a human immunoglobulin gene minilocus that encodes unrearranged human heavy (μ and γ) and κ light chain immunoglobulin sequences, together with targeted mutations that inactivate the endogenous μ and κ chain loci (Lonberg et al., *Nature* 368:856-859 (1994)). Accordingly, the mice exhibit reduced expression of mouse IgM or κ chains and in response to immunization, the introduced human heavy and light chain transgenes undergo class switching and somatic mutation to generate high affinity human IgG κ monoclonal antibodies (Lonberg et al., *supra*; Lonberg and Huszar, *Intern. Ref. Immunol.* 13:65-93 (1995); Harding and Lonberg,

Ann. N.Y. Acad. Sci. 764:536-546 (1995)). The preparation of HuMAb mice is described in detail in Taylor et al., Nucl. Acids Res. 20:6287-6295 (1992); Chen et al., Int. Immunol. 5:647-656 (1993); Tuailon et al., J. Immunol. 152:2912-2920 (1994); Lonberg et al., supra; Lonberg, Handbook of Exp. Pharmacol. 113:49-101 (1994); Taylor et al., Int. Immunol. 6:579-591 (1994); Lonberg and Huszar, Intern. Ref. Immunol. 13:65-93 (1995); Harding and Lonberg, Ann. N.Y. Acad. Sci. 764:536-546 (1995); Fishwild et al., Nat. Biotechnol. 14:845-851 (1996); the foregoing references are herein incorporated by reference in their entirety for all purposes. See further, U.S. Pat. Nos. 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,789,650; 5,877,397; 5,661,016; 5,814,318; 5,874,299; 5,770,429; and 5,545,807; as well as International Patent Application Publication Nos. WO 93/1227; WO 92/22646; and WO 92/03918, the disclosures of all of which are hereby incorporated by reference in their entirety for all purposes. Technologies utilized for producing human antibodies in these transgenic mice are disclosed also in WO 98/24893, and Mendez et al., Nat. Genetics 15:146-156 (1997), which are herein incorporated by reference. For example, the HCo7 and HCo12 transgenic mice strains can be used to generate human antibodies.

[00242] Using hybridoma technology, antigen-specific humanized monoclonal antibodies with the desired specificity can be produced and selected from the transgenic mice such as those described above. Such antibodies may be cloned and expressed using a suitable vector and host cell, or the antibodies can be harvested from cultured hybridoma cells. Fully human antibodies can also be derived from phage-display libraries (as disclosed in Hoogenboom et al., J. Mol. Biol. 227:381 (1991); and Marks et al., J. Mol. Biol. 222:581 (1991)). One such technique is described in International Patent Application Publication No. WO 99/10494 (herein incorporated by reference), which describes the isolation of high affinity and functional agonistic antibodies for MPL- and msk-receptors using such an approach.

2. Antibody Fragment Production

[00243] Antibody fragments that retain the ability to recognize the antigen of interest will also find use herein. A number of antibody fragments are known in the art that comprise antigen-binding sites capable of exhibiting immunological binding properties of an intact antibody molecule and can be subsequently modified by methods known in the arts. Functional fragments, including only the variable regions of the heavy and light chains, can also be produced using standard techniques such as recombinant production or preferential proteolytic

cleavage of immunoglobulin molecules. These fragments are known as Fv. See, e.g., Inbar et al., *Proc. Nat. Acad. Sci. USA* 69:2659-2662 (1972); Hochman et al., *Biochem.* 15:2706-2710 (1976); and Ehrlich et al., *Biochem.* 19:4091-4096 (1980).

[00244] Single-chain variable fragments (scFvs) may be prepared by fusing DNA encoding a peptide linker between DNAs encoding the two variable domain polypeptides (VL and VH). scFvs can form antigen-binding monomers, or they can form multimers (e.g., dimers, trimers, or tetramers), depending on the length of a flexible linker between the two variable domains (Kortt et al., *Prot. Eng.* 10:423 (1997); Kort et al., *Biomol. Eng.* 18:95-108 (2001)). By combining different VL- and VH-comprising polypeptides, one can form multimeric scFvs that bind to different epitopes (Kriangkum et al., *Biomol. Eng.* 18:31-40 (2001)). Antigen-binding fragments are typically produced by recombinant DNA methods known to those skilled in the art. Although the two domains of the Fv fragment, VL and VH, are coded for by separate genes, they can be joined using recombinant methods by a synthetic linker that enables them to be made as a single chain polypeptide (known as single chain Fv (sFv or scFv); see e.g., Bird et al., *Science* 242:423-426 (1988); and Huston et al., *Proc. Natl. Acad. Sci. USA* 85:5879-5883 (1988). Design criteria include determining the appropriate length to span the distance between the C-terminus of one chain and the N-terminus of the other, wherein the linker is generally formed from small hydrophilic amino acid residues that do not tend to coil or form secondary structures. Suitable linkers generally comprise polypeptide chains of alternating sets of glycine and serine residues, and may include glutamic acid and lysine residues inserted to enhance solubility. Antigen-binding fragments are screened for utility in the same manner as intact antibodies. Such fragments include those obtained by amino-terminal and/or carboxy-terminal deletions, where the remaining amino acid sequence is substantially identical to the corresponding positions in the naturally occurring sequence deduced, for example, from a full-length cDNA sequence.

[00245] Antibodies may also be generated using peptide analogs of the epitopic determinants disclosed herein, which may consist of non-peptide compounds having properties analogous to those of the template peptide. These types of non-peptide compound are termed “peptide mimetics” or “peptidomimetics”. Fauchere, J. *Adv. Drug Res.* 15:29 (1986); Veber and Freidinger *TINS* p. 392 (1985); and Evans et al., *J. Med. Chem.* 30:1229 (1987). Liu et al. (2003) also describe “antibody like binding peptidomimetics” (ABiPs), which are peptides that act as

pared-down antibodies and have certain advantages of longer serum half-life as well as less cumbersome synthesis methods. These analogs can be peptides, non-peptides or combinations of peptide and non-peptide regions. Fauchere, Adv. Drug Res. 15:29 (1986); Veber and Freidiner, TINS p. 392 (1985); and Evans et al., J. Med. Chem. 30:1229 (1987), which are incorporated herein by reference in their entirety for any purpose. Peptide mimetics that are structurally similar to therapeutically useful peptides may be used to produce a similar therapeutic or prophylactic effect. Such compounds are often developed with the aid of computerized molecular modeling. Generally, peptidomimetics of the invention are proteins that are structurally similar to an antibody displaying a desired biological activity, such as the ability to bind a protein, but have one or more peptide linkages optionally replaced by a linkage selected from: $\text{—CH}_2\text{NH—}$, $\text{—CH}_2\text{S—}$, $\text{—CH}_2\text{—CH}_2\text{—}$, —CH=CH— (cis and trans), $\text{—COCH}_2\text{—}$, $\text{—CH(OH)CH}_2\text{—}$, and $\text{—CH}_2\text{SO—}$ by methods well known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (e.g., D-lysine in place of L-lysine) may be used in certain embodiments of the invention to generate more stable proteins. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo and Gierasch, Ann. Rev. Biochem. 61:387 (1992), incorporated herein by reference), for example, by adding internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

[00246] Once generated, a phage display library can be used to improve the immunological binding affinity of the Fab molecules using known techniques. See, e.g., Figini et al., J. Mol. Biol. 239:68 (1994). The coding sequences for the heavy and light chain portions of the Fab molecules selected from the phage display library can be isolated or synthesized and cloned into any suitable vector or replicon for expression. Any suitable expression system can be used.

Polypeptide Expression

[00247] In some aspects, there are nucleic acid molecule encoding polypeptides or peptides of the disclosure (e.g antibodies, TCR genes, and immunogenic peptides). These may be generated by methods known in the art, e.g., isolated from B cells of mice that have been immunized and isolated, phage display, expressed in any suitable recombinant expression system and allowed to assemble to form antibody molecules or by recombinant methods.

1. Expression

[00248] The nucleic acid molecules may be used to express large quantities of polypeptides. If the nucleic acid molecules are derived from a non-human, non-transgenic animal, the nucleic acid molecules may be used for humanization of the antibody or TCR genes.

2. Vectors

[00249] In some aspects, contemplated are expression vectors comprising a nucleic acid molecule encoding a polypeptide of the desired sequence or a portion thereof (e.g., a fragment containing one or more CDRs or one or more variable region domains). Expression vectors comprising the nucleic acid molecules may encode the heavy chain, light chain, or the antigen-binding portion thereof. In some aspects, expression vectors comprising nucleic acid molecules may encode fusion proteins, modified antibodies, antibody fragments, and probes thereof. In addition to control sequences that govern transcription and translation, vectors and expression vectors may contain nucleic acid sequences that serve other functions as well.

[00250] To express the polypeptides or peptides of the disclosure, DNAs encoding the polypeptides or peptides are inserted into expression vectors such that the gene area is operatively linked to transcriptional and translational control sequences. In some aspects, a vector that encodes a functionally complete human CH or CL immunoglobulin sequence with appropriate restriction sites engineered so that any VH or VL sequence can be easily inserted and expressed. In some aspects, a vector that encodes a functionally complete human TCR alpha or TCR beta sequence with appropriate restriction sites engineered so that any variable sequence or CDR1, CDR2, and/or CDR3 can be easily inserted and expressed. Typically, expression vectors used in any of the host cells contain sequences for plasmid or virus maintenance and for cloning and expression of exogenous nucleotide sequences. Such sequences, collectively referred to as “flanking sequences” typically include one or more of the following operatively linked nucleotide sequences: a promoter, one or more enhancer sequences, an origin of replication, a transcriptional termination sequence, a complete intron sequence containing a donor and acceptor splice site, a sequence encoding a leader sequence for polypeptide secretion, a ribosome binding site, a polyadenylation sequence, a polylinker region for inserting the nucleic acid encoding the polypeptide to be expressed, and a selectable marker element. Such sequences and methods of using the same are well known in the art.

3. Expression Systems

[00251] Numerous expression systems exist that comprise at least a part or all of the expression vectors discussed above. Prokaryote- and/or eukaryote-based systems can be employed for use with an embodiment to produce nucleic acid sequences, or their cognate polypeptides, proteins and peptides. Commercially and widely available systems include in but are not limited to bacterial, mammalian, yeast, and insect cell systems. Different host cells have characteristic and specific mechanisms for the post-translational processing and modification of proteins. Appropriate cell lines or host systems can be chosen to ensure the correct modification and processing of the foreign protein expressed. Those skilled in the art are able to express a vector to produce a nucleic acid sequence or its cognate polypeptide, protein, or peptide using an appropriate expression system.

4. Methods of Gene Transfer

[00252] Suitable methods for nucleic acid delivery to effect expression of compositions are anticipated to include virtually any method by which a nucleic acid (e.g., DNA, including viral and nonviral vectors) can be introduced into a cell, a tissue or an organism, as described herein or as would be known to one of ordinary skill in the art. Such methods include, but are not limited to, direct delivery of DNA such as by injection (U.S. Patents 5,994,624, 5,981,274, 5,945,100, 5,780,448, 5,736,524, 5,702,932, 5,656,610, 5,589,466 and 5,580,859, each incorporated herein by reference), including microinjection (Harland and Weintraub, 1985; U.S. Patent 5,789,215, incorporated herein by reference); by electroporation (U.S. Patent No. 5,384,253, incorporated herein by reference); by calcium phosphate precipitation (Graham and Van Der Eb, 1973; Chen and Okayama, 1987; Rippe et al., 1990); by using DEAE dextran followed by polyethylene glycol (Gopal, 1985); by direct sonic loading (Fechheimer et al., 1987); by liposome mediated transfection (Nicolau and Sene, 1982; Fraley et al., 1979; Nicolau et al., 1987; Wong et al., 1980; Kaneda et al., 1989; Kato et al., 1991); by microprojectile bombardment (PCT Application Nos. WO 94/09699 and 95/06128; U.S. Patents 5,610,042; 5,322,783, 5,563,055, 5,550,318, 5,538,877 and 5,538,880, and each incorporated herein by reference); by agitation with silicon carbide fibers (Kaeppler et al., 1990; U.S. Patents 5,302,523 and 5,464,765, each incorporated herein by reference); by Agrobacterium mediated transformation (U.S. Patents 5,591,616 and 5,563,055, each incorporated herein by reference); or by PEG mediated transformation of protoplasts (Omirulleh et al., 1993; U.S. Patents 4,684,611 and 4,952,500, each incorporated herein by reference); by desiccation/inhibition mediated DNA

uptake (Potrykus et al., 1985). Other methods include viral transduction, such as gene transfer by lentiviral or retroviral transduction.

5. Host Cells

[00253] In another aspect, contemplated are the use of host cells into which a recombinant expression vector has been introduced. Antibodies can be expressed in a variety of cell types. An expression construct encoding an antibody can be transfected into cells according to a variety of methods known in the art. Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. Some vectors may employ control sequences that allow it to be replicated and/or expressed in both prokaryotic and eukaryotic cells. In certain aspects, the antibody expression construct can be placed under control of a promoter that is linked to T-cell activation, such as one that is controlled by NFAT-1 or NF- κ B, both of which are transcription factors that can be activated upon T-cell activation. Control of antibody expression allows T cells, such as tumor- targeting T cells, to sense their surroundings and perform real-time modulation of cytokine signaling, both in the T cells themselves and in surrounding endogenous immune cells. One of skill in the art would understand the conditions under which to incubate host cells to maintain them and to permit replication of a vector. Also understood and known are techniques and conditions that would allow large-scale production of vectors, as well as production of the nucleic acids encoded by vectors and their cognate polypeptides, proteins, or peptides.

[00254] For stable transfection of mammalian cells, it is known, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a selectable marker (e.g., for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die), among other methods known in the arts.

6. Isolation

[00255] The nucleic acid molecule encoding either or both of the entire heavy and light chains of an antibody or the variable regions thereof may be obtained from any source that produces antibodies. Methods of isolating mRNA encoding an antibody are well known in the art. See e.g., Sambrook et al., *supra*. The sequences of human heavy and light chain constant

region genes are also known in the art. See, e.g., Kabat et al., 1991, supra. Nucleic acid molecules encoding the full-length heavy and/or light chains may then be expressed in a cell into which they have been introduced and the antibody isolated.

Cancer-Related Diseases and Disorders

[00256] In some embodiments, the disease or disorder is a type of cancer. In some embodiments, the type of cancer is melanoma (skin cancer), breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS).

[00257] As related to treating a type of cancer (e.g., melanoma, ovarian, head and neck, breast, skin, lung, adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, uterine cancer, and combinations thereof, and the like), treating can include but is not limited to prophylactic treatment and therapeutic treatment. As such, treatment can include, but is not limited to: preventing cancer (e.g., melanoma, ovarian, head and neck, breast, lung adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, uterine cancer, and combinations thereof, and the like); reducing the risk of cancer (e.g., melanoma, ovarian, head and neck, breast, lung adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, uterine cancer, and combinations thereof, and the like); ameliorating or relieving symptoms of cancer (e.g., melanoma, ovarian, head and neck, breast, lung adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, uterine cancer, and combinations thereof, and the like); eliciting a bodily response against cancer (e.g., melanoma, ovarian, head and neck, breast, lung adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, uterine cancer, and combinations thereof, and the like); inhibiting the development or progression of cancer (e.g., melanoma, ovarian, head and neck, breast, lung adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder,

testicular, uterine cancer, and combinations thereof, and the like) in animals that have intrinsic or acquired resistance to other cancer treatments; or preventing relapse of melanoma, ovarian, head and neck, breast, lung adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, uterine cancer, and the like, in animals that have intrinsic or acquired resistance to other cancer treatments. In some embodiments, treating does not include prophylactic treatment of cancer (e.g., preventing or ameliorating future cancer).

[00258] Treatment of a subject can occur using any suitable administration method (such as those disclosed herein) and using any suitable amount of a treatment. In some embodiments, methods of treatment comprise treating an animal for a type of cancer (e.g., melanoma, ovarian, head and neck, breast, skin, lung, adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, and/or uterine cancer), and combinations thereof, and the like). Other embodiments include treatment after one or more of an occurrence of chemical exposure, an exposure to ionizing radiation, or a treatment for a hematopoietic cancer (e.g., with chemotherapy, ionizing radiation, or both). Some embodiments of the disclosure include a method for treating a subject (e.g., an animal such as a human or primate) with a treatment which comprises one or more administrations of one or more such compositions; the compositions may be the same or different if there is more than one administration.

Methods of Treatment

[00259] Some embodiments of the disclosure further comprise administering a treatment to a subject. In some embodiments, the treatment is a type of immunotherapy. Some embodiments of the disclosure further comprise administering to the subject an immunotherapy in combination with one or more additional therapies.

[00260] In some embodiments, the immunotherapy and the one or more additional therapies are administered together in one administration or composition. In an embodiment, the immunotherapy and the one or more additional therapies are administered separately in more than one administration or more than one composition.

[00261] In some embodiments, methods of treatment include administering to a subject a therapeutically effective amount of an immunotherapy comprising a neoantigen target identified as described herein, such as, for example, the polypeptides and/or proteins listed in Tables 3 and 4. As used herein, the term “effective amount” refers to a dosage or a series of dosages sufficient to affect treatment (e.g., to treat a type of cancer, such as melanoma, ovarian, head and neck, breast, skin, lung, adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, uterine cancer, and the like). In some embodiments, an effective amount can encompass a therapeutically effective amount. In certain embodiments, an effective amount can vary depending on the subject and the particular treatment being affected. The exact amount that is required can, for example, vary from subject to subject, depending on the age and general condition of the subject, the particular adjuvant being used (if applicable), administration protocol, and the like. As such, the effective amount can, for example, vary based on the particular circumstances, and an appropriate effective amount can be determined in a particular case. An effective amount can, for example, include any dosage or composition amount determined to be clinically relevant, as would be appreciated by those skilled in the art.

[00262] In some embodiments, the treatments can also include one or more of surgical intervention, chemotherapy, radiation therapy, hormone therapies, immunotherapy, and adjuvant systematic therapies. Adjuvants may include but are not limited to chemotherapy (e.g., temozolomide), radiation therapy, antiangiogenic therapy (e.g., bevacizumab), and hormone therapies, such as administration of LHRH agonists; antiestrogens, such as tamoxifen; high-dose progestogens; aromatase inhibitors; and/or adrenalectomy. Chemotherapy can be used as a single-agent or as a combination with known or new therapies.

[00263] Adjuvant treatments include treatments by the mechanisms disclosed herein and of cancers as disclosed herein, including, but not limited to tumors. Corresponding primary therapies can include, but are not limited to, surgery, chemotherapy, or radiation therapy. In some instances, the adjuvant treatment can be a combination of chemokine receptor antagonists with traditional chemotoxic agents or with immunotherapy that increases the specificity of treatment to the cancer and potentially limits additional systemic side effects.

[00264] In some embodiments, the administration to a subject may decrease the incidence of one or more symptoms associated with a type of cancer. In some embodiments, the type of cancer comprises melanoma, ovarian, head and neck, breast, skin, lung, adult and pediatric acute myeloid leukemia, myelodysplastic syndrome, bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, and/or uterine cancer. In some embodiments, the administration may decrease cancer-associated illness severity or mortality, or combinations thereof in said subject, as compared to a subject not receiving said composition.

[00265] In some embodiments, the method may decrease a marker of viability of cancer cells in a subject. In one aspect, the method may decrease a marker of viability of cancer cells. The marker may be selected from survival over time, proliferation, growth, migration, formation of colonies, chromatic assembly, DNA binding, RNA metabolism, cell migration, cell adhesion, inflammation, or a combination thereof.

Immunotherapies

[00266] Certain embodiments of the disclosure relate to immunotherapies including one or more immunogenic target. In some embodiments, the immunogenic target is identified via the methods described herein, or a nucleotide coding therefor. In some embodiments, the immunogenic target is a shared splicing neoantigen or protein including a neoantigen, wherein the neoantigen is identified via the methods described herein, or a nucleotide coding therefor. In some embodiments, the immunogenic target includes a neoantigen target identified as described herein, such as, for example, the polypeptides and/or proteins listed in Tables 3 and 4, or a nucleotide coding therefor.

[00267] Cancer immunotherapy is the use of the immune system to treat cancer. Immunotherapies can be categorized as active, passive or hybrid (active and passive). These approaches exploit the fact that cancer cells often have molecules on their surface that can be detected by the immune system, known as tumor-associated antigens (TAAs); they are often proteins or other macromolecules (e.g. carbohydrates). Active immunotherapy directs the immune system to attack tumor cells by targeting TAAs. Passive immunotherapies enhance existing anti-tumor responses and include the use of monoclonal antibodies, lymphocytes and cytokines.

[00268] In some embodiments, immunotherapies include cell-based immunotherapies, such as those involving cells which effect an immune response (such as, for example, lymphocytes, macrophages, natural killer (NK) cells, dendritic cells, cytotoxic T lymphocytes (CTL), and the like), CAR-T therapies, antibodies and antibody derivatives (such as, for example, monoclonal antibodies, conjugated monoclonal antibodies, polyclonal antibodies, antibody fragments, radiolabeled antibodies, chemolabeled antibodies, etc.), immune checkpoint inhibitors, vaccines (such as, for example, cancer vaccines (e.g. tumor cell vaccines, antigen vaccines, dendritic cell vaccines, vector-based vaccines, etc.), e.g. oncopage, sipuleucel-T, and the like), immunomodulators (such as, for example, interleukins, cytokines, chemokines, etc.), topical immunotherapies (such as, for example, imiquimod, and the like), injection immunotherapies, adoptive cell transfer, oncolytic virus therapies (such as, for example, talimogene laherparepvec (T-VEC), and the like), immunosuppressive drugs, helminthic therapies, other non-specific immunotherapies, and the like. Immune checkpoint inhibitor immunotherapies are those that target one or more specific proteins or receptors, such as PD-1, PD-L1, CTLA-4, and the like. Immune checkpoint inhibitor immunotherapies include ipilimumab (Yervoy), nivolumab (Opdivo), pembrolizumab (Keytruda), and the like. Non-specific immunotherapies include cytokines, interleukins, interferons, and the like. In some embodiments, an immunotherapy assigned or administered to a subject can include an interleukin, and/or interferon (IFN), and/or one or more suitable antibody-based reagent, such as denileukin diftitox and/or administration of an antibody-based reagent selected from the group consisting of ado- trastuzumab emtansine, alemtuzumab, atezolizumab, bevacizumab, blinatumomab, brentuximab vedotin, cetuximab, catumaxomab, gemtuzumab, ibritumomab tiuxetan, ilipimumab, natalizumab, nimotuzumab, nivolumab, ofatumumab, panitumumab, pembrolizumab, rituximab, tositumomab, trastuzumab, vivatuxin, and the like. In some embodiments, an immunotherapy assigned or administered to a subject can include an indoleamine 2,3-dioxygenase (IDO) inhibitor, adoptive T-cell therapy, virotherapy (T-VEC), and/or any other immunotherapy whose efficacy extensively depends on anti-tumor immunity.

[00269] Those skilled in the art can determine appropriate immunotherapy options, including treatments that have been approved and those that in clinical trials or otherwise under development. Any relevant immunotherapy treatment strategies, alone or in combination with

one or more additional cancer therapy, can be utilized in the practice of the present disclosure. Additional detail regarding various types of immunotherapy is provided below.

1. Checkpoint Inhibitors and Combination Treatment

[00270] PD-1, PDL1, and PDL2 inhibitors

[00271] PD-1 can act in the tumor microenvironment where T cells encounter an infection or tumor. Activated T cells upregulate PD-1 and continue to express it in the peripheral tissues. Cytokines such as IFN-gamma induce the expression of PDL1 on epithelial cells and tumor cells. PDL2 is expressed on macrophages and dendritic cells. The main role of PD-1 is to limit the activity of effector T cells in the periphery and prevent excessive damage to the tissues during an immune response. Inhibitors of the disclosure may block one or more functions of PD-1 and/or PDL1 activity.

[00272] Alternative names for “PD-1” include CD279 and SLEB2. Alternative names for “PDL1” include B7-H1, B7-4, CD274, and B7-H. Alternative names for “PDL2” include B7-DC, Btde, and CD273. In some embodiments, PD-1, PDL1, and PDL2 are human PD-1, PDL1 and PDL2.

[00273] In some embodiments, the PD-1 inhibitor is a molecule that inhibits the binding of PD-1 to its ligand binding partners. In a specific aspect, the PD-1 ligand binding partners are PDL1 and/or PDL2. In another embodiment, a PDL1 inhibitor is a molecule that inhibits the binding of PDL1 to its binding partners. In a specific aspect, PDL1 binding partners are PD-1 and/or B7-1. In another embodiment, the PDL2 inhibitor is a molecule that inhibits the binding of PDL2 to its binding partners. In a specific aspect, a PDL2 binding partner is PD-1. The inhibitor may be an antibody, an antigen binding fragment thereof, an immunoadhesin, a fusion protein, or oligopeptide. Exemplary antibodies are described in U.S. Patent Nos. 8,735,553, 8,354,509, and 8,008,449, all incorporated herein by reference. Other PD-1 inhibitors for use in the methods and compositions provided herein are known in the art such as described in U.S. Patent Application Nos. US2014/0294898, US2014/022021, and US2011/0008369, all incorporated herein by reference.

[00274] In some embodiments, the PD-1 inhibitor is an anti-PD-1 antibody (e.g., a human antibody, a humanized antibody, or a chimeric antibody). In some embodiments, the anti-PD-1 antibody is selected from the group consisting of nivolumab, pembrolizumab, and pidilizumab. In some embodiments, the PD-1 inhibitor is an immunoadhesin (e.g., an

immunoadhesin comprising an extracellular or PD-1 binding portion of PDL1 or PDL2 fused to a constant region (e.g., an Fc region of an immunoglobulin sequence). In some embodiments, the PDL1 inhibitor comprises AMP- 224. Nivolumab, also known as MDX-1106-04, MDX-1106, ONO-4538, BMS-936558, and OPDIVO®, is an anti-PD-1 antibody described in WO2006/121168. Pembrolizumab, also known as MK-3475, Merck 3475, lambrolizumab, KEYTRUDA®, and SCH-900475, is an anti-PD-1 antibody described in WO2009/114335. Pidilizumab, also known as CT-011, hBAT, or hBAT-1, is an anti-PD-1 antibody described in WO2009/101611. AMP-224, also known as B7-DCIg, is a PDL2-Fc fusion soluble receptor described in WO2010/027827 and WO2011/066342. Additional PD-1 inhibitors include MEDI0680, also known as AMP-514, and REGN2810.

[00275] In some embodiments, the immune checkpoint inhibitor is a PDL1 inhibitor such as Durvalumab, also known as MEDI4736, atezolizumab, also known as MPDL3280A, avelumab, also known as MSB00010118C, MDX-1105, BMS-936559, or combinations thereof. In certain aspects, the immune checkpoint inhibitor is a PDL2 inhibitor such as rHIgM12B7.

[00276] In some embodiments, the inhibitor comprises the heavy and light chain CDRs or VRs of nivolumab, pembrolizumab, or pidilizumab. Accordingly, in one embodiment, the inhibitor comprises the CDR1, CDR2, and CDR3 domains of the VH region of nivolumab, pembrolizumab, or pidilizumab, and the CDR1, CDR2 and CDR3 domains of the VL region of nivolumab, pembrolizumab, or pidilizumab. In another embodiment, the antibody competes for binding with and/or binds to the same epitope on PD-1, PDL1, or PDL2 as the above- mentioned antibodies. In another embodiment, the antibody has at least about 70, 75, 80, 85, 90, 95, 97, or 99% (or any derivable range therein) variable region amino acid sequence identity with the above-mentioned antibodies.

[00277] CTLA-4, B7-1, and B7-2

[00278] Another immune checkpoint that can be targeted in the methods provided herein is the cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), also known as CD152. The complete cDNA sequence of human CTLA-4 has the Genbank accession number L15006. CTLA-4 is found on the surface of T cells and acts as an “off” switch when bound to B7-1 (CD80) or B7-2 (CD86) on the surface of antigen-presenting cells. CTLA4 is a member of the immunoglobulin superfamily that is expressed on the surface of Helper T cells and transmits an inhibitory signal to T cells. CTLA4 is similar to the T-cell co-stimulatory protein, CD28, and

both molecules bind to B7-1 and B7-2 on antigen-presenting cells. CTLA-4 transmits an inhibitory signal to T cells, whereas CD28 transmits a stimulatory signal. Intracellular CTLA-4 is also found in regulatory T cells and may be important to their function. T cell activation through the T cell receptor and CD28 leads to increased expression of CTLA-4, an inhibitory receptor for B7 molecules. Inhibitors of the disclosure may block one or more functions of CTLA-4, B7-1, and/or B7-2 activity. In some embodiments, the inhibitor blocks the CTLA-4 and B7-1 interaction. In some embodiments, the inhibitor blocks the CTLA-4 and B7-2 interaction.

[00279] In some embodiments, the immune checkpoint inhibitor is an anti-CTLA-4 antibody (e.g., a human antibody, a humanized antibody, or a chimeric antibody), an antigen binding fragment thereof, an immunoadhesin, a fusion protein, or oligopeptide.

[00280] Anti-human-CTLA-4 antibodies (or VH and/or VL domains derived therefrom) suitable for use in the present methods can be generated using methods well known in the art. Alternatively, art recognized anti-CTLA-4 antibodies can be used. For example, the anti-CTLA-4 antibodies disclosed in: US 8,119,129, WO 01/14424, WO 98/42752; WO 00/37504 (CP675,206, also known as tremelimumab; formerly ticilimumab), U.S. Patent No. 6,207,156; Hurwitz et al., 1998; can be used in the methods disclosed herein. The teachings of each of the aforementioned publications are hereby incorporated by reference. Antibodies that compete with any of these art-recognized antibodies for binding to CTLA-4 also can be used. For example, a humanized CTLA-4 antibody is described in International Patent Application No. WO2001/014424, WO2000/037504, and U.S. Patent No. 8,017,114; all incorporated herein by reference.

[00281] A further anti-CTLA-4 antibody useful as a checkpoint inhibitor in the methods and compositions of the disclosure is ipilimumab (also known as 10D1, MDX- 010, MDX- 101, and Yervoy®) or antigen binding fragments and variants thereof (see, e.g., WO01/14424).

[00282] In some embodiments, the inhibitor comprises the heavy and light chain CDRs or VRs of tremelimumab or ipilimumab. Accordingly, in one embodiment, the inhibitor comprises the CDR1, CDR2, and CDR3 domains of the VH region of tremelimumab or ipilimumab, and the CDR1, CDR2 and CDR3 domains of the VL region of tremelimumab or ipilimumab. In another embodiment, the antibody competes for binding with and/or binds to the

same epitope on PD-1, B7-1, or B7-2 as the above-mentioned antibodies. In another embodiment, the antibody has at least about 70, 75, 80, 85, 90, 95, 97, or 99% (or any derivable range therein) variable region amino acid sequence identity with the above-mentioned antibodies.

2. Inhibition of Co-Stimulatory Molecules

[00283] In some embodiments, the immunotherapy comprises an inhibitor of a co-stimulatory molecule. In some embodiments, the inhibitor comprises an inhibitor of B7-1 (CD80), B7-2 (CD86), CD28, ICOS, OX40 (TNFRSF4), 4-1BB (CD137; TNFRSF9), CD40L (CD40LG), GITR (TNFRSF18), and combinations thereof. Inhibitors include inhibitory antibodies, polypeptides, compounds, and nucleic acids.

3. Dendritic Cell Therapy

[00284] Dendritic cell therapy provokes anti-tumor responses by causing dendritic cells to present tumor antigens to lymphocytes, which activates them, priming them to kill other cells that present the antigen. Dendritic cells are antigen presenting cells (APCs) in the mammalian immune system. In cancer treatment they aid cancer antigen targeting. One example of cellular cancer therapy based on dendritic cells is sipuleucel-T.

[00285] One method of inducing dendritic cells to present tumor antigens is by vaccination with autologous tumor lysates or short peptides (small parts of protein that correspond to the protein antigens on cancer cells). These peptides are often given in combination with adjuvants (highly immunogenic substances) to increase the immune and anti-tumor responses. Other adjuvants include proteins or other chemicals that attract and/or activate dendritic cells, such as granulocyte macrophage colony-stimulating factor (GM-CSF).

[00286] Dendritic cells can also be activated in vivo by making tumor cells express GM-CSF. This can be achieved by either genetically engineering tumor cells to produce GM-CSF or by infecting tumor cells with an oncolytic virus that expresses GM-CSF.

[00287] Another strategy is to remove dendritic cells from the blood of a patient and activate them outside the body. The dendritic cells are activated in the presence of tumor antigens, which may be a single tumor-specific peptide/protein or a tumor cell lysate (a solution of broken down tumor cells). These cells (with optional adjuvants) are infused and provoke an immune response.

[00288] Dendritic cell therapies include the use of antibodies that bind to receptors on the surface of dendritic cells. Antigens can be added to the antibody and can induce the dendritic cells to mature and provide immunity to the tumor. Dendritic cell receptors such as TLR3, TLR7, TLR8 or CD40 have been used as antibody targets.

4. CAR-T Cell Therapy

[00289] Chimeric antigen receptors (CARs, also known as chimeric immunoreceptors, chimeric T cell receptors or artificial T cell receptors) are engineered receptors that combine a new specificity with an immune cell to target cancer cells. Typically, these receptors graft the specificity of a monoclonal antibody onto a T cell. The receptors are called chimeric because they are fused of parts from different sources. CAR-T cell therapy refers to a treatment that uses such transformed cells for cancer therapy.

[00290] The basic principle of CAR-T cell design involves recombinant receptors that combine antigen-binding and T-cell activating functions. The general premise of CAR-T cells is to artificially generate T-cells targeted to markers found on cancer cells. Scientists can remove T-cells from a person, genetically alter them, and put them back into the patient for them to attack the cancer cells. Once the T cell has been engineered to become a CAR-T cell, it acts as a “living drug”. CAR-T cells create a link between an extracellular ligand recognition domain to an intracellular signalling molecule which in turn activates T cells. The extracellular ligand recognition domain is usually a single-chain variable fragment (scFv). An important aspect of the safety of CAR-T cell therapy is how to ensure that only cancerous tumor cells are targeted, and not normal cells. The specificity of CAR-T cells is determined by the choice of molecule that is targeted.

[00291] Exemplary CAR-T therapies include Tisagenlecleucel (Kymriah) and Axicabtagene ciloleucel (Yescarta). In some embodiments, the CAR-T therapy targets CD19.

5. Cytokine Therapy

[00292] Cytokines are proteins produced by many types of cells present within a tumor. They can modulate immune responses. The tumor often employs them to allow it to grow and reduce the immune response. These immune-modulating effects allow them to be used as drugs to provoke an immune response. Two commonly used cytokines are interferons and interleukins.

[00293] Interferons are produced by the immune system. They are usually involved in anti-viral response, but also have use for cancer. They fall in three groups: type I (IFN α and IFN β), type II (IFN γ) and type III (IFN λ).

[00294] Interleukins have an array of immune system effects. IL-2 is an exemplary interleukin cytokine therapy.

6. Adoptive T-Cell Therapy

[00295] Adoptive T cell therapy is a form of passive immunization by the transfusion of T-cells (adoptive cell transfer). They are found in blood and tissue and usually activate when they find foreign pathogens. Specifically they activate when the T-cell's surface receptors encounter cells that display parts of foreign proteins on their surface antigens. These can be either infected cells, or antigen presenting cells (APCs). They are found in normal tissue and in tumor tissue, where they are known as tumor infiltrating lymphocytes (TILs). They are activated by the presence of APCs such as dendritic cells that present tumor antigens. Although these cells can attack the tumor, the environment within the tumor is highly immunosuppressive, preventing immune-mediated tumour death.[60]

[00296] Multiple ways of producing and obtaining tumour targeted T-cells have been developed. T-cells specific to a tumor antigen can be removed from a tumor sample (TILs) or filtered from blood. Subsequent activation and culturing is performed ex vivo, with the results reinfused. Activation can take place through gene therapy, or by exposing the T cells to tumor antigen.

Chemotherapies / Targeted Therapies / Alternative Therapies

[00297] Cancers are commonly treated with chemotherapy and/or targeted therapy and/or alternative therapy. Chemotherapies act by indiscriminately targeting rapidly dividing cells, including healthy cells as well as tumor cells, whereas targeted cancer therapies rather act by interfering with specific molecules, or molecular targets, which are involved in cancer growth and progression. Targeted therapy generally targets cancer cells exclusively, having minimal damage to normal cells. Chemotherapies and targeted therapies which are approved and/or in the clinical trial stage are known to those skilled in the art. Any such compound can be utilized in the practice of the present disclosure.

[00298] For example, approved chemotherapies include abitrexate (Methotrexate Injection), abraxane (Paclitaxel Injection), adcetris (Brentuximab Vedotin Injection), adriamycin

(Doxorubicin), adrucil Injection (5-FU (fluorouracil)), afinitor (Everolimus), afinitor Disperz (Everolimus), alimta (PEMETREXED), alkeran Injection (Melphalan Injection), alkeran Tablets (Melphalan), aredia (Pamidronate), arimidex (Anastrozole), aromasin (Exemestane), arranon (Nelarabine), arzerra (Ofatumumab Injection), avastin (Bevacizumab), beleodaq (Belinostat Injection), bexxar (Tositumomab), BiCNU (Carmustine), blenoxane (Bleomycin), blincyto (Blinatumoma b Injection), bosulif (Bosutinib), busulfex Injection (Busulfan Injection), campath (Alemtuzumab), camptosar (Irinotecan), caprelsa (Vandetanib), casodex (Bicalutamide), CeeNU (Lomustine), CeeNU Dose Pack (Lomustine), cerubidine (Daunorubicin), clolar (Clofarabine Injection), cometriq (Cabozantinib), cosmegen (Dactinomycin), cotellic (Cobimetinib), cyramza (Ramucirumab Injection), cytosarU (Cytarabine), cytoxan (Cytosan), cytoxan Injection (Cyclophosphamide Injection), dacogen (Decitabine), daunoXome (Daunorubicin Lipid Complex Injection), decadron (Dexamethasone), depoCyt (Cytarabine Lipid Complex Injection), dexamethasone Intensol (Dexamethasone), dexpak Taperpak (Dexamethasone), docefrez (Docetaxel), doxil (Doxorubicin Lipid Complex Injection), droxia (Hydroxyurea), DTIC (Decarbazine), eligard (Leuprolide), ellence (Ellence (epirubicin)), eloxatin (Eloxatin (oxaliplatin)), elspar (Asparaginase), emcynt (Estramustine), erbitux (Cetuximab), erivedge (Vismodegib), erwinaze (Asparaginase Erwinia chrysanthemi), ethyol (Amifostine), etopophos (Etoposide Injection), eulexin (Flutamide), fareston (Toremifene), farydak (Panobinostat), faslodex (Fulvestrant), femara (Letrozole), firmagon (Degarelix Injection), fludara (Fludarabine), folex (Methotrexate Injection), folotylin (Pralatrexate Injection), FUDR (FUDR (floxuridine)), gazyva (Obinutuzumab Injection), gemzar (Gemcitabine), gilotrif (Afatinib), gleeevec (Imatinib Mesylate), Gliadel Wafer (Carmustine wafer), Halaven (Eribulin Injection), Herceptin (Trastuzumab), Hexalen (Altretamine), Hycamtin (Topotecan), Hycamtin (Topotecan), Hydrea (Hydroxyurea), Ibrance (Palbociclib), Iclusig (Ponatinib), Idamycin PFS (Idarubicin), Ifex (Ifosfamide), Imbruvica (Ibrutinib), Inlyta (Axitinib), Intron A alfab (Interferon alfa-2a), Iressa (Gefitinib), Istodax (Romidepsin Injection), Ixemptra (Ixabepilone Injection), Jakafi (Ruxolitinib), Jevtana (Cabazitaxel Injection), Kadcyla (Ado-trastuzumab Emtansine), Keytruda (Pembrolizumab Injection), Kyprolis (Carfilzomib), Lanvima (Lenvatinib), Leukeran (Chlorambucil), Leukine (Sargramostim), Leustatin (Cladribine), Lonsurf (Trifluridine and Tipiracil), Lupron (Leuprolide), Lupron Depot (Leuprolide), Lupron DepotPED (Leuprolide), Lynparza (Olaparib), Lysodren (Mitotane), Marqibo Kit (Vincristine Lipid Complex Injection),

Matulane (Procarbazine), Megace (Megestrol), Mekinist (Trametinib; for more information, see Borthakur, G. et al., Blood, 2012, 120:677, which is incorporated by reference herein), Mesnex (Mesna), Mesnex (Mesna Injection), Metastron (Strontium-89 Chloride), Mexate (Methotrexate Injection), Mustargen (Mechlorethamine), Mutamycin (Mitomycin), Myleran (Busulfan), Mylotarg (Gemtuzumab Ozogamicin, for more information, see Norsworthy, K. J. et al., Oncologist, 2018, 23:1103-1108, which is incorporated herein by reference), Navelbine (Vinorelbine), Neosar Injection (Cyclophosphamide Injection), Neulasta (filgrastim), Neulasta (pegfilgrastim), Neupogen (filgrastim), Nexavar (Sorafenib), Nilandron (Nilandron (nilutamide)), Nipent (Pentostatin), Nolvadex (Tamoxifen), Novantrone (Mitoxantrone, for more information, see Fox, E. J., Neurology, 2004, 28(12 Suppl 6):S15-8, which is incorporated herein by reference), Odomzo (Sonidegib), Oncaspar (Pegaspargase), Oncovin (Vincristine), Ontak (Denileukin Diftitox), onxol (Paclitaxel Injection), opdivo (Nivolumab Injection), panretin (Alitretinoin), paraplatin (Carboplatin), perjeta (Pertuzumab Injection), platinol (Cisplatin), platinol (Cisplatin Injection), platinolAQ (Cisplatin), platinolAQ (Cisplatin Injection), pomalyst (Pomalidomide), prednisone Intensol (Prednisone), proleukin (Aldesleukin), purinethol (Mercaptopurine), reclast (Zoledronic acid), revlimid (Lenalidomide; for more information see Krönke, J. et al., Nature, 2015, 523:183-188, which is incorporated by reference herein), actimid (Pomalidomid), rheumatrex (Methotrexate), rituxan (Rituximab), roferonA alfaa (Interferon alfa-2a), rubex (Doxorubicin), sandostatin (Octreotide), sandostatin LAR Depot (Octreotide), soltamox (Tamoxifen), sprycel (Dasatinib; for more information, see Duong, V. H. et al., Leukemia Research, 2013, 37:300-304, which is incorporated herein by reference), sterapred (Prednisone), sterapred DS (Prednisone), stivarga (Regorafenib), supprelin LA (Histrelin Implant), sutent (Sunitinib), sylatron (Peginterferon Alfa-2b Injection (Sylatron)), sylvant (Siltuximab Injection), synribo (Omacetaxine Injection), tabloid (Thioguanine), taflinar (Dabrafenib), tarceva (Erlotinib), targretin Capsules (Bexarotene), tassigna (Decarbazine), taxol (Paclitaxel Injection), taxotere (Docetaxel), temodar (Temozolomide), temodar (Temozolomide Injection), tepadina (Thiotepa), thalomid (Thalidomide), theraCys BCG (BCG), thioplex (Thiotepa), TICE BCG (BCG), toposar (Etoposide Injection), torisel (Temsilolimus), treanda (Bendamustine hydrochloride), trelstar (Triptorelin Injection), trexall (Methotrexate), trisenox (Arsenic trioxide), tykerb (lapatinib), unituxin (Dinutuximab Injection), valstar (Valrubicin Intravesical), vantas (Histrelin Implant), vectibix (Panitumumab), velban (Vinblastine), velcade

(Bortezomib), vepesid (Etoposide), vepesid (Etoposide Injection), vesanoid (Tretinoin), vidaza (Azacitidine), vincasar PFS (Vincristine), vincrex (Vincristine), votrient (Pazopanib), vumon (Teniposide), wellcovorin IV (Leucovorin Injection), xalkori (Crizotinib), xeloda (Capecitabine), xtandi (Enzalutamide), yervoy (Ipilimumab Injection), yondelis (Trabectedin Injection), zaltrap (Ziv-aflibercept Injection), zanosar (Streptozocin), zelboraf (Vemurafenib), zevalin (Ibritumomab Tiuxetan), zoladex (Goserelin), zolinza (Vorinostat), zometa (Zoledronic acid), zortress (Everolimus), zydelig (Idelalisib), zykadia (Ceritinib), zytiga (Abiraterone), and the like, in addition to analogs and derivatives thereof. For example, approved targeted therapies include ado-trastuzumab emtansine (Kadcyla), afatinib (Gilotrif), aldesleukin (Proleukin), alectinib (Alecensa), alemtuzumab (Campath), axitinib (Inlyta), bosutinib (Bosulif), brentuximab vedotin (Adcetris), cabozantinib (Cabometyx [tablet], Cometriq [capsule]), canakinumab (Ilaris), carfilzomib (Kyprolis), ceritinib (Zykadia), cetuximab (Erbix), cobimetinib (Cotellic), crizotinib (Xalkori), dabrafenib (Tafinlar), daratumumab (Darzalex), dasatinib (Sprycel), denosumab (Xgeva), dinutuximab (Unituxin), elotuzumab (Empliciti), erlotinib (Tarceva, for more information, see Boehringer, S. et al., Blood, 2008, 111:2170-2180, which is incorporated by reference herein), everolimus (Afinitor), gefitinib (Iressa), ibritumomab tiuxetan (Zevalin), ibrutinib (Imbruvica), idelalisib (Zydelig), imatinib (Gleevec), ipilimumab (Yervoy), ixazomib (Ninlaro), lapatinib (Tykerb), lenvatinib (Lenvima), necitumumab (Portrazza), nilotinib (Tasigna), nivolumab (Opdivo), obinutuzumab (Gazyva), ofatumumab (Arzerra, HuMax-CD20), olaparib (Lynparza), osimertinib (Tagrisso), palbociclib (Ibrance), panitumumab (Vectibix), panobinostat (Farydak), pazopanib (Votrient), pembrolizumab (Keytruda), pertuzumab (Perjeta), ponatinib (Iclusig), ramucirumab (Cyramza), rapamycin, regorafenib (Stivarga), rituximab (Rituxan, Mabthera), romidepsin (Istodax), ruxolitinib (Jakafi), siltuximab (Sylvant), sipuleucel-T (Provenge), sirolimus, sonidegib (Odomzo), sorafenib (Nexavar), sunitinib, tamoxifen, temsirolimus (Torisel), tocilizumab (Actemra), tofacitinib (Xeljanz), tositumomab (Bexxar), trametinib (Mekinist), trastuzumab (Herceptin), vandetanib (Caprelsa), vemurafenib (Zelboraf), venetoclax (Venclexta), vismodegib (Erivedge), vorinostat (Zolinza), ziv-aflibercept (Zaltrap), and the like, in addition to analogs and derivatives thereof. In an embodiment, the approved chemotherapy is an anthracycline, such as Doxorubicin, Daunorubicin, Epirubicin, and/or Idarubicin.

[00299] Those skilled in the art can determine appropriate additional therapy options (e.g. chemotherapy, targeted therapy, alternative therapy, and the like), including treatments that have been approved and those that in clinical trials or otherwise under development. Some targeted therapies are also immunotherapies. Any relevant chemotherapy, target therapy, and alternative therapy treatment strategies can be utilized, alone or in combination with one or more additional cancer therapy, in the practice of the present disclosure. thereof.

Other Cancer Treatments

[00300] In addition to chemotherapies, targeted therapies, alternative therapies, and immunotherapies, cancer can additionally be treated by other strategies. These include surgery, radiation therapy, hormone therapy, stem cell transplant, precision medicine, and the like; such treatments and the compounds and compositions utilized therein are known to those skilled in the art. Any such treatment strategies can be utilized in the practice of the present disclosure.

[00301] In some embodiments, the additional therapy or prior therapy comprises radiation, such as ionizing radiation. As used herein, “ionizing radiation” means radiation comprising particles or photons that have sufficient energy or can produce sufficient energy via nuclear interactions to produce ionization (gain or loss of electrons). An exemplary and preferred ionizing radiation is an x-radiation. Means for delivering x-radiation to a target tissue or cell are well known in the art.

[00302] In some embodiments, the additional therapy comprises surgery. Approximately 60% of persons with cancer will undergo surgery of some type, which includes preventative, diagnostic or staging, curative, and palliative surgery. Curative surgery includes resection in which all or part of cancerous tissue is physically removed, excised, and/or destroyed and may be used in conjunction with other therapies, such as the treatment of the present embodiments, chemotherapy, radiotherapy, hormonal therapy, gene therapy, immunotherapy, and/or alternative therapies. Tumor resection refers to physical removal of at least part of a tumor. In addition to tumor resection, treatment by surgery includes laser surgery, cryosurgery, electrosurgery, and microscopically-controlled surgery (Mohs’ surgery).

[00303] Upon excision of part or all of cancerous cells, tissue, or tumor, a cavity may be formed in the body. Treatment may be accomplished by perfusion, direct injection, or local application of the area with an additional anti-cancer therapy. Such treatment may be repeated, for example, every 1, 2, 3, 4, 5, 6, or 7 days, or every 1, 2, 3, 4, and 5 weeks or every 1, 2, 3, 4, 5,

6, 7, 8, 9, 10, 11, or 12 months (or any range derivable therein). These treatments may be of varying dosages as well.

[00304] Alternative treatment strategies have also been used with various types of cancers. Such treatment can be used alone or in combination with any other treatment modality. These include exercise, massage, relaxation techniques, yoga, acupuncture, aromatherapy, hypnosis, music therapy, dietary changes, nutritional and dietary supplements, and the like; such treatments are known to those skilled in the art. Any such treatment strategies can be utilized, alone or in combination with one or more additional cancer therapy, in the practice of the present disclosure.

Dosage and Administration Routes

[00305] Other embodiments of the disclosure can include methods of administering or treating an animal, which can involve administering an amount of at least one treatment, such as an immunotherapy, that is effective to treat the disease, condition, or disorder that the organism has, or is suspected of having, or is susceptible to, or to bring about a desired physiological effect. In some embodiments, the composition or pharmaceutical composition comprises at least one treatment, such as an immunotherapy, which can be administered to an animal (e.g., mammals, primates, monkeys, or humans) in an amount of about 0.005 to about 50 mg/kg body weight, about 0.01 to about 15 mg/kg body weight, about 0.1 to about 10 mg/kg body weight, about 0.5 to about 7 mg/kg body weight, about 0.005 mg/kg, about 0.01 mg/kg, about 0.05 mg/kg, about 0.1 mg/kg, about 0.5 mg/kg, about 1 mg/kg, about 3 mg/kg, about 5 mg/kg, about 5.5 mg/kg, about 6 mg/kg, about 6.5 mg/kg, about 7 mg/kg, about 7.5 mg/kg, about 8 mg/kg, about 10 mg/kg, about 12 mg/kg, or about 15 mg/kg. In regard to some conditions, the dosage can be about 0.5 mg/kg human body weight or about 6.5 mg/kg human body weight. In some instances, some subjects (e.g., mammals, mice, rabbits, feline, porcine, or canine) can be administered a dosage of about 0.005 to about 50 mg/kg body weight, about 0.01 to about 15 mg/kg body weight, about 0.1 to about 10 mg/kg body weight, about 0.5 to about 7 mg/kg body weight, about 0.005 mg/kg, about 0.01 mg/kg, about 0.05 mg/kg, about 0.1 mg/kg, about 1 mg/kg, about 5 mg/kg, about 10 mg/kg, about 20 mg/kg, about 30 mg/kg, about 40 mg/kg, about 50 mg/kg, about 80 mg/kg, about 100 mg/kg, or about 150 mg/kg. Of course, those skilled in the art will appreciate that it is possible to employ many concentrations in the methods of the present disclosure, and using, in part, the guidance provided herein, will be able to adjust and test any

number of concentrations in order to find one that achieves the desired result in a given circumstance. In some embodiments, a dose or a therapeutically effective dose of a compound disclosed herein will be that which is sufficient to achieve a plasma concentration of the compound or its active metabolite(s) within a range set forth herein, e.g., about 1-10 nM, 10-100 nM, 0.1-1 μ M, 1-10 μ M, 10-100 μ M, 100-200 μ M, 200-500 μ M, or even 500-1000 μ M, preferably about 1-10 nM, 10-100 nM, or 0.1-1 μ M.

[00306] In other embodiments, a treatment can be administered in combination with one or more other therapeutic agents for a given disease, condition, or disorder.

[00307] The compounds and pharmaceutical compositions are preferably prepared and administered in dose units. Solid dose units are tablets, capsules and suppositories. For treatment of a subject, depending on activity of the compound, manner of administration, nature and severity of the disease or disorder, age and body weight of the subject, different daily doses can be used.

[00308] Under certain circumstances, however, higher or lower daily doses can be appropriate. The administration of the daily dose can be carried out both by single administration in the form of an individual dose unit or else several smaller dose units and also by multiple administrations of subdivided doses at specific intervals.

[00309] A treatment, such as an immunotherapy treatment as described herein can be administered locally or systemically in a therapeutically effective dose. Amounts effective for this use will, of course, depend on the severity of the disease or disorder and the weight and general state of the subject. Typically, dosages used *in vitro* can provide useful guidance in the amounts useful for *in situ* administration of the pharmaceutical composition, and animal models can be used to determine effective dosages for treatment of particular disorders.

[00310] Various considerations are described, e. g. , in Langer, 1990, *Science*, 249: 1527; Goodman and Gilman's (eds.), 1990, Id., each of which is herein incorporated by reference and for all purposes. Dosages for parenteral administration of active pharmaceutical agents can be converted into corresponding dosages for oral administration by multiplying parenteral dosages by appropriate conversion factors. As to general applications, the parenteral dosage in mg/mL times 1.8 = the corresponding oral dosage in milligrams ("mg"). As to oncology applications, the parenteral dosage in mg/mL times 1.6 = the corresponding oral dosage in mg.

An average adult weighs about 70 kg. See e.g., Miller-Keane, 1992, Encyclopedia & Dictionary of Medicine, Nursing & Allied Health, 5th Ed., (W. B. Saunders Co.), pp.1708 and 1651.

[00311] It will be understood, however, that the specific dose level for any particular patient will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, sex, diet, time of administration, route of administration, rate of excretion, drug combination and the severity of the particular disease undergoing therapy.

[00312] In some embodiments, the administration can include a unit dose of one or more treatments in combination with a pharmaceutically acceptable carrier and, in addition, can include other medicinal agents, pharmaceutical agents, carriers, adjuvants, diluents, and excipients. In certain embodiments, the carrier, vehicle or excipient can facilitate administration, delivery and/or improve preservation of the composition. In other embodiments, the one or more carriers, include but are not limited to, saline solutions such as normal saline, Ringer's solution, PBS (phosphate-buffered saline), and generally mixtures of various salts including potassium and phosphate salts with or without sugar additives such as glucose. Carriers can include aqueous and non-aqueous sterile injection solutions that can contain antioxidants, buffers, bacteriostats, bactericidal antibiotics, and solutes that render the formulation isotonic with the bodily fluids of the intended recipient; and aqueous and non-aqueous sterile suspensions, which can include suspending agents and thickening agents. In other embodiments, the one or more excipients can include, but are not limited to water, saline, dextrose, glycerol, ethanol, or the like, and combinations thereof. Nontoxic auxiliary substances, such as wetting agents, buffers, or emulsifiers may also be added to the composition. Oral formulations can include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, and magnesium carbonate. [00275] The quantity of active component in a unit dose preparation can be varied or adjusted from 0.1 mg to 10000 mg, more typically 1.0 mg to 1000 mg, most typically 10 mg to 500 mg, according to the particular application and the potency of the active component. The composition can, if desired, also contain other compatible therapeutic agents.

[00313] A treatment, such as an immunotherapy, can be administered to subjects by any number of suitable administration routes or formulations. The treatment, such as an immunotherapy, can also be used to treat subjects for a variety of diseases. Subjects include but

are not limited to mammals, primates, monkeys (e.g., macaque, rhesus macaque, or pig tail macaque), humans, canine, feline, bovine, porcine, avian (e.g., chicken), mice, rabbits, and rats. As used herein, the term “subject”, unless stated otherwise, encompasses both human and non-human subjects.

[00314] The route of administration of the compounds of the treatments described herein can be of any suitable route. Administration routes can be, but are not limited to the oral route, the parenteral route, the cutaneous route, the nasal route, the rectal route, the vaginal route, and the ocular route. In other embodiments, administration routes can be parenteral administration, a mucosal administration, intravenous administration, subcutaneous administration, topical administration, intradermal administration, oral administration, sublingual administration, intranasal administration, or intramuscular administration. The choice of administration route can depend on the compound identity (e.g., the physical and chemical properties of the compound) as well as the age and weight of the animal, the particular disease (e.g., type of cancer), and the severity of the disease (e.g., stage or severity of cancer). Of course, combinations of administration routes can be administered, as desired.

[00315] Some embodiments of the disclosure include a method for providing a subject with a treatment which comprises one or more administrations of one or more compositions; the compositions may be the same or different if there is more than one administration.

Cell Culture and Cellular Compositions

[00316] In particular embodiments, the cells of the disclosure may be specifically formulated and/or they may be cultured in a particular medium. The cells may be formulated in such a manner as to be suitable for delivery to a recipient without deleterious effects.

[00317] The medium in certain aspects can be prepared using a medium used for culturing animal cells as their basal medium, such as any of AIM V, X-VIVO-15, NeuroBasal, EGM2, TeSR, BME, BGJb, CMRL 1066, Glasgow MEM, Improved MEM Zinc Option, IMDM, Medium 199, Eagle MEM, α MEM, DMEM, Ham, RPMI-1640, and Fischer's media, as well as any combinations thereof, but the medium may not be particularly limited thereto as far as it can be used for culturing animal cells. Particularly, the medium may be xeno-free or chemically defined.

[00318] The medium can be a serum-containing or serum-free medium, or xeno-free medium. From the aspect of preventing contamination with heterogeneous animal-derived

components, serum can be derived from the same animal as that of the stem cell(s). The serum-free medium refers to medium with no unprocessed or unpurified serum and accordingly, can include medium with purified blood-derived components or animal tissue-derived components (such as growth factors).

[00319] The medium may contain or may not contain any alternatives to serum. The alternatives to serum can include materials which appropriately contain albumin (such as lipid-rich albumin, bovine albumin, albumin substitutes such as recombinant albumin or a humanized albumin, plant starch, dextrans and protein hydrolysates), transferrin (or other iron transporters), fatty acids, insulin, collagen precursors, trace elements, 2-mercaptoethanol, 3'-thioglycerol, or equivalents thereto. The alternatives to serum can be prepared by the method disclosed in International Publication No. 98/30679, for example (incorporated herein in its entirety). Alternatively, any commercially available materials can be used for more convenience. The commercially available materials include knockout Serum Replacement (KSR), Chemically-defined Lipid concentrated (Gibco), and Glutamax (Gibco).

[00320] In certain embodiments, the medium may comprise one, two, three, four, five, six, seven, eight, nine, ten, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more of the following: Vitamins such as biotin; DL Alpha Tocopherol Acetate; DL Alpha-Tocopherol; Vitamin A (acetate); proteins such as BSA (bovine serum albumin) or human albumin, fatty acid free Fraction V; Catalase; Human Recombinant Insulin; Human Transferrin; Superoxide Dismutase; Other Components such as Corticosterone; D-Galactose; Ethanolamine HCl; Glutathione (reduced); L-Carnitine HCl; Linoleic Acid; Linolenic Acid; Progesterone; Putrescine 2HCl; Sodium Selenite; and/or T3 (triiodo-L-thyronine). In specific embodiments, one or more of these may be explicitly excluded.

[00321] In some embodiments, the medium further comprises vitamins. In some embodiments, the medium comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 of the following (and any range derivable therein): biotin, DL alpha tocopherol acetate, DL alpha-tocopherol, vitamin A, choline chloride, calcium pantothenate, pantothenic acid, folic acid nicotinamide, pyridoxine, riboflavin, thiamine, inositol, vitamin B12, or the medium includes combinations thereof or salts thereof. In some embodiments, the medium comprises or consists essentially of biotin, DL alpha tocopherol acetate, DL alpha-tocopherol, vitamin A, choline chloride, calcium pantothenate, pantothenic acid, folic acid nicotinamide, pyridoxine, riboflavin, thiamine, inositol,

and vitamin B12. In some embodiments, the vitamins include or consist essentially of biotin, DL alpha tocopherol acetate, DL alpha-tocopherol, vitamin A, or combinations or salts thereof. In some embodiments, the medium further comprises proteins. In some embodiments, the proteins comprise albumin or bovine serum albumin, a fraction of BSA, catalase, insulin, transferrin, superoxide dismutase, or combinations thereof. In some embodiments, the medium further comprises one or more of the following: corticosterone, D-Galactose, ethanolamine, glutathione, L-carnitine, linoleic acid, linolenic acid, progesterone, putrescine, sodium selenite, or triodo-I-thyronine, or combinations thereof. In some embodiments, the medium comprises one or more of the following: a B-27® supplement, xeno-free B-27® supplement, GS21TM supplement, or combinations thereof. In some embodiments, the medium comprises or further comprises amino acids, monosaccharides, inorganic ions. In some embodiments, the amino acids comprise arginine, cystine, isoleucine, leucine, lysine, methionine, glutamine, phenylalanine, threonine, tryptophan, histidine, tyrosine, or valine, or combinations thereof. In some embodiments, the inorganic ions comprise sodium, potassium, calcium, magnesium, nitrogen, or phosphorus, or combinations or salts thereof. In some embodiments, the medium further comprises one or more of the following: molybdenum, vanadium, iron, zinc, selenium, copper, or manganese, or combinations thereof. In certain embodiments, the medium comprises or consists essentially of one or more vitamins discussed herein and/or one or more proteins discussed herein, and/or one or more of the following: corticosterone, D-Galactose, ethanolamine, glutathione, L-carnitine, linoleic acid, linolenic acid, progesterone, putrescine, sodium selenite, or triodo-I-thyronine, a B-27® supplement, xeno-free B-27® supplement, GS21TM supplement, an amino acid (such as arginine, cystine, isoleucine, leucine, lysine, methionine, glutamine, phenylalanine, threonine, tryptophan, histidine, tyrosine, or valine), monosaccharide, inorganic ion (such as sodium, potassium, calcium, magnesium, nitrogen, and/or phosphorus) or salts thereof, and/or molybdenum, vanadium, iron, zinc, selenium, copper, or manganese. In specific embodiments, one or more of these may be explicitly excluded.

[00322] The medium can also contain one or more externally added fatty acids or lipids, amino acids (such as non-essential amino acids), vitamin(s), growth factors, cytokines, antioxidant substances, 2-mercaptoethanol, pyruvic acid, buffering agents, and/or inorganic salts. In specific embodiments, one or more of these may be explicitly excluded.

[00323] One or more of the medium components may be added at a concentration of at least, at most, or about 0.1, 0.5, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 180, 200, 250 ng/L, ng/ml, µg/ml, mg/ml, or any range derivable therein.

[00324] In specific embodiments, the cells of the disclosure are specifically formulated. They may or may not be formulated as a cell suspension. In specific cases they are formulated in a single dose form. They may be formulated for systemic or local administration. In some cases the cells are formulated for storage prior to use, and the cell formulation may comprise one or more cryopreservation agents, such as DMSO (for example, in 5% DMSO). The cell formulation may comprise albumin, including human albumin, with a specific formulation comprising 2.5% human albumin. The cells may be formulated specifically for intravenous administration; for example, they are formulated for intravenous administration over less than one hour. In particular embodiments the cells are in a formulated cell suspension that is stable at room temperature for 1, 2, 3, or 4 hours or more from time of thawing.

[00325] In some embodiments, the method further comprises priming the T cells. In some embodiments, the T cells are primed with antigen presenting cells. In some embodiments, the antigen presenting cells present tumor antigens or peptides, such as those disclosed herein.

[00326] In particular embodiments, the cells of the disclosure comprise an exogenous TCR, which may be of a defined antigen specificity. In some embodiments, the TCR can be selected based on absent or reduced alloreactivity to the intended recipient (examples include certain virus-specific TCRs, xeno-specific TCRs, or cancer-testis antigen-specific TCRs). In the example where the exogenous TCR is non-alloreactive, during T cell differentiation the exogenous TCR suppresses rearrangement and/or expression of endogenous TCR loci through a developmental process called allelic exclusion, resulting in T cells that express only the non-alloreactive exogenous TCR and are thus non-alloreactive. In some embodiments, the choice of exogenous TCR may not necessarily be defined based on lack of alloreactivity. In some embodiments, the endogenous TCR genes have been modified by genome editing so that they do not express a protein. Methods of gene editing such as methods using the CRISPR/Cas9 system are known in the art and described herein.

[00327] In some embodiments, the cells of the disclosure further comprise one or more chimeric antigen receptors (CARs). Examples of tumor cell antigens to which a CAR may be

directed include at least 5T4, 8H9, $\alpha\beta 6$ integrin, BCMA, B7-H3, B7-H6, CAIX, CA9, CD19, CD20, CD22, CD30, CD33, CD38, CD44, CD44v6, CD44v7/8, CD70, CD123, CD138, CD171, CEA, CSPG4, EGFR, EGFR family including ErbB2 (HER2), EGFRvIII, EGP2, EGP40, ERBB3, ERBB4, ErbB3/4, EPCAM, EphA2, EpCAM, folate receptor-a, FAP, FBP, fetal AchR, FR, GD2, G250/CAIX, GD3, Glypican-3 (GPC3), Her2, IL-13R $\alpha 2$, Lambda, Lewis-Y, Kappa, KDR, MAGE, MCSP, Mesothelin, Muc1, Muc16, NCAM, NKG2D Ligands, NY-ESO-1, PRAME, PSC1, PSCA, PSMA, ROR1, SP17, Survivin, TAG72, TEMs, carcinoembryonic antigen, HMW-MAA, AFP, CA-125, ETA, Tyrosinase, MAGE, laminin receptor, HPV E6, E7, BING-4, Calcium-activated chloride channel 2, Cyclin-B1, 9D7, EphA3, Telomerase, SAP-1, BAGE family, CAGE family, GAGE family, MAGE family, SAGE family, XAGE family, NY-ESO-1/LAGE-1, PAME, SSX-2, Melan-A/MART-1, GP100/pmel17, TRP-1/-2, P. polypeptide, MC1R, Prostate-specific antigen, β -catenin, BRCA1/2, CML66, Fibronectin, MART-2, TGF- β RII, or VEGF receptors (e.g., VEGFR2), for example. The CAR may be a first, second, third, or more generation CAR. The CAR may be bispecific for any two nonidentical antigens, or it may be specific for more than two nonidentical antigens.

Toxicity

[00328] The ratio between toxicity and therapeutic effect for a particular treatment is its therapeutic index and can be expressed as the ratio between LD50 (the amount of compound lethal in 50% of the population) and ED50 (the amount of compound effective in 50% of the population). Compounds that exhibit high therapeutic indices are preferred. Therapeutic index data obtained from in vitro assays, cell culture assays and/or animal studies can be used in formulating a range of dosages for use in humans. The dosage of such compounds preferably lies within a range of plasma concentrations that include the ED50 with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. See, e.g. Fingl et al., In: THE PHARMACOLOGICAL BASIS OF THERAPEUTICS, Ch.1, p.1, 1975. The exact formulation, route of administration, and dosage can be chosen by the individual practitioner in view of the patient's condition and the particular method in which the compound is used. For in vitro formulations, the exact formulation and dosage can be chosen by the individual practitioner in view of the patient's condition and the particular method in which the compound is used.

Kits

[00329] Certain aspects of the present invention also concern kits containing compositions of the disclosure or compositions to implement methods of the invention. In some embodiments, kits can be used to evaluate one or more biomarkers or HLA types. In certain embodiments, a kit contains, contains at least or contains at most 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, 50, 100, 500, 1,000 or more probes, primers or primer sets, synthetic molecules or inhibitors, or any value or range and combination derivable therein.

[00330] Kits may comprise components, which may be individually packaged or placed in a container, such as a tube, bottle, vial, syringe, or other suitable container means.

[00331] Individual components may also be provided in a kit in concentrated amounts; in some embodiments, a component is provided individually in the same concentration as it would be in a solution with other components. Concentrations of components may be provided as 1x, 2x, 5x, 10x, or 20x or more.

[00332] In certain aspects, negative and/or positive control nucleic acids, probes, and inhibitors are included in some kit embodiments. In addition, a kit may include a sample that is a negative or positive control for methylation of one or more biomarkers.

[00333] It is contemplated that any method or composition described herein can be implemented with respect to any other method or composition described herein and that different embodiments may be combined. The claims originally filed are contemplated to cover claims that are multiply dependent on any filed claim or combination of filed claims.

Computer Implemented System

[00334] In various embodiments, the systems and methods for recognizing home activities by deep learning subtle vibrations on an interior surface of a house from a single point using vibration sensing devices can be implemented via computer software or hardware. Refer to the Appendix for further information regarding the system, devices and methods provided herein, in accordance with various embodiments.

[00335] Fig. 1 is a block diagram illustrating a computer system 100 upon which embodiments of the present teachings may be implemented. In various embodiments of the present teachings, computer system 100 can include a bus 102 or other communication

mechanism for communicating information and a processor 104 coupled with bus 102 for processing information. In various embodiments, computer system 100 can also include a memory, which can be a random-access memory (RAM) 106 or other dynamic storage device, coupled to bus 102 for determining instructions to be executed by processor 104. Memory can also be used for storing temporary variables or other intermediate information during execution of instructions to be executed by processor 104. In various embodiments, computer system 100 can further include a read only memory (ROM) 108 or other static storage device coupled to bus 102 for storing static information and instructions for processor 104. A storage device 110, such as a magnetic disk or optical disk, can be provided and coupled to bus 102 for storing information and instructions.

[00336] In various embodiments, computer system 100 can be coupled via bus 102 to a display 112, such as a cathode ray tube (CRT) or liquid crystal display (LCD), for displaying information to a computer user. An input device 114, including alphanumeric and other keys, can be coupled to bus 102 for communication of information and command selections to processor 104. Another type of user input device is a cursor control 116, such as a mouse, a trackball or cursor direction keys for communicating direction information and command selections to processor 104 and for controlling cursor movement on display 112. This input device 114 typically has two degrees of freedom in two axes, a first axis (i.e., x) and a second axis (i.e., y), that allows the device to specify positions in a plane. However, it should be understood that input devices 114 allowing for 3-dimensional (x, y and z) cursor movement are also contemplated herein.

[00337] Consistent with certain implementations of the present teachings, results can be provided by computer system 100 in response to processor 104 executing one or more sequences of one or more instructions contained in memory 106. Such instructions can be read into memory 106 from another computer-readable medium or computer-readable storage medium, such as storage device 110. Execution of the sequences of instructions contained in memory 106 can cause processor 104 to perform the processes described herein. Alternatively, hard-wired circuitry can be used in place of or in combination with software instructions to implement the present teachings. Thus, implementations of the present teachings are not limited to any specific combination of hardware circuitry and software.

[00338] The term “computer-readable medium” (e.g., data store, data storage, etc.) or “computer-readable storage medium” as used herein refers to any media that participates in providing instructions to processor 104 for execution. Such a medium can take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Examples of non-volatile media can include, but are not limited to, dynamic memory, such as memory 106. Examples of transmission media can include, but are not limited to, coaxial cables, copper wire, and fiber optics, including the wires that comprise bus 102.

[00339] Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, a RAM, PROM, and EPROM, a FLASH-EPROM, another memory chip or cartridge, or any other tangible medium from which a computer can read.

[00340] In addition to computer-readable medium, instructions or data can be provided as signals on transmission media included in a communications apparatus or system to provide sequences of one or more instructions to processor 104 of computer system 100 for execution. For example, a communication apparatus may include a transceiver having signals indicative of instructions and data. The instructions and data are configured to cause one or more processors to implement the functions outlined in the disclosure herein. Representative examples of data communications transmission connections can include, but are not limited to, telephone modem connections, wide area networks (WAN), local area networks (LAN), infrared data connections, NFC connections, etc.

[00341] It should be appreciated that the methodologies described herein, flow charts, diagrams and accompanying disclosure can be implemented using computer system 100 as a standalone device or on a distributed network or shared computer processing resources such as a cloud computing network.

[00342] The methodologies described herein may be implemented by various means depending upon the application. For example, these methodologies may be implemented in hardware, firmware, software, or any combination thereof. For a hardware implementation, the processing unit may be implemented within one or more application specific integrated circuits (ASICs), digital signal processors (DSPs), digital signal processing devices (DSPDs), programmable logic devices (PLDs), field programmable gate arrays (FPGAs), processors,

controllers, micro-controllers, microprocessors, electronic devices, other electronic units designed to perform the functions described herein, or a combination thereof.

[00343] In various embodiments, the methods of the present teachings may be implemented as firmware and/or a software program and applications written in conventional programming languages such as C, C++, Python, etc. If implemented as firmware and/or software, the embodiments described herein can be implemented on a non-transitory computer-readable medium in which a program is stored for causing a computer to perform the methods described above. It should be understood that the various engines described herein can be provided on a computer system, such as computer system 100, whereby processor 104 would execute the analyses and determinations provided by these engines, subject to instructions provided by any one of, or a combination of, memory components 106/108/110 and user input provided via input device 114.

[00344] In various embodiments, the methods of the present teachings can involve deep learning and/or machine learning and/or one or more neural network, such as a deep neural network, and the like. It should be understood that while deep learning and such processes may be discussed in conjunction with various embodiments herein, the various embodiments herein are not limited to being associated only with deep learning tools. As such, machine learning and/or artificial intelligence tools generally may be applicable as well. Moreover, the terms deep learning, machine learning, and artificial intelligence may even be used interchangeably in generally describing the various embodiments of systems, software and methods herein.

[00345] A deep neural network (DNN) generally, such as a convolutional neural network (CNN), generally accomplishes an advanced form of image processing and classification/detection by first looking for low level features such as, for example, edges and curves, and then advancing to more abstract (e.g., unique to the type of images being classified) concepts through a series of convolutional layers. A DNN/CNN can do this by passing an image through a series of convolutional, nonlinear, pooling (or downsampling, as will be discussed in more detail below), and fully connected layers, and get an output. Again, the output can be a single class or a probability of classes that best describes the image or detects objects on the image.

[00346] Regarding layers in a CNN, for example, the first layer is generally a convolutional layer (Conv). This first layer will process the image's representative array using a

series of parameters. Rather than processing the image as a whole, a CNN will analyze a collection of image sub-sets using a filter (or neuron or kernel). The sub-sets will include a focal point in the array as well surrounding points. For example, a filter can examine a series of 5×5 areas (or regions) in a 32×32 image. These regions can be referred to as receptive fields. Since the filter must possess the same depth of the input, an image with dimensions of $32 \times 32 \times 3$ would have a filter of the same depth (e.g., $5 \times 5 \times 3$). The actual step of convolving, using the exemplary dimensions above, would involve sliding the filter along the input image, multiplying filter values with the original pixel values of the image to compute element wise multiplications, and summing these values to arrive at a single number for that examined portion of the image.

[00347] After completion of this convolving step, using a $5 \times 5 \times 3$ filter, an activation map (or filter map) having dimensions of $28 \times 28 \times 1$ will result. For each additional layer used, spatial dimensions are better preserved such that using two filters will result in an activation map of $28 \times 28 \times 2$. Each filter will generally have a unique feature it represents (e.g., colors, edges, curves, etc.) that, together, represent the feature identifiers required for the final image output. These filters, when used in combination, allow the CNN to process an image input to detect those features present at each pixel. Therefore, if a filter serves as a curve detector, the convolving of the filter along the image input will produce an array of numbers in the activation map that correspond to high likelihood of a curve (high summed element wise multiplications), low likelihood of a curve (low summed element wise multiplications) or a zero value where the input volume at certain points provided nothing that would activate the curve detector filter. As such, the greater number of filters (also referred to as channels) in the Conv, the more depth (or data) that is provided on the activation map, and therefore more information about the input that will lead to a more accurate output.

[00348] Balanced with accuracy of the CNN is the processing time and power needed to produce a result. In other words, the more filters (or channels) used, the more time and processing power needed to execute the Conv. Therefore, the choice and number of filters (or channels) to meet the needs of the CNN method are specifically chosen to produce as accurate an output as possible while considering the time and power available.

[00349] To enable further a CNN to detect more complex features, additional Conv layers can be added to analyze what outputs from the previous Conv layer (i.e., activation maps). For example, if a first Conv layers looks for a basic feature such as a curve or an edge, a second

Conv layer can look for a more complex feature such as shapes, which can be a combination of individual features detected in an earlier Conv layer. By providing a series of Conv layers, the CNN can detect increasingly higher-level features to arrive eventually at the specific desired object detection. Moreover, as the Conv layers stack on top of each other, analyzing the previous activation map output, each Conv layer in the stack is naturally going to analyze a larger and larger receptive field by virtue of the scaling down that occurs at each Conv level, thereby allowing the CNN to respond to a growing region of pixel space in detecting the object of interest.

[00350] A CNN architecture generally consists of a group of processing blocks, including at least one processing block for convoluting an input volume (image) and at least one for deconvolution block (or transpose convolution). Additionally, the processing blocks can include at least one pooling block and unpooling block. Pooling blocks can be used to scale down an image in resolution to produce an output available for Conv. This can provide computational efficiency (efficient time and power), which can in turn improve actual performance of the CNN. Those these pooling, or subsampling, blocks keep filters small and computational requirements reasonable, these blocks coarsen the output (can result in lost spatial information within a receptive field), reducing it from the size of the input by a factor equal to the pixel stride of the receptive fields of the output units.

[00351] Unpooling blocks can be used to reconstruct a these coarse outputs to produce an output volume with the same dimensions as the input volume. An unpooling block can be considered a reverse operation of a convoluting block to return an activation output to the original input volume dimension.

[00352] However, the unpooling process generally just simply enlarges the coarse outputs into a sparse activation map. To avoid this result, the deconvolution block densifies this sparse activation map to produce both and enlarged and dense activation map that eventually, after any further necessary processing, a final output volume with size and density much closer to the input volume. As a reverse operation of the convolution block, rather than reducing multiple array points in the receptive field to a single number, the deconvolution block associate a single activation output point with a multiple outputs to enlarge and densify the resulting activation output.

[00353] It should be noted that while pooling blocks can be used to scale down an image and unpooling blocks can be used to enlarge these scaled down activation maps, convolution and deconvolution blocks can be structured to both convolve/deconvolve and scale down/enlarge without separate pooling and unpooling blocks.

[00354] The pooling and unpooling process can be limited depending on the objects of interest being detected in an image input. Since pooling generally scales down an image by looking at sub-image windows without overlap of windows, there is a clear loss in spatial info as the scaling down occurs.

[00355] A processing block can include other layers that are packaged with a convolutional or deconvolutional layer. These can include, for example, a rectified linear unit layer (ReLU) or exponential linear unit layer (ELU), which are activation functions that examine the output from a Conv layer in its processing block. The ReLU or ELU layer acts as a gating function to advance only those values corresponding to positive detection of the feature of interest unique to the Conv layer its processing block.

[00356] Given a basic architecture, the CNN is then prepared for a training process to hone its accuracy in image classification/detection (of objects of interest). Using training data sets, or sample images used to train the CNN so that it updates its parameters in reaching an optimal, or threshold, accuracy, a process called backpropagation (backprop) occurs. Backpropagation involves a series of repeated steps (training iterations) that, depending on the parameters of the backprop, either will slowly or quickly train the CNN. Backprop steps generally include forward pass, loss function, backward pass, and parameter (weight) update according to a given learning rate. The forward pass involves passing a training image through the CNN. The loss function is a measure of error in the output. The backward pass determines the contributing factors to the loss function. The weight update involves updating the parameters of the filters to move the CNN towards optimal. The learning rate determines the extent of weight update per iteration to arrive at optimal. If the learning rate is too low, the training may take too long and involve too much processing capacity. If the learning rate is too fast, each weight update may be too large to allow for precise achievement of a given optimum or threshold.

[00357] The backprop process can cause complications in training, thus leading to the desire for lower learning rates and more specific and carefully determined initial parameters

upon start of training. One such complication is that, as weight updates occur at the conclusion of each iteration, the changes to the parameters of the Conv layers amplify the deeper the network goes. For example, if a CNN has a plurality of Conv layers that, as discussed above, allows for higher-level feature analysis, the parameter update to the first Conv layer is multiplied at each subsequent Conv layer. The net effect is that the smallest changes to parameters have large impact depending on the depth of a given CNN. This phenomenon is referred to as internal covariate shift.

[00358] It should be noted that even though CNNs are spoken about in detail above, the various embodiments discussed herein could utilize any neural network type or architecture.

[00359] In describing the various embodiments, the specification may have presented a method and/or process as a particular sequence of steps. However, to the extent that the method or process does not rely on the particular order of steps set forth herein, the method or process should not be limited to the particular sequence of steps described. As one of ordinary skill in the art would appreciate, other sequences of steps may be possible. Therefore, the particular order of the steps set forth in the specification should not be construed as limitations on the claims. In addition, the claims directed to the method and/or process should not be limited to the performance of their steps in the order written, and one skilled in the art can readily appreciate that the sequences may be varied and still remain within the spirit and scope of the various embodiments. Similarly, any of the various system embodiments may have been presented as a group of particular components. However, these systems should not be limited to the particular set of components, now their specific configuration, communication and physical orientation with respect to each other. One skilled in the art should readily appreciate that these components can have various configurations and physical orientations (e.g., wholly separate components, units and subunits of groups of components, different communication regimes between components).

[00360] Although specific embodiments and applications of the disclosure have been described in this specification, these embodiments and applications are exemplary only, and many variations are possible. Having described the disclosure in detail, it will be apparent that modifications, variations, and equivalent embodiments are possible without departing from the scope of the disclosure defined in the appended claims. Furthermore, it should be appreciated that all examples in the present disclosure are provided as non-limiting examples.

EXAMPLES

[00361] The following non-limiting examples are provided to further illustrate embodiments of the disclosure disclosed herein. It should be appreciated by those of skill in the art that the techniques disclosed in the examples that follow represent approaches that have been found to function well in the practice of the disclosure, and thus can be considered to constitute examples of modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments that are disclosed and still obtain a like or similar result without departing from the spirit and scope of the disclosure.

EXAMPLE 1

Materials and Methods

[00362] The materials and methods used in Examples 2-11 are described below.

SNAF Architecture

[00363] Splicing Neo-Antigen Finder (SNAF) was designed as a modular python package to automate splicing neoantigen identification, using a series of embedded workflows. This workflow consists of distinct steps (see below) which are divided into functionally distinct modules which can be called on a single SNAF python object or independently produced data files, which can be mixed and matched to identify T-cell or B-cell neoantigens or perform orthogonal evaluations (e.g., survival, MS proteomics, long-read analysis). Additional documentation and tutorials are provided from the GitHub repository (<https://github.com/frankligy/SNAF>).

Alternative Splicing Quantification (Step 1)

[00364] SNAF depends on the identification and annotation of gene-associated exon-exon and exon-intron junctions, in addition to alternative splicing quantification. As it has been extensively validated for splicing quantification (experimental, synthetic), relative to other well-described approaches, the previously described method Multipath-PSI embedded in the software suite AltAnalyze was integrated for junction and splicing analyses.

[00365] Specifically, AltAnalyze iterates over each BAM file, to extract exon-exon junctions from alignments (STAR 2.6.1 in these analyses), based on the presence of an “N” in the CIGAR string, exported to a sample-specific exon-exon junction genome coordinate file,

considering predicted strand (Figure 3A). AltAnalyze re-iterates over these BAM files to identify high-confidence intron retention events, considering reads that span the exon-intron boundary, have paired-end reads in the intron (Figure 3B).

[00366] Prior to exporting exon-intron spanning read counts to a separate file, each intron is required to have evidence of intron retention on both ends of the intron (5' and 3') with at least 5 supporting paired-end reads (exon-intron spanning and intron only pair) on each side of the intron. The method also eliminates reads that meet this criterion, but have a splice-site in one of the examined paired-ends within the intron (e.g., novel exons within the intron or alternative splice-sites). Both exon-exon and exon-intron splicing events have been previously experimentally validated. When single-end sequencing data is provided, the same rules are enforced, but no requirement for a paired-end read is fully contained within the intron.

[00367] AltAnalyze applies a rigorous annotation model to curate detected exon-exon junctions and retention introns, based on an Ensembl gene model supplemented by transcripts downloaded from UCSC (annotated mRNAs) for matched genome and transcriptome builds (Figure 3C). This is the same gene model used by SNAF; however, the software can effectively use any Ensembl version database via AltAnalyze. Only exon-exon junctions in which one splice-site is matched to this database of known isoforms are confidently assigned to a gene model. This gene model has standardized annotations for each exon, divided into blocks (E1, E2, E3, etc.), where each block represents a consensus exon from different transcript sources, and regions (E1.1, E1.2, E1.3, etc.) are defined by different annotated splice-sites.

[00368] Novel splice-sites from exon-exon junctions sequenced are annotated according to their proximity to known splice sites or as novel exons within an intron. Such novel splice sites are prime targets for neoantigen predictions. The final junction IDs are noted based on their gene, exon block, regions and genomic coordinates. The junction read counts are used directly for analysis in SNAF and in conjunction with alternative splicing measurements derived from AltAnalyze automatically curated events (cassette exon, 3'/5' splice site exon, intron retention, alternative terminal exon, trans-splicing, alternative promoter), using the MultiPath-PSI. In MultiPath-PSI, each junction is compared to those it overlaps within the genome and that are associated with the same gene (exon-exon or exon-intron) to compute a ratio of that junction relative to all overlapping junctions (Percent Spliced-In or PSI). For each splicing event, only those with a read count > 19, considering the sum of counts for all overlapping junctions, are

computed. For all others, a missing value is reported. The selected thresholds were supported as being robust, when benchmarked on simulated differential splicing analysis versus other prior reported local splicing variation analysis tools (see Example 1, as well as Figure 4).

[00369] This workflow was applied to all cancer bulk RNA-Seq samples evaluated in this study, as well as a subset of GTEx healthy tissue samples from 54 body sites and TCGA control samples, comprising a total of 2,629 samples (25/tissue on average). Two additional normal cohorts were included: 1) TCGA para-tumor sites of 705 samples from 22 tissue types, and 2) an additional 313 GTEx skin samples to complement the lack of TCGA skin control in the database. The addition of these controls further improved the specificity of SNAF predictions (fewer reported hits) (see Figure 5). The workflow optionally provides a proliferating tissue database as an optional control dataset; however, this is not included by default as aberrant cell proliferation is indeed a hallmark of various cancer types (Figure 17). As AltAnalyze v.2.1.4 is only compatible with Python 2.7, it has been integrated with SNAF via Docker containerization.

[00370] SNAF and other splicing neoantigen approaches uses standardized alignment of FASTQ files to the genome and transcriptome. In particular, STAR version 2.4.0, with two-pass aligned to GENCODE version 36, was applied for the control database to match the detection thresholds used for the TCGA samples analyzed.

Prediction of Immunogenic MHC-Bound T-cell Neoantigens (Step 2)

[00371] To identify splice-junctions that are specific to tumor samples, relative to a database of healthy control tissues (GTEx, 54 body sites), SNAF computes the tumor specificity of each junction. This score is defined as the average read count for a splice junction in GTEx as N_{mean} , and the maximum read count for a junction across samples in the evaluated tumor cohort as T_{max} . By default, a splice event is considered tumor-specific (neojunction) if and only if the following two criteria are met:

$$N_{\text{mean}} < 3 \quad (1)$$

$$T_{\text{max}} - N_{\text{mean}} > 20 \quad (2)$$

[00372] These two cutoffs are adjustable based on specific datasets and conditions. The default cutoffs were selected in accordance with those previously recommended from a pan-cancer splicing neoantigen analysis. If a user supplies matching controls for one or more tumor

samples, the matched control will be automatically concatenated to the existing GTEx database to calculate the integrated average read count. The filtered tumor-specific splice junction is referred to as a neojunction.

[00373] For each neojunction, the junction RNA sequence is first retrieved by concatenating the 5' and 3' exon/intron sequences, followed by *in silico* translation into possible k-mers, considering all the three possible reading frames (Figure 2). A valid short peptide should meet two criteria: (1) absence of a stop codon, and (2) at least one amino acid should span the splice-junction site, such that the junction peptide will not derive from only the 5' donor or the 3' acceptor exon/intron. All possible strand-specific reading frames are used to account for both potential tumor-specific alternative protein translations and to increase the sensitivity to detect novel neoantigens.

[00374] SNAF next restricts the number of T cell neoantigen candidates by considering MHC binding prediction and T cell immunogenicity. Only a few MHC-bound peptides are likely to elicit T cell response (Figure 2). Specifically, SNAF provides interfaces to two popular MHC binding prediction models – NetMHCpan4.1 and MHCflurry2.0 to prioritize junction-peptides that are likely to bind with patient-specific HLA sequences. NetMHCpan4.1 and MHCflurry2.0 have the highest reported accuracy for MHC binding based on recent benchmarking. MHCflurry is directly embedded with SNAF, while NetMHCpan4.1 is available for direct integration, following authorization of the software's license from the software's website.

[00375] To produce the most accurate immunogenicity predictions, SNAF leverages the inventors' recently published deep convolutional neural network approach DeepImmuno, which ranked among the best-performing models in a recent independent benchmarking study. The resulting predicted neoantigens are expected to have a high probability of MHC binding and immunogenicity (Figure 2).

Prediction of ExNeoEpitopes in SNAF-B (Step 3)

[00376] Extracellular epitopes (B cell antigens) identified by SNAF use the same initial neoantigen discovery workflows described above (Steps 1, 2), with the exclusion of MHC peptide binding and immunogenicity prediction SNAF module calls. To further predict full-length protein coding ExNeoEpitopes, SNAF-B can limit neojunctions to only those which encode for transmembrane spanning protein coding genes.

[00377] A heuristic approach is used to infer full-length isoform from all existing transcripts and local splicing events. For example, if a tumor-specific splicing event E2.1-E3.1 is observed, along with a documented transcript whose exon composition is E1.1|E2.1|E2.2|E3.1|E4.1, the potential full-length isoform resulting from this splice-event will be E1.1|E2.1|E3.1|E4.1 (E2.2 skipping) (Figure 2).

[00378] This workflow is biased to detect exon-skipping or alternative splice-sites that result in shorter peptides, but not isoforms with longer protein coding sequences (novel exons and splice-sites). To account for this class of insertion peptides, SNAF implements a parallel strategy leveraging input cancer long-read sequencing data from a gtf file. Specifically, for a specific alternative splicing event associated junction (i.e. E2.1-E3.1), SNAF identifies an associated full-length isoform by matching the exon-exon junction splice-site coordinates to those present in the long-read isoform (presumably tumor-associated) GTF file. For intron retention, SNAF determines whether both intron boundaries fall within a long-read supported single-exon region. Such matching long-read evidenced isoforms represent additional candidate ExNeoEpitopes.

[00379] **SNAF-B reference-augmented workflow (de novo prediction):** When no matching mRNA isoform contains the neojunction, the most appropriate reference isoform is first selected for a given neojunction. This analysis (short_read mode), produces a theoretical mRNA isoform based on the software selected reference isoform. To accomplish this, SNAF-B leverages a unique feature of the software AltAnalyze, which is the explicit annotation of exon and intron blocks and regions. Specifically, all AltAnalyze documented isoforms are decomposed into program annotated exon blocks and regions, where each exon block is subdivided into regions defined by alternative splice site locations, using a series of prior described rules. When RNA-Seq data is analyzed, all exon-exon and exon-intron junctions are annotated according to this schema, with novel exons or splice sites assigned unique IDs relative to the block and region they are situated within. In SNAF, if the neojunction represents the excision of an exon block or specific exon region, the software identifies the longest protein-coding isoform from AltAnalyze's reference isoform collection that contains both exon regions that comprise the junction. The collection consolidates mRNA and protein isoforms from Ensembl the UCSC genome browser database (Ensembl version 91 and matched UCSC database for this paper). For example, the neojunction ENSG00000056291:E2.2-E5.1 represents an exon

skipping isoform. The corresponding longest protein coding isoform that contains these two exons is ENST00000395999 (ENSP00000379321), which is represented as E2.1|E2.2|E3.1|E4.1|E5.1|E7.1|E7.2|E7.3. SNAF represents the ExNeoEpitope from this reference form as E2.1|E2.2|E5.1|E7.1|E7.2|E7.3, by excluding the skipped exon block and regions. If the isoform is an inclusion form, occurring due to the presence of a novel splice-site, exon or intron-retention event, this function lacks the specificity to infer the full-length isoform and neojunction prediction is skipped.

[00380] SNAF-B Long-Read Alignment workflow: As an alternative approach to directly identify novel mRNA isoforms associated with short-read bulk RNA-Seq-derived neojunctions from SNAF, the SNAF-B pipeline includes specialized methods to identify novel mRNA isoforms that can encode for full-length transmembrane proteins. This analysis (long_read mode), produces an inferred protein isoform based on a long-read sequence match. This workflow leverages an externally provided mRNA isoform gene transfer format (GTF) text files, containing the genomic coordinates corresponding to the isoform exon positions for the genome database being analyzed (e.g., hg38). The SNAF-B predicted neojunction genomic coordinates are matched to the provided GTF to identify the longest cDNA junction match. These can include bulk or single-cell long-read sequencing predictions or ESTs from one or more studies/repositories. In this study, two GTF long-read isoform files from a prior cancer cell-line compendium dataset produced using Iso-Seq (PacBio Sequel) were used, along with a separate internal melanoma cell line Iso-Seq dataset, detailed below.

[00381] To assess the translational potential of these candidate ExNeoEpitopes, an efficient Open Reading Frame (ORF) finding algorithm was implemented based on three criteria: (1) ORF length, (2) GC content, and (3) codon usage bias. ORFs of higher length are more likely to be the most valid translation. When the length of two alternative ORF predictions are less than 8nt, the ORF with higher GC content and higher codon usage bias is prioritized. As guanine-cytosine (GC) pairs have three hydrogen bonds compared to two between adenine-thymine (AT) pairs, higher GC content is an indication of stable transcripts in most gene-finding prediction algorithms.

[00382] Codon usage bias measures the species-specific tendency to prefer one codon over the other, which can be used to prioritize valid protein coding genes. The codon frequency table is obtained from <https://www.genscript.com/tools>

<slash> codon-frequency-table. The cumulative frequency value across the ORF (each triplet will have a codon frequency value) is normalized to a value between 0-1 and added to the GC content percentage as the score for ORF prioritization. The predicted ORF is assessed for Nonsense Mediated Decay (NMD) potential (Figure 2).

[00383] To assess NMD, SNAF measures the distance between the end position of the ORF and the last nucleotide of the transcript and removes any ORF predictions that result in a premature stop code more than one exon away from the terminal exon in that transcript ($n_stride=2$), where n_stride is an adjustable parameter. The NMD check is performed for all inferred short-read predicted inferred isoforms, as a reference isoform with an established ORF, but not for long-read isoforms, as the ORF must be inferred. Any predicted isoforms that have a 100% match with a documented membrane protein isoform in the Uniprot database are excluded, as these are presumed to be expressed in non-malignant tissues.

[00384] To increase the accuracy of these predictions, the resultant novel proteins are subjected to TMHMM (Figure 2), a Hidden Markov Model (HMM)-based method to predict the topology of the membrane proteins. Proteins without predicted transmembrane domains will be eliminated for further analysis. As SNAF is modular, these can be optionally skipped by the user to identify additional candidates. The user can optionally supply long-read sequencing data (GTF) to the program to validate predicted full-length isoforms derived from junctions alone (Figure 2). The resultant novel membrane isoform will be reported as high-confidence B cell neoepitopes.

[00385] To determine whether a predicted novel protein will result specifically in extracellular impacts or alter transmembrane domain composition, the individual TMHMM predictions can be explored and compared to different possible reference protein isoform sequences from the SNAF interactive viewer (external linkout) on a case-by-case basis. In this analysis, a secondary manual inspection of the included or deleted polypeptide was performed in UniProt to determine which topological domain and transmembrane features they localize with, relative to the reference protein curation for that gene.

[00386] To further identify candidate ExNeoEpitopes in these analyses, peptides that were contained within existing isoform protein sequences from Ensembl or UCSC that represent a 100% string match (AltAnalyze build database) were eliminated. These principally comprise novel alternative promoter exons that are not predicted to encode for protein sequences.

Alignment Software and Reference Evaluation

[00387] The previously described GDC-TCGA workflow was replicated for sequence alignment for the alignment of control samples to GENCODE version 36. To confirm the validity of these predictions, two TCGA breast cancer samples, FASTQ files (bedtools) were converted using the TCGA workflow and alternative settings or genome reference versions, followed by processing in AltAnalyze.

Benchmark of MultiPath-PSI on Splicing Detection

[00388] Benchmarking of MultiPath-PSI was performed against several recently reported methods for local splicing variation analyses (MAJIQ (version 1.0), LeafCutter (version 0.2.1), and rMATS (version 3.0.9)). As a gold-standard to evaluate prediction accuracy and error rates, a simulated RNA-Seq dataset derived from known and novel isoform predictions was created for pluripotent stem cells (PSC) and *in vitro* derived day 30 cardiomyocytes (CM) (<https://www.synapse.org/#!Synapse:syn12104338>).

[00389] To create this simulation dataset, known and novel isoforms and expression estimates were predicted using Cufflinks on biological triplicate samples (PSC and CM). The Cufflinks isoform GTF and expression values (isoforms.fpkms_tracking) were supplied as input for the software Polyester to produce simulated RNA-Seq reads at a depth of 100 million paired end reads using the software default parameters. These subsequent FASTA files were supplied as input for genome alignment (hg19) with the software STAR to produce BAM files with aligned, stranded, paired end reads. The resulting aligned sequencing data were evaluated with each analysis method and benchmarked against isoform-derived alternative splicing events (exon-exon and exon-intron junctions) comparing CM to PSC. Supervised group comparison analyses with MultiPath-PSI were performed using an FDR corrected (Benjamini-Hochberg) empirical Bayes moderated t-test $P\text{-value} < 0.05$.

[00390] To evaluate the performance of the different algorithms in reporting the differential splicing events, precision-recall curves were generated using Matlab for the synthetic dataset. For each method, the significant splicing events were ranked in descending order according to the empirical score or p-value reported by each algorithm. Each algorithm reported only a portion of the gold-standard splicing events as significant. To generate precision-recall curves that extend to recall=1, the non-reported gold-standard splicing events were included additionally with a p-value=0 or lowest empirical score. This inclusion leads to the sudden peak

in precision with recall identified in the tail of the precision-recall curves for each tool. Comparison of unique splice junction clusters produced by each algorithm (determined from the junction-graph of junction genomic coordinates output by each method) to these gold-standards indicated that MultiPath-PSI has the greatest overall accuracy and lowest error rates, based on precision and recall estimates (Figure 4).

Bulk and Single-Cell Melanoma Splicing Evaluation Datasets

[00391] To evaluate shared and unique splicing neoantigens identified by SNAF, prior reported bulk and single-cell RNA-Seq datasets were reanalyzed through SNAF, using the same genome alignment and splicing quantification workflows applied to TCGA samples. To assess the association of melanoma splicing neoantigens with non-cancerous proliferative skin cell splicing events, five proliferative melanocyte RNA-Seq datasets were reanalyzed in the GEO database (GSE102983, GSE111786, GSE149189, GSE197471, GSE202700) (Figure 17).

[00392] To determine the cell of origin for melanoma splicing antigens, the analysis obtained access to the controlled access raw sequencing data (DUOS-000002), for 4,645 individual cell transcriptomes corresponding to 19 patients with melanoma (GSE72056). Only 3,877 with cell annotations were retained for further analysis. These individual cell-level FASTQ files were re-analyzed in STAR and AltAnalyze to produce aggregate junction read counts for each patient and author annotated cell-populations. These junction read counts were summed per cell-population to identify tumor versus immune neojunction enrichments (fold>2 enriched). These analyses are biased towards immune cells, as twice as many immune cells (n=2,605) versus tumor (n=1,174) were present.

Calculating Tumor Specificity

[00393] SNAF can calculate two optional continuous scores (maximum likelihood estimation and Bayesian hierarchical), measuring the tumor specificity of each splice junction based on its expression across diverse tissues. These methods are designed to rank and prioritize SNAF neoantigens for experimental validation, using different normal tissue references. These scores can be optionally used to filter results by the user (not by default). Mathematically, tumor specificity for a neojunction (nj) is defined as $\sigma_{nj} \in [0,1]$ given its raw read counts as a vector $X_{nj} \in \mathbb{R}^n$ where n indicates the total number of samples in GTEx and TCGA paratumor

samples (in the present case, $n=3,644$). Since GTEx contains over 50 different body site regions, the collection of normal tissues was defined as $TT \in \{brain, skin, heart, lung, \dots\}$, so that tissue-specific raw read counts can be expressed as a vector X_{nj}^{tt} .

[00394] **Maximum Likelihood Estimation (Frequentist Approach):** To account for the sequencing depth bias of different GTEx samples, first the normalized raw read counts $\hat{X}_{nj}$ are computed, as follows in Equation 3:

$$\hat{X}_{nj} = \frac{X_{nj}}{\text{total counts per sample}} \times 10^6 \quad (3)$$

[00395] This value is equivalent to the widely used Count Per Million (CPM) in single-cell sequencing, as there is no need to account for length bias for each splice junction. Next, the following is assumed as in Equation 4:

$$\hat{X}_{nj} \sim \text{HalfNorm}(\sigma = 1 - \sigma_{nj}) \quad (4)$$

[00396] The Likelihood function for observing $\hat{X}_{nj}$ given σ_{nj} can be formalized as in Equation 5:

$$LL = n \ln \frac{\sqrt{2}}{\sigma\sqrt{\pi}} - \frac{1}{2\sigma^2} \sum_{j=1}^n \hat{X}_{nj}(l) \quad (5)$$

[00397] The parameter σ_{nj} can be estimated using Maximum Likelihood Estimation (MLE) according to Equations 6-7:

$$\hat{\sigma} = \operatorname{argmax} \left(\frac{1}{2\sigma^2} \sum_{j=1}^n \hat{X}_{nj}(l) - n \ln \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \right) \quad (6)$$

$$\hat{\sigma}_{nj} = 1 - \hat{\sigma} \quad (7)$$

[00398] The MLE procedure is implemented using the Python `scipy.stats.halfnorm.logpdf` function.

[00399] **Bayesian Hierarchical Approach:** Although the MLE approach takes into account the whole expression vector across GTEx samples, the distribution of splicing expressions across tissue types was not explicitly modeled. The present study therefore implemented a hierarchical Bayesian model to consider the different distributions of the splicing expression values, which has been further developed into an unified tumor specificity score

model called BayesTS for all types of molecular tumor targets including protein-coding genes. Briefly, this method assigns a higher tumor specificity score to predicted neojunctions with infrequent versus frequent expression in healthy controls (GTEx) across all tissues and donors. While an ideal neojunction will never be detected in healthy tissues, low-level and infrequent expression in a subset of donor tissues should not disqualify a potential neoantigen, while frequent low-level neojunction expression represents an unacceptable risk of off-target effects.

[00400] Herein, a random variable C_{nj} has been defined, representing the number of samples with non-zero read counts for a neojunction (nj) in each tissue (tt). For example, if there are 5 out of 25 hippocampus samples that have non-zero read counts, then $C_{nj}^{hippocampus} = 5$. This assumes:

$$C_{nj} \sim \text{Poisson}(\lambda) \quad (8)$$

[00401] The rate parameter λ , is determined by the underlying tumor specificity score σ and coefficient β^c_{nj} to account for the magnitude difference between σ and observed count C :

$$\lambda_{nj} = \beta^c_{nj} \sigma \quad (9)$$

[00402] where β^c follows a gamma distribution $\text{Gamma}(25,1)$ with the intuition that a NeoJunction present at a medium level should have around 12.5 samples expressing it within each tissue. Leveraging the advantages of Bayesian approach to incorporate multiple types of observations, this analysis assumes the underlying σ also gives rise to the normalized count data as defined in the MLE section above, such that:

$$\hat{X}_{nj} \sim \text{LogNorm}(\beta^x_{nj}, \sigma) \quad (4)$$

where β^x follows another gamma distribution $\text{Gamma}(10,1)$ to account for the magnitude difference. Log Norm distribution was chosen due to its exponentially increased space, which shows better convergence in bayesian inference than half norm distribution. A weakly informative prior $\text{Beta}(2,2)$ was chosen for the underlying σ . And finally,

$$\hat{\phi}_{nj} = 1 - \hat{\phi} \quad (7)$$

[00403] Using the same assumptions as the MLE model, this again assumes $\hat{X}_{nj}$ follows the Harf Norm distribution:

$$\hat{x}_{nj} \sim \text{HalfNorm}(\sigma = 1 - \sigma_{nj}) \quad (4)$$

[00404] This study leverages the Stochastic Variational Inference solver implemented in Python pymc3 package to efficiently estimate the σ_{nj} and the confidence interval. The mean value of the posterior distribution of σ_{nj} serves as the final tumor specificity score.

[00405] In this study, results from both scoring methods are provided; however, to be inclusive, the analysis has not restricted neoantigens using either tumor specificity exclusion method. Each tumor specificity measurement has its own strength and weakness. While the MLE approach is fast and can distinguish splice-junctions that are frequently expressed in normal tissue versus being tumor specific, it tends to report zero for junctions that are extremely rare in normal tissues (i.e., splice junctions with an average read count < 1.5 - empirically observed). This is partially due to the fact that MLE does not take into account the distribution of reads across all tissues. Alternatively, the Bayesian approach can identify difficult-to-identify non-tumor-specific junctions, such as in the context of lowly expressed genes, by considering the distribution of junction expression across tissue types, but results in a significantly longer runtime. While still an ongoing area of evaluation, these methods can be used in combination when desiring the highest confidence tumor-specific junctions (i.e., initial MLE and Bayesian on remaining junctions). To verify tumor specificity predictions, IGV visualization was performed on an independent deeply sequenced RNA-Seq dataset of healthy control tissue (ArrayExpress: E-MTAB-2836).

RBP Activity Inference

[00406] For both training and testing of RBP activity inference methods, extensive prior RBP knockdown, CRISPR and eCLIP *in vitro* profiling datasets were downloaded and reanalyzed. To produce the prior network for evidenced RBP-junction interactions, K562 cell line knockdown and CRISPR RNA-Seq data (pair-end) were downloaded from the ENCODE website (<https://www.encodeproject.org/>). With a few exceptions, all RBPs analyzed had two replicates present and two matching controls.

[00407] In total, 484 RNA-Seq samples were downloaded spanning 191 RBPs. These FASTQ files were aligned to human genome version hg38 using STAR version 2.6.1, followed by AltAnalyze 2.1.4.2 to obtain splicing junctions from the resultant BAM files. As batch effects in these data were previously reported and corrected for, these corrections were reproduced using

limma on the splice junction count results prior to computing the PSI matrix in AltAnalyze. Only RBP perturbation alternative splicing events that were predicted using both the non-batch effects corrected data (comparing the two replicates and matched controls) and the batch effects corrected data (versus all 194 combined controls), were retained to ensure specificity of the predictions. K562 eCLIP bed files were downloaded from the ENCODE website (<https://www.encodeproject.org/>) spanning over 120 RBPs. Only high-confidence eCLIP peaks were retained from technical replicate samples, using Irreproducible Discovery Rate (IDR) analysis (idr 2.0.4 software) using an FDR threshold of 0.1.

[00408] From the resultant IDR peaks, RBPs that have direct regulatory evidence for an alternative splicing event were defined, where >1 IDR peak is proximal (exon/intron plus the intermediate downstream and upstream flanking regions). From these predictions, a K562 prior splicing regulatory network was produced using rule-based criteria. Specifically, edge weight = 1 if the RBP is evidenced by both eCLIP and perturbation for a given splice junction, 0.5 if only one evidence modality is present and 0 if neither.

[00409] To assess the performance of commonly used TF activity inference algorithms, the K562 prior network was used. Combined with the context-specific PSI splicing matrix (AltAnalyze), activity inference algorithms will compute the activity of each RBP in each queried sample. Ten commonly used TF activity inference algorithms implemented in decouplerPy package were tested: AUCell, GSEA, GSVA, Multivariate Decision Tree (MDT), Multivariate Linear Model (MLM), Over Representation Analysis (ORA), Univariate Linear Model (ULM), VIPER, Weighted Sum (Wsum) and Weighted Mean (Wmean). GSEA analysis was run using either the “estimate” (GSEA) or the normalized GSEA score (GSEA_norm). GSVA analysis was run using `mx_diff = False` or `True`, the former setting will penalize the events with opposite predictions than expected (increased versus decreased regulation), whereas the latter will erase the directionality, resulting in a total of 12 tested methods. The accuracy of inferred RBP activity was evaluated in ENCODE HepG2 cell line RNA-Seq knockdown data, spanning 193 RBPs across 474 samples. Splicing events with less than 75% non-zero PSI values were filtered out and the missing PSI values were imputed by the median of present values per event. Finally, splicing events with Coefficient of Variation (CV) less than 0.1 were excluded (uninformative).

[00410] An accurate inference should reflect the reduced activity of each RBP in the knockdown samples compared to controls, based on the K562 prior network. This is analogous to the multi-label ranking problem and hence, Label Ranking Average Precision (LRAP) and Normalized Discounted Cumulative Gain (NDCG) can be used to quantify performance. For each RBP, the ground-state truth label in the knockdown sample should equal 0, whereas the controls should equal 1. The total number of knockdown and control samples was denoted as S , the ground state truth label as a binary vector $\mathbf{X}_S$, and the predicted RBP activity as a vector $\mathbf{Y}_S$ where the higher value indicates higher RBP activity. The metrics are defined as below:

$$LRAP = \sum_{k=1}^S \frac{|L_k|}{rank_k} \quad (15)$$

where $L_k = \{k: X_k = 1, Y_k \geq Y_j\}$ and $rank_k = \{k: Y_k \geq Y_j\}$.

[00411] The intuition is that for each control sample we identify samples whose predicted RBP activity is higher than itself. A higher LRAP means most of the control samples receive a higher RBP activity score and vice versa.

[00412] NDCG is originally used to assess the usefulness of a web research result, by determining the fraction of top results which are relevant to the query, times a decay factor based on the rank of the results. First, all samples are ranked based on the predicted RBP activity in descending order, so the highest sample will receive the rank 1. The Discounted Cumulative Gain (DCG) is computed as below:

$$DCG = \sum_{k=1}^S \frac{r_k}{\log(1+r)} \quad (16)$$

[00413] NDCG normalizes the DCG value by considering the perfect DCG score it can get, such that NDCG is ranging from 0 to 1 where as a higher values indicate a higher accuracy.

MS Validation of Ovarian and Melanoma Tumor Biopsies

[00414] Immunopeptidome Mass Spectrum sequencing data from 14 ovarian tumor biopsies was obtained from [ftp <colon slash slash> ftp <dot> ebi <dot> ebi <dot> ac <dot> uk <slash> pride-archive <slash> 2017 <slash> 11 <slash> PXD007635](ftp://ftp.ebi.ac.uk/pride/archive/2017/11/PXD007635). Matched RNA-Seq data was downloaded from the Sequence Read Archive project PRJNA398141. This single-end RNA-Seq data was processed in AltAnalyze as described above (Step 1). To increase sensitivity for

this single-end data, SNAF was run on all quantified exon-exon and exon-intron junctions, rather than additionally filtering on differential splicing events from MultiPath-PSI. Patient HLA type was provided previously.

[00415] The MS profiles were analyzed using the software MaxQuant 2.0.3.1 ([https <colon slash slash> www <dot> maxquant <dot> org](https://www.maxquant.org/)), run on a Linux High Performance Compute (HPC) environment using mono version 5.20.1. A customized MHC-bound splicing neoantigen database was built using `snaf.proteomics.remove_redundant` and `snaf.proteomics.compare_two_fasta` function. The `snaf.proteomics.compare_two_fasta` function removes all redundant short peptides from the prediction, whereas the `snaf.proteomics.remove_redundant` function discards splicing neoantigens overlapping with non-diseased human proteome sequences (UniProt). The custom splicing neoantigen database (FASTA files) used for Peptide Spectrum Matching (PSM) search is available in synapse ([https <colon slash slash> www <dot> synapse <dot> org <slash> #!synapse:syn32057176](https://www.synapse.org/#!synapse:syn32057176)), and the MaxQuant configuration files `mqpal.xml` and the code SNAF uses to automatically generate the xml files are available in GitHub ([https <colon slash slash> github <dot> com <slash> frankligy <SNAF> reproduce](https://github.com/frankligy/SNAF)).

[00416] The configuration file was created using the MaxQuant Windows Desktop application and modified using `snaf.proteomics.set_maxquant_configuration` function. Specifically, MaxQuant was run using the custom splicing neoantigen database as described above. Key parameters are as following: `enzyme=None`, `enzyme_mode=5` (no digestion), `protein_fdr=1`, `peptide_fdr=0.05`, `site_fdr=1`, `minPepLen=8`, `minPeptideLengthForUnspecificSearch=8`, `maxPeptideLengthForUnspecificSearch=25`. Other parameters are the default value automatically set by the program.

[00417] Raw MS sequencing data from 25 independent melanoma patients was obtained from [ftp <dot> ebi <dot> ac <dot> uk <slash> pride-archive <slash> 2017 <slash> 04 <slash> PXD004894](ftp://ebi.ac.uk/pride-archive/2017/04/PXD004894) and analyzed using MaxQuant as described above. Patient 16 (Mel-16) was excluded due to potentially corrupted raw data.

[00418] The custom splicing neoantigen database (FASTA files) used for Peptide Spectrum Matching (PSM) search is available in synapse ([https <colon slash slash> www <dot> synapse <dot> org <slash> #!Synapse:syn32057176](https://www.synapse.org/#!Synapse:syn32057176)). In brief, all observed shared splicing neoantigens (n=114) and unique (patient-specific) neoantigens were combined into a single

combined database, to ensure these two sets had an equal chance to be detected. Here, neoantigens which correspond to peptides in UniProt (non-diseased) were eliminated, and only non-redundant peptides (from different junctions) were reported.

Peptide Synthesis and MS Spike-In Aalidation

[00419] The 36 peptide candidates were synthesized (GenScript, Piscataway, NJ) at minimum 70% purity with an average yield of 0.2-0.5 mg. Peptides were reconstituted with water to a final stock concentration of 1 pmol/ μ L. Peptides were pooled (except for LELLVKGTV and STLEFGLRV, which did not solubilize sufficiently) at a concentration of 1 pmol/ μ L and then diluted 1:10 for a 100 fmol/ μ L working solution. LC-MS analysis was performed on a 50 fmol injection of pooled peptides using a Ultimate 3000 nanoflow HPLC (Dionex) and Orbitrap Eclipse Tribrid mass spectrometer (ThermoFisher Scientific) as described below.

[00420] Injections were loaded onto an Acclaim PepMap 100 trap column (300 μ m x 5 mm x 5 μ m C18) and gradient-eluted from an Acclaim PepMap 100 analytical column (75 μ m x 25 cm, 3 μ m C18) equilibrated in 96% solvent A (0.1% formic acid in water) and 4% solvent B (80% acetonitrile in 0.1% formic acid). The peptides were eluted at 300 nL/min using the following gradient: 4% B from 0-5 min, 4 to 10% B from 5-10 min, 10-35% B from 10-60 min, 35-55% B from 60-70 min, 55-90% B from 70-71 min, 90% B from 71-73 min, 90-4%B from 73-74 min and 4% B from 74-90 min.

[00421] The Orbitrap Eclipse was operated in positive ion mode with 2.0 kV at the spray source, RF lens at 30% and data dependent MS/MS acquisition with XCalibur version 4.3.73.11. Positive ion Full MS scans were acquired in the Orbitrap from 375-1500 m/z with 120,000 resolution. Data dependent selection of precursor ions was performed in Cycle Time mode, with 3 seconds in between Master Scans, using an intensity threshold of 2×10^4 ion counts and applying dynamic exclusion (n=1 scans within 30 seconds for an exclusion duration of 60 seconds and +/- 10 ppm mass tolerance). Monoisotopic peak determination was applied and charge states 2-6 were included for HCD MS2 scans (quadrupole isolation mode; 1.6 m/z isolation window, Normalized collision energy at 30%). The resulting fragments were detected in the Orbitrap at 15,000 resolution with Standard AGC target and Dynamic maximum injection time mode.

[00422] The MS2 raw file was imported into Skyline v.22.2.0.351 and a template for synthetic peptide sequences was applied to visualize MS2 spectra. The ion intensities for MS2 b- and y- fragment ions were plotted for each synthetic peptide sequence. Peptides with coverage for at least six fragment ions were considered “observed”. The Skyline software enables visualization of the observed fragment ions in the context of the entire raw MS2 spectrum via color-coding and labeling of the b- and –y ions. The synthetic peptide MS2 spectra were then compared to experimental spectra which were identified by search algorithm software to have the same peptide sequence. The b- and y- ions identified from both the synthetic and experimental spectrums were selected, and the quantitative correspondence were calculated based on their associated ion densities. Particularly, denoting each fragment ions in synthetic spectrum as x and the corresponding ions in experimental spectrum as y , with the total number of I , and the associated relative abundance is dx_i and dy_i . Pearson correlation (r) and the cosine similarity ($\cos$) was computed as below.

$$r = \frac{\sum (x_{ai} - \bar{x}_{ai})(y_{ai} - \bar{y}_{ai})}{\sqrt{\sum (x_{ai} - \bar{x}_{ai})^2 \sum (y_{ai} - \bar{y}_{ai})^2}} \quad (10)$$

$$\cos = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2} \sqrt{\sum y_i^2}} \quad (11)$$

where X and Y are the vector representation of the ion relative density. The p-value was reported as the probability of observing a higher correlation of randomly drawn X and Y from two independent normal distribution (zero correlation). The analysis used $r > 0.70$, $p \leq 0.05$ and $\cos > 0.9$ as a rough threshold instead of a hard criterion to evaluate the similarity of observed spectrums.

[00423] Amongst all the MS2 spectrums that were assigned to a splicing neoantigen by the program ($FDR < 0.05$), the matched ions and the associated ion density were manually inspected. A series of rules were applied to rule out potential false positive match and only retain high-confidence peptides for MS spike-in experimental validation. Criteria:

- [1] At least 6 matched fragment ions;
- [2] No more than 25% unassigned ions are greater than 10% relative abundance;

[3] Fragment ions make up to at least 20% of total ions in the spectrum;

[4] Presence of sequential fragment ions (i.e. y4-y5-y6); and

[5] y-ions ending in proline were higher density than other ions.

Peptide Spectrum matchings (PSM) that meet all the of the above were retained as high-confident peptides for further experimental validation using synthetic sequence.

Validation of Peptide-MHC Binding by MHC Stabilization Assay

[00424] To test MHC-I binding of synthesized neoantigen peptides, TAP deficient T2 cells that are defective in transporters were used for endogenous peptide loading. T2 cells were obtained from ATCC and grown at 37°C, 5% CO₂ in Iscove's Modified Dulbecco's Medium supplemented with 20% FBS and pen/strep. 1x10⁵ T2 cells were used without and loaded with 100ug/ml peptides and thereafter incubated overnight. All T2 cells were harvested and stained with HLA-A2-PE antibody (clone BB7.2; Biolegend) for 30 mins on ice. Cells were washed once with cell culture medium and acquired on a Fortessa II flow cytometer. Median fluorescence intensity was determined using FlowJo software.

Immunogenicity Assay

[00425] Immunogenicity of predicted splicing neoantigen peptides was determined as described previously. In short, HLA-typed PBMCs from leukocyte reduction system (LRS) chambers were isolated using Ficoll hypaque density gradient centrifugation, aliquoted in 20x10⁶ cells per vial and frozen in liquid nitrogen until use. At day 0, PBMC were thawed and used to set up monocyte derived dendritic cells by plating 4x10⁶ PBMC in a 24 well. Cells were incubated at 37°C, 5% CO₂ in DC medium (RPMI 1640 supplemented with 10% FBS, 1% L-Glutamine (200mM) + IL-4 (1000 U/ml) and GM-CSF (800U/ml). After 4h, the non-adherent fraction was removed by rinsing the wells twice with PBS. Adherent cells were cultured for 7 days in DC medium. On day 7, DCs were loaded with 10ug/ml peptides dissolved in DC medium and incubated at 37°C, 5% CO₂. After 4h, 1.5ml DC maturation medium was added (RPMI 1640 supplemented with 10% FBS, 1% L-Glutamine, IL-4 (1000U/ml), GM-CSF (800U/ml), IL-1 β (10ng/ml), IL-6 (10ng/ml), TNF- α (10ng/ml) and LPS (30ng/ml)). After 16h of DC maturation, peptide loaded DCs were used to stimulate autologous PBMC, by adding 1x10⁶ PBMC of the same donor to the DC cultures. DC and PBMC co-cultures were grown in T cell medium (60% RPMI 1640, 40% Click's medium supplemented with 10% FBS, 1% L-glutamine, IL-6 (100ng/ml), IL-7 (10ng/ml), IL-12 (10ng/ml), and IL-15 (5ng/ml). Medium was changed on day

3 and day 6 based on medium color change. On day 14 and 21 T cells were harvested and stimulated with new autologous peptide loaded DCs.

[00426] After three rounds of T cell priming, at day 28, T cells were harvested and tested for their IFN γ response to peptide loaded 721.221 (221) single HLA antigen target cells. First the target cells 221.A*02:01, 221.C*04:01, and 221.C*08:01 were loaded with relevant peptides at 10 μ g/ml in separate wells. 221 and peptides were incubated for 4h at 37°C, 5% CO₂ in RPMI 1640 supplemented with 10% FBS. Peptide loaded 221 target cells with and without peptides were co-cultured with primed T cells at 3:1 ratio in the presence of 50 ng/ml PMA for 6h at 37°C, 5% CO₂ in RPMI 1640 supplemented with 10% FBS and monensin. PMA and ionomycin stimulation (each at 1 μ g/ml) was used at positive control. Thereafter the cells were harvested and stained for extracellular CD45-BV786 (Biolegend), CD14-PerCP (Biolegend) and CD8-PE (biolegend) for 30 mins on ice. Cells were fixed and permeabilized using the CytoFix/CytoPerm kit (BD) according to manufacturer instructions and thereafter stained for intracellular IFN γ -APC expression (clone 4S.B3; Biolegend) for 20 mins on ice and directly analyzed on a BD Fortessa flow cytometer. Analysis of CD45+CD14-CD8+ IFN γ + cells was determined using FlowJo software.

Validation of ExNeoEpitope Localization

[00427] To confirm the expression and cell membrane localization of SNAF-B novel isoforms, the long-read sequencing evidenced alternative isoforms and their annotated reference isoforms that are C-terminal tagged with either mNeon-Green or eGFP on a plasmid vector (VectorBuilder, USA). Streak LB agar plates with 100 μ g/mL Ampicillin were made for each isoform. A single colony was picked from each plate and expanded in 1 mL of LB broth for 8 hours at 37 Celsius respectively. 20 μ L of the pre-expansion broth was then taken and pipetted into a Erlenmeyer flask with 50 mL of LB broth. The competent cells were expanded overnight at 37 Celsius. Midi-preps were performed for each construct with the ZymoPure II plasmid midiprep kit. On an Ibidi 4-well chamber μ -slide, HEK-293T cells were seeded in prior with a concentration of 0.15 M/mL. When HEK-293T cells reached 60% of confluency, these constructs (CMV promoter) were transfected into the cells separately with 1 μ g of plasmid (99% pUC19 negative control plasmid+1% engineered plasmid) with TransIT-LT1 (Mirus) following the manufacturer's protocol.

[00428] The cells were fixed and permeabilized with 4% PFA 24 hours post-transfection. After two rounds of washing, the cells were treated with a warm 1x citrate buffer (diluted from 10X stock; Sigma-Aldrich) to break the protein cross-links. The cells were then washed once with PBS again and a membrane actin stain was performed with 1x Phalloidin 647 in PBS with 1% BSA (Abcam) at room temperature for an hour. To stain the nuclei of the fixed cells, the cells were washed with PBS and stained with DAPI (Thermo Scientific) (1:4000 diluted in PBS with 1% BSA) at room temperature for 5 minutes. The stained cells were immediately washed with PBS, and the PBS removed by vacuum. Prolong gold mounting media (Thermo Fisher Scientific) was evenly applied to the fixed cells surface. 24 hours post transfection, the expression and co-localization of the reference and alternative isoforms were assessed by a confocal spinning disk microscopy (Yokogawa SoRa W1 dual camera system) using the Nikon Elements software. The confocal raw imaging files with all z-stacks are provided in Synapse (<https://www.synapse.org/#!Synapse:syn52063953>).

Melanoma Cell-Line Long-Read Isoform Sequencing

[00429] For long-read mRNA isoform sequencing, cells were grown and isolated from five independent Melanoma cell lines: A375, SKMEL, MeWo, UACC62, and UACC257 (ATCC). Total RNA was isolated (Trizol) and analyzed on a Thermo Nanodrop UV-Vis and an Agilent Bioanalyzer to confirm the nominal concentration and ensure RNA integrity. From the RNA, cDNA was synthesized using the Clontech SMARTer cDNA Synthesis Kit, in which a barcode was added to the oligo-dT at the 3' end. Each melanoma cell line cDNA was pooled and then converted into a SMRTbell library using the Iso-Seq Express Kit SMRT Bell Express Template prep kit 2.0 (Pacific Biosciences). Each library was sequenced on a SMRT cell on the Sequel II system using a 30 hour movie collection time. The “ccs” command from the PacBio SMRTLink suite (SMRTLink version 9) was used to convert Raw reads into Circular Consensus Sequence (CCS) reads. The resulting data was analyzed in SQUANTI to assign reads to full-length collapsed reference or novel isoforms. The isoform GTF files, barcode sequences and raw data are available in Synapse (<https://www.synapse.org/#!Synapse:syn32057176>). Results are shown in Table 1.

Melanoma RNA-Seq Analyses

[00430] RNA-Seq paired-end BAM files from 472 Melanoma patients collected by TCGA (SKCM) were obtained from the GDC portal following dbGAP approval (phs000178.v10.p8). A second collection of 40 RNA-Seq FASTQ files from patients who underwent immunotherapy (archival formalin-fixed, paraffin-embedded) in Van Allen cohort (33) files were obtained from the dbGAP database (phs000452.v2.p1). These FASTQ files were aligned to the reference human genome (hg38) and transcriptome (Ensembl 91) using STAR. Among these 40 Van Allen RNA-Seq samples, patient 41 was excluded (only partial sequencing data available), as previously reported. The HLA genotype of each patient sample was determined from the RNA-Seq FASTQ files using the software Optitype 1.3.3. Optitype was chosen based on its superior performance on calling HLA-I alleles (over 99% accuracy) from RNA-Seq data based on recent large-scale benchmark, evaluated on “gold-standard” HLA genotyping data.

[00431] AltAnalyze v 2.1.4 was used to quantify splicing independently in these two cohorts using the Ensembl version 91 database. MultiPath-PSI identified splicing events were used as inputs for SNAF. The TCGA survival and mutation data were downloaded from Xena Browser. Survival analysis was performed using the `snaf.survival_analysis` function with stratification argument `n=2` (high burden is equivalent to $>$ the median burden, low burden $<$ the median burden). Mutation analysis was conducted using `snaf.mutation_analysis` function.

[00432] To identify individual neojunctions or neoantigens associated with survival, a univariate Cox Regression analysis was used to identify events/antigens with a parental PSI value that are positively or negatively associated with patient outcome. Here, the neojunction and its parental PSI value is ignored if the neoantigen was not predicted to be presented in that sample, resulting in different survival associations for different neoantigens produced from the same neojunction. For analysis of Melanoma RNA-Seq TCGA samples in the SNAF-B workflow, long-read Iso-Seq cDNA sequences were obtained from pan-cancer cell line sequencing, using the PacBio provided isoform GTF file (<https://downloads.pacbcloud.com/public/dataset/UHRRisoseq2021/Final-MappedTranscripts>).

Normal Tissue Reference RNA-Seq

[00433] A subset of represented samples (n=1,215 samples) for all GTEx healthy tissues were obtained as RNA-Seq FASTQ files from dbGAP (phs000424.v8.p2). These FASTQ files were aligned to the reference human genome (hg38) and transcriptome (Ensembl 91) using STAR prior to analysis in SNAF.

Selective Amino Acid Motif Enrichment Analysis

[00434] The shared and unique neoantigens were defined by accounting for the frequency of their parental splicing junction. The occurrence frequency of a neoantigen were denoted as f_{na} and its parental splicing junction (neojunction) as f_{nj} . By default, a splicing neoantigen was classified as shared if:

$$f_{na} > 0.9 * f_{nj} \quad (f_{nj} > 0.1) \quad (12)$$

Instead, a splicing neoantigen was classified as unique if:

$$f_{na} < 0.1 * f_{nj} \quad (f_{nj} > 0.1) \quad (13)$$

[00435] Each amino acid in the neoantigens was encoded into the numerical vector of length 12 using a reduced representation of over 500 physicochemical properties derived from the AAindex database. A detailed description of the feature selection criteria and dimension reduction procedures is provided in DeepImmuno. Briefly speaking, each amino acid was associated with over 500 numerical properties. A Principal Component Analysis (PCA) was performed to reduce the dimension to 12, which was further used for encoding the peptide sequence.

[00436] The encoded neoantigens were embedded into a 2D space using Uniform Manifold Approximation and Projection (UMAP). The most statistically significant motifs were derived using the MEME suite differential enrichment mode on the official website, the minimum and maximum length of the motif were set to 9 to reflect the queried 9mer sequences. The plots were produced in Python or Prism version 9.

SNAF Interactive Web Application

[00437] An interactive web application for both SNAF-T and SNAF-B visualization results was created using the Python Dash framework. The front-end was written in React.js and the backend in the Flask framework. Plotly-py was used for interactive visualization and the

Ensembl REST API used for retrieving transcript reference sequences and Needleman-Wunsch global alignment. The peptide weblogo plot was implemented. For each position i , the height of the stack (i) represents the difference between the maximum entropy and the actual entropy among the observed amino acids distribution.

$$\begin{aligned} Height_i &= Entropy_{max} - Entropy_{obs} \\ &= \log_2 20 - \left(- \sum_{n=1}^{20} p_n \log_2 p_n \right) \end{aligned} \quad (14)$$

[00438] Where n denotes one of the 20 amino acids and p_n denotes the observed frequency of amino acids n at position i . Within each stack, the height of each amino acid was determined by their relative frequencies.

AlphaFold2 3D modeling

[00439] AlphaFold2 was obtained from <https://github.com/deepmind/alphafold> and implemented as a Singularity container from the provided Dockerfile, to be run on a Linux HPC, along with all training databases. For the indicated examples, the ranked_0.pdb was chosen as the 3D modeling result, and the protein cartoon visualization and coloring were performed using PyMol software.

EXAMPLE 2

SNAF Computational Workflows – Inferring New Classes of Neoepitopes from RNA-Seq

[00440] The SNAF workflow begins with user-supplied BAM files from tumor samples or cancer cell lines, followed by the identification and quantification of diverse classes of post-transcriptional regulation. In particular, the workflow applies a highly accurate approach for local splicing variation (MultiPath-PSI) from the AltAnalyze framework, to detect known and novel alternative splicing (cassette exon, 3'/5' splice site exon, intron retention, alternative terminal exon, trans-splicing) and alternative promoter regulatory events, which can produce unique exon-exon or exon-intron junctions for *in silico* translation (Figure 3).

[00441] This approach has been extensively benchmarked against diverse local-splicing variation (LSV) approaches, including rMATs, Leafcutter and MAJIQ, with specialized methods to accurately quantify retained introns (Figure 4). The produced splice-junction/sample

count matrix is compared against a MultiPath-PSI pre-processed database of normal human healthy tissues (GTEx and TCGA) to identify those that are tumor-specific (23) (Figure 5). Tumor-specific splice junctions can be analyzed in parallel with SNAF-T and SNAF-B.

[00442] SNAF-T consists of:

- 1) HLA type prediction from sample FASTQ files (user provided);
- 2) *in-silico* translation;
- 3) MHC-binding prediction (NetMHCpan or embedded calls to MHCflurry);

and

- 4) HLA-allele specific immunogenicity prediction (DeepImmuno).

[00443] SNAF-B consists of:

- 1) full-length isoform prediction for each tumor-specific splice-junction by augmenting existing isoform references;
- 2) exclusion of isoforms predicted to induce nonsense-mediated decay (NMD);
- 3) transmembrane topology prediction; and
- 4) long-read isoform sequence validation and augmented prediction (optional).

[00444] For both of the SNAF-T and SNAF-B workflows, a Maximum Likelihood Estimation and separate hierarchical Bayesian model (BayesTS) can optionally be applied to assess the tumor specificity of each neojunction, e.g. relative to GTEx, in SNAF's default and an optional custom healthy tissue reference RNA-Seq data, with custom tissue weighting assigned by the user. Normal healthy tissue references can include but are not limited to 54 tissue regions from the GTEx consortium, human fetal development through adult independent pan-tissue bulk RNA-Seq cohorts, single-cell RNA-Seq of diverse human tissues, and long-read isoform sequencing from normal healthy tissues (e.g., GTEx). Finally, to identify causal regulators of splicing neoantigen production, RNA-SPRINT (RNA-based Splicing PRotein activity INference from multivariate decision Trees) was developed to infer splicing factor activity directly from tumor RNA-Seq splicing profiles (see Example 1).

[00445] The workflows described herein are unique in both design and functionality (see Table 2). Unlike prior T cell-based splicing neoantigen prediction approaches, SNAF is fully automated, supports any human genome version, has an embedded database of healthy

reference RNA-Seq profiles (such as, for example, GTEx and fetal human development), identifies intron retention associated antigens, and enables more accurate prediction of immunogenicity. As the program has a modular design with well-described Python classes, it can be extensively customized to incorporate additional reference datasets for verification (e.g., control RNA-Seq, long-read sequencing) and alternative algorithms (e.g., MHC binding prediction).

[00446] While existing splice-neoantigen methods only consider T cell antigens, SNAF-B provides independent evidence of tumor-specific transmembrane proteins (ExNeoEpitopes) that can uncover novel extracellular epitopes for antibody recognition. SNAF predicts ExNeoEpitopes from a reference mRNA transcript database through the removal of alternative spliced-out transcript sequences in the longest reference model and through matching of tumor-specific splice-junctions to large long-read isoform sequencing experiments or repositied expressed sequence tags (ESTs).

[00447] SNAF automates the analysis of survival within a cancer patient cohort to prioritize neoantigens for further evaluation. SNAF further automates comparative proteomics analysis of produced candidate splicing neoantigen peptide sequences to produced mass-spectrum profiling datasets, such as immunopeptidomics and surface proteomics datasets, to confirm neoantigen cell-surface or MHC presentation. Such predictions are used to derive a ranking of splicing neoantigen importance, in conjunction with other SNAF produced outputs, including, but not limited to, immunogenicity score, HLA-binding score, neoantigen tumor specificity scores relative to healthy control samples (maximum likelihood Estimation, Bayesian hierarchical model, control cohort mean counts), alternative splicing PSI value, mRNA junction occurrence in the reference transcriptome (Ensembl version 91), patient cohort occurrence frequency, recurrence in independent cancers and cohorts, splice-event type, full-length peptide prediction length versus the reference, peptide sequence similarity to reference peptides, the presence of signal peptide sequence, secondary structure, 3D protein structure prediction, insertion versus deletion of amino acids, absolute junction read counts, topology predictions and other programmatic outputs.

[00448] SNAF can identify both precision (patient specific) and shared neoantigens, which are frequently presented by MHC and are immunogenic or which produce ExNeoEpitopes. The present inventors have found that shared splicing neoantigens can be

distinguished from patient-specific neoantigens in their amino acid sequence and corresponding physicochemical parameters, further enabling systemic prioritization.

[00449] SNAF has a modular design with well-described Python classes, enabling extensive customization to incorporate additional reference datasets for verification (e.g., custom control RNA-Seq, long-read sequencing) and alternative algorithms (e.g., MHC binding prediction). Further, interactive visualization tools enable the user to explore sequence features shared among common versus unique neoantigens or explore background signals in specific GTEx tissues (outlier junction reads) on a case-by-case basis, following prioritization.

[00450] A central component of the SNAF prediction workflow is the ability to more accurately predict which splicing neoantigens will selectively induce a T-cell response based on which HLA alleles are present in a given patient specimen. As previously demonstrated, the DeepImmuno workflow can predict immunogenic tumor neoantigens with up-to a 2-fold greater sensitivity than alternative approaches to assess antigen immunogenicity. This workflow has been independently validated as a top-performer in predicting immunogenicity versus other publicly available tools. Unpublished and commercial analytical platforms also designed to identify splicing neoantigens for T-cell and B-cell based immunotherapies include SpliceCore and IRIS, but such tools are limited in their features and capabilities, which limits their utility (see Table 2).

Table 2. Comparison of features in SNAF to other published splicing neoantigen workflows.

	SNAF	IRIS	ASNEO	NeoSplice
Automated Program Features				
in-silico translation				
MHC binding predictions				
Tumor-specific expression				
Immunogenicity predictions				
Surface antigen predictions				
Tumor specificity score				
Custom normal tissue reference				
Long-read validation				
Differential gene expression				
Gene-set enrichment				
Splicing factor activity prediction				
Advanced visualization				
Integrated survival analysis				
Stand-alone python module				
Interactive web app				
Parallization				
Interface to proteomics analysis				

EXAMPLE 3

Preliminary Evaluation of the Computational Workflows – Validation of Predicted MHC-Bound Neoantigens

[00451] The DeepImmuno workflow can predict immunogenic tumor neoantigens with up-to a 2-fold greater sensitivity than alternative approaches. To determine whether SNAF-T can identify bonafide MHC-presented splicing neoantigens, two prior produced cancer immunopeptidome datasets to validate its predictions were selected. First, bulk RNA-Seq (single-end) was evaluated and immunopeptidome profiling data (HLA-bound peptides) was matched from 14 ovarian cancer patients (27). To expand these predictions, SNAF-T was applied to skin cutaneous melanoma (SKCM) biopsy RNA-Seq from The Cancer Genome Atlas (TCGA) initiative (n=472), and unmatched immunopeptidome data from 24 melanoma patients, herein referred to as the Bassani Sternberg cohort (28) (Figure 6A). In ovarian tumors, SNAF identified 46 splicing neoantigens on average per sample detected by MS as MHC presented in all 14 patients (7.40% of all detected), ranging from 12 to 160 (Figure 6B).

[00452] This number is expected to be an underestimate, as HLA-bound peptides arise from non-specific protease cleavage (which alters the resultant MS spectra), and hence traditional tryptic-based search engines cannot confidently recover all neoantigens. Here, the absolute number of MS-supported splicing neoantigens is higher than previously reported (average 2 peptides per sample) using untargeted proteome data (CPTAC), indicating increased sensitivity of targeted immunopeptidome for identifying valid and rare neoantigens.

[00453] As MS-based predictions are subject to inherent type-1 errors, select ovarian and melanoma SNAF-T neoantigens were validated using targeted MS on synthesized neopeptides. Specifically, 14 ovarian and 22 melanoma peptides were selected with high-confidence spectra (see Example 1). Of the 36 tested, 27 were sufficiently detected by MS. Comparison of the synthetic to the original MS2 spectra found 11 matches, of which 7 were high scoring based on all match criteria (see Example 1).

[00454] These seven neoantigens derive from multiple mechanisms, including known cassette exons, alternative 3' and 5' splice sites, intron retention and novel cassette exons (AltAnalyze defined). For example, the shared melanoma splicing neoantigen HAAASFETL in the gene FCRLA occurs due to a known cassette exon-exon junction in an isoform that is weakly detected in blood and spleen (average read count = 0.51, BayesTS: 0.03) (Figure 6C).

[00455] Expression of *FCRLA*, which is a member of the F-receptor-like immunoglobulins, is correlated with good prognosis in Melanoma. Interestingly, *FCRLA* gene expression is only weakly tumor-specific (BayesTS: 0.13). However, this junction is detected in >34% of TCGA melanoma patients (162/472, Average read count = 52.91) (Figure 6D). The resultant MS2 spectrum of this antigen has a high-confidence match (Andromeda score: 149.06, P-value: 0.0), with a synthetic spectrum Pearsonr similarity of 0.63 (p-value: 0.05, cosine similarity: 0.87). The other MS2 confirmed neoantigens derive from ubiquitin protein ligase complex (*FBXO7*), asparagine amidase (*NGLY1*), cytoskeletal motor protein (*DYNLT5*), negative regulator of RAS signaling (*RASA3*) and currently uncharacterized protein coding genes (*C20orf204*, *C6orf52*) (Figure 7). Lower spectrum similarity scores are observed in the Ovarian cohort (Pearsonr 0.55) compared to the Melanoma cohort (Pearsonr: 0.84), which can be attributed to differences in the MS technology applied (Ion trap MS versus Fourier transform MS) and fragmentation methods (CID versus HID).

EXAMPLE 4

Splicing Neoantigen Burden Predicts Overall Survival and Response to Immunotherapy

[00456] To broadly assess the clinical relevance of splicing neoantigens, SNAF was applied to >500 melanoma patient biopsies with (i.e., Van Allen cohort) (33) and without immunotherapy (TCGA-SKCM). These analyses separately considered the number of predicted neojunctions, MHC-bound peptides, immunogenic neoantigens and overall neoantigen burden, considering RNA-Seq determined HLA alleles for each patient.

[00457] In the TCGA cohort of 472 samples, 528 tumor-specific splicing junctions (neojunctions) were found per patient on average, ranging from 28 to 1,549 neojunctions in each patient. From these neojunctions, an average of 1,090 MHC-bound peptides were predicted, ranging from 75-2,981 peptides per patient. DeepImmuno predictions reduced the number of neoantigens to 915 on average (ranging from 74-2,486/patient), filtering out 16% of potentially non-immunogenic bound peptides, as current immunogenicity approaches still suffer from low precision, which may explain the relatively low filtering rates from netMHCpan-reported bound peptides. To only prioritize the most high-confident immunogenic neoantigens for downstream experimental validation, SNAF further allows users to increase the default immunogenicity cutoffs (i.e. Deepimmuno default cutoff = 0.5) and supply external immunogenicity predictions (i.e. IEDB, DeepHLApan).

[00458] To investigate the relationship between splicing neoantigen burden and clinical outcome, survival analyses were performed on both the TCGA and Van Allen cohorts. This analysis surprisingly found that patients in TCGA with high MHC-bound neoantigen burden trends towards poor overall survival (log rank $p < 0.05$) (Figure 8A). Conversely, the analysis found that patients with a high neoantigen burden that received ICB (Van Allen cohort) have not statistically significant but improved overall survival (log rank $p = 0.18$). These trends were also reflected in neojunction and immunogenic peptide burden (Figure 8B). Since the bulk RNA-Seq data profiled in these cohorts was obtained prior to therapy, one plausible explanation is that patients with high neoantigen burden exhibit frequent tumor immune escape, which is overcome by CTLA-4 inhibition.

[00459] Examination of differential gene expression in TCGA patients with high splicing neoantigen burden versus low found 597 up- and 227 down-regulated genes (fold > 1.5 and eBayes t-test $p < 0.05$, FDR) (Figure 8C). Regarding the survival results, the top most-differentially up-regulated genes were those previously implicated in poor response to immunotherapy, immune evasion and tumor invasion (e.g., *ADAM10*, *PTPN11*, *TGFBRI*, *TNPO1*, *ANKRD52*, *MIB1*, *KIF3B*).

[00460] In contrast, down-regulated genes were markers of active tumor-immune cell infiltration, in particular plasma cells and T-cells (Figure 8C) with corresponding enrichment of this immune infiltration signature, along with genes involved in antigen binding and complement activation (Figure 8D). Enrichment of this immune infiltration signature was observed in the low-burden patients by gene-set enrichment.

[00461] Enrichment analysis of high burden induced transcripts identified upregulation of p53 signaling, mitotic cycle-cycle, cell motility and signaling MET, among others (Figure 9A). Hence, patients with splicing neoantigen burden express genes that promote immune evasion while low burden patients show evidence of immunoreactivity.

[00462] Since both chemotherapy and radiotherapy target cancer cells via directly inducing DNA damage or reactive oxygen species, reprogramming of metabolic pathways and increased DNA repair machinery have been recognized as possible mechanisms of chemo- and radio-therapy resistance. Combination radiotherapy and immunotherapy have been proposed as a strategy to overcome single therapy drug resistance, which aligns with the observation of elevated immune evasion genes in high-burden groups. Thus, high splicing neoantigen patients

(TCGA) can represent chemo- and radio-therapy resistant tumors with upregulated immune evasion capacity, representing candidates for combination therapy. In contrast, low neoantigen burden patients evidence immune reactivity, which can partially explain their favorable prognosis.

EXAMPLE 5

CAMKK2 Mutations Correlate with Increased Neoantigen Burden in Melanoma

[00463] To determine whether splicing neoantigen burden in melanoma is associated with specific mutations, mutations in high- versus low-burden neoantigen patients were compared. While no individual mutations were significant based on FDR correction, one of the most significant hits relates to mutations in the gene CAMKK2 versus wild type. Specifically comparing patients with and without CAMKK2 mutations, a significantly higher neoantigen burden was found with mutations (Mann Whitney test $p=0.0001$) (Figure 8E). Out of 19 patients with reported CAMKK2 mutations, 18 were found to occur in the high burden group, out of a total of 222 patients.

[00464] CAMKK2 has been previously identified as a target to improve immunotherapy outcomes, as its mutation or inhibition leads to increased anti-PD1 immunotherapy efficacy. This effect is believed to occur via CAMKK2's ability to negatively regulate ferroptosis, a mechanism of cell death that is induced by iron-dependent lipid peroxidation, via the AMPK–NRF2 pathway. These results indicate that CAMKK2 mutations can contribute to splicing neoantigen burden and immunotherapy outcome, either through direct or indirect splicing regulatory networks, in a subset of melanoma patients.

EXAMPLE 6

Individual Splicing Neoantigens can Predict Response to Immunotherapy

[00465] In addition to neoantigen burden associations with survival, the individual PSI profiles of SNAF immunogenic neoantigens were compared with patient survival. This analysis identified 2,970 parental junctions associated with poor overall survival in TCGA (likelihood-ratio test (LRT) $p<0.05$ and $z \geq 1$) (Figure 8F). Among these neoantigens, it was found that a subset ($n=108$ junctions) was present in over 15% of patients (shared neoantigens), indicating that these represent novel survival biomarkers. Although a much smaller number of neoantigens was associated with survival in the Van Allen cohort due to the limited sample size, the analysis

found 1,755 poor and 227 good prognosis associated neoantigen junctions following CTLA-4 inhibitor treatment (LRT $p < 0.05$ and $z \geq 1$ or $z \leq -1$, respectively).

[00466] Importantly, 7 unique neoantigen junctions in TCGA and Van Allen were found to be associated with opposite survival associations (poor in TCGA and good in Van Allen) and present in >10 TCGA patients (Figure 8F). For example, a splicing neoantigen in the melanin synthesis associated glucose transporter SLC45A2 (TEFQTRRAM) was detected in 212/472 TCGA SKCM patients; this neoantigen is associated with poor overall survival in TCGA (Wald $p < 0.05$, $z\text{-score} > 2$) and is associated with good overall survival in the Van Allen cohort (Wald $p < 0.05$, $z\text{-score} < -2$), indicating that it represents a relatively common target for therapy and prognosis (Figure 8F).

[00467] Other splicing neoantigens were associated with poor overall survival in TCGA; these were identified in the majority of patients in both cohorts (e.g., SLC45A2 - FQTRRAMTL). Other overlapping neoantigens produced from this same junction were present in $>64\%$ patients in both melanoma cohorts (FQTRRAMTL), with differences in their abundances due to varying MHC-I allele binding and immunogenicity preferences. Dozens of other shared neoantigens were associated with poor prognosis in TCGA and trending towards good prognosis in the Van Allen cohorts. The occurrence and tumor specificity of these and other similar neojunctions were readily confirmed by IGV visualization, and frequently associated with alternative splice sites in conjunction with tumor-specific gene expression (Figure 9B). These data find that individual shared splicing neoantigens can predict outcomes in patients prior to and after ICB.

EXAMPLE 6

Disruptive splicing factor activity underlies high splicing neoantigen burden

[00468] The coordinated regulation of alternative splicing and alternative promoters are orchestrated by the expression or activity of conserved cis- and trans-regulatory interactions (e.g., splicing factor binding to RNA recognition elements). Splicing neoantigens are likely produced by modulation of such interactions, driven by direct mutations or RNA-editing that result in unique tumor-specific mRNAs.

[00469] The most likely mediator of differential splicing neoantigen burden is a change in the expression or activity of one or more splicing factors. Examination of differential gene expression in the high versus low splicing neoantigen group found a large number of

upregulated splicing regulators (n=20) (Figure 8C). These included 117 mRNA splicing regulators, upregulated with a fold > 1.2 and eBayes t-test $p < 0.05$, FDR. Patients with splicing factor mutations in SF3B1 were further enriched in the high-burden group (Mann Whitney $p = 0.02$).

[00470] This surprising finding led to an investigation of whether splicing factor activity was also increased with burden. Existing methods for transcription regulatory network inference, which rely on a defined set of target genes or regulons, were used to infer splicing factor activity. The up- and down-regulation of these regulons can indirectly provide evidence of TF activity. A large dataset of RNA-Binding Proteins (RBP) knockdowns (n=191) in K562 cells from the ENCODE project (see Example 1) was used to establish a reliable link between splicing factors and downstream splicing junctions. Initial splicing regulons were derived by observing the changes in splicing events upon the knockdown of specific RBP and accounting for batch effects. To refine these regulons, evidence of direct RBP-target regulation was incorporated using eCLIP sequencing from 120 in the K562 cell line. The resulting data were used to construct a prior network to infer splicing factor activity (see Example 1, Figure 10A).

[00471] Multivariate Decision Tree (MDT) was found to produce the most accurate RBP activity predictions, compared to 11 other highly-used transcription factor (TF) activity inference methods, when considering 116 RBP knockdown splicing profiles from HepG2 cells (ENCODE) (Figure 11A-C). Application of this approach, called RNA-SPRINT, to all TCGA melanoma patient PSI profiles, finds that 209 out of 221 examined RBPs have reduced versus increased activity in the high versus low splicing neoantigen group (Mann-Whitney $p < 0.05$) (Figure 10B-10C).

[00472] While contrary to expectations based on RBP gene expression, these data indicate that there exists a coordinated failure to properly splice transcripts, as previously described. This result is further supported by an observed increase in intron-retention in the high-burden patients (failure to excise introns) (Figure 10D). It is also possible that broad upregulation of splicing factors is a compensatory mechanism to counteract global splicing defects or simply a byproduct of cell-type heterogeneity within the tumor (e.g., immune infiltration).

EXAMPLE 7

Shared splicing neoantigens are presented by MHC and display amino acid compositional bias

[00473] As an important observation from the SNAF-T analyses is the co-occurrence of splicing neoantigens among many patients, the study examined broadly shared splicing neoantigens observed in melanoma, versus those observed in individual patients, to understand such regulation. Intriguingly, the analysis found 940 T-cell neoantigens predicted in >15% of patients in TCGA (shared) (Figure 12A). Although a much smaller cohort, 439 of 8,422 shared neoantigens were found in Van Allen, overlapping with the 940 in TCGA.

[00474] TCGA shared neoantigens were found to more frequently result from cassette exon as compared to unique neoantigens (only found in one patient) (Figure 12B). Genes with shared neoantigens were significantly enriched in gene sets for melanocyte biology, such as melanocyte differentiation ($p = 9.3 \text{ e-}4$, FDR), melanin biosynthetic process ($p = 3.4 \text{ e-}3$, FDR) and cell division ($p = 1.0 \text{ e-}4$, FDR) (Figure 12C). Thus, this analysis revealed that splicing neoantigens tend to occur in resident melanocyte genes and those which regulate NMD regulation and are predominantly associated with alternative splice-site selection.

[00475] As shared splicing neoantigens are rare relative to all unique predicted neoantigens, it would be expected that such peptides are more frequently detected using immunopeptidome profiling in an independent cohort. Hence, a previously published melanoma dataset of HLA-I bound peptides detected by MS in 24 independent patients was analyzed (the Bassani-Sternberg cohort).

[00476] Considering both shared ($n=613$) and unique ($n=16,753$) peptides in the reference peptide database (combined unique reference database, excluding peptides in UniProt), the study indeed found that shared splicing neoantigens can be detected at a higher rate than unique neoantigens (paired t-test, $p=0.006$) (Figure 12D). Additionally, of the 613 shared neoantigens examined, immunopeptidome evidence was found for 34% ($n=210$), with 98 shared neoantigens observed in at least 15% of patients by immunopeptidome analysis in Bassani-Sternberg, as illustrated using kernel density estimates or estimator of the empirical cumulative distribution (eCDF) function (Figure 12E, Figure 13A). Inspection of neojunctions produced from these MHC-presented neoantigens showed that they were tumor-specific (Figure 13B). Taken together, this MS validation from an independent melanoma cohort further demonstrates that those common neoantigens can be leveraged as high-value therapeutic targets.

[00477] In addition to exploring the regulation of alternative splicing, the study investigated whether selective amino acid preferences exist between shared and unique splicing neoantigens. While HLA genes are highly polymorphic in humans and have different binding preferences for neoantigens, the existence of shared splicing neoantigens suggests that diverse HLA genotypes can bind in a more promiscuous manner to these peptides versus those present only in one or few individuals.

[00478] To evaluate the potential recurrence of amino acid sequences, the shared and unique neoantigens were first redefined by normalizing the frequency of their parental splicing junctions (Figure 12F, see also the methods described in Example 1). Since 9-mer neoantigens have a balanced number of shared and unique neoantigens, the focus was specifically on these peptides. To derive a physicochemical profile of the amino acids associated with each, each 9-mer was encoded as a numerical vector based on all amino acid physicochemical parameters in the AAIndex database (566 parameters) (see the methods described in Example 1), and each neoantigen vector was projected as a point in UMAP space (Figure 12G). While the majority of shared and unique splicing neoantigens have similar physicochemical characteristics, a few empirically observed clusters were found to be enriched in shared versus unique neoantigens.

[00479] To find preferentially detected amino acids present in the shared and unique neoantigens, motif analysis was performed with MEME in these two sets. MEME finds that shared neoantigens frequently end in lysine and phenylalanine (e-value=9.8e-14), whereas the most dominant motif found in unique neoantigens disproportionately end in arginine (e-value=0.14); the less significant E-value also indicates the more diverse binding modes for unique neoantigens (Figure 12H). Near identical motif enrichments were observed for shared vs. unique splicing neoantigens in the Van Allen cohort (Figure 13C). These data indicate that amino acid sequence bias can be used to find shared neoantigens that are bound by most MHC alleles.

[00480] To demonstrate that such splicing neoantigens represent viable targets for broadly applicable immunotherapies, their ability to directly bind MHC and induce T-cell reactivity was experimentally verified. For validation, 5 shared splicing neoantigens were selected, derived from three neojunctions in *PMEL*, *SLC45A2* and *CDH19*. While *PMEL* is an existing target for immunotherapy in melanoma (in phase I/II clinical trials), the predicted splicing neoantigen has not been previously described (novel exon-exon junction). These

neoantigens were selected based on availability of HLA matched donor cells, high neoantigen frequency in melanoma and tumor specificity.

[00481] An MHC stabilization assay using TAP deficient HLA-A*01 containing T2 cells confirmed the binding of two neoantigens to HLA-A*02 (Figure 14A-B). To assess immunogenicity, HLA-typed healthy blood PBMC was leveraged to prime T cells and test IFN γ responses upon pMHC stimulation (Figure 15). Frozen PBMC were primed 3 times with autologous peptide loaded DCs and thereafter tested for their IFN γ response against single HLA expressing 721.221 (221) cells, to ensure immunogenicity is specific for a single HLA genotype. The 5 neoantigen peptides loaded on 221 cells expressing their respective allele, generated similar IFN γ responses compared to known immunogenic FLU and/or HMCV peptide antigens in 3 donors (Figure 14C-D). Only weak IFN γ responses were detected in unstimulated T cells and in response to 221 without peptide, showing peptide specificity. Thus, these analyses provide strong evidence of T-cell immunogenicity by all 5 shared splicing neoantigens tested. Notably, prior described neoantigen prediction workflows have reported far lower validation accuracies.

EXAMPLE 8

Shared splicing neoantigens derive from tumor cells rather than the tumor microenvironment

[00482] Emerging data suggests that neoantigens can also be derived from the tumor microenvironment, including immune cells. While bulk tumor RNA-Seq enables the detection of splicing neoantigens, it cannot clarify the precise cellular source. To determine which cell-types splicing neoantigens derive from, single-cell (sc)-RNA-Seq from melanoma tumor biopsies from 17 patients (4,454 cells) was re-analyzed. These data were profiled using SmartSeq2 chemistry, allowing for the detection of over 340,000 exon-exon and exon-intron junctions present throughout the gene body, for 7 prior annotated cell-types (tumor, endothelial, cancer associated fibroblasts, B cell, T cell, natural killer cells, and macrophages).

[00483] This analysis found that only a small proportion of scRNA-Seq detected junctions are enriched in tumor cells >2 fold (n=30,523) versus those enriched in immune cells (n=236,954) (Figure 16A). Intriguingly, roughly equal numbers of immune and tumor enriched neojunctions were associated with TCGA melanoma neoantigens (1,289 and 979, respectively). However, restricting this analysis to shared splicing neoantigens in >15% of patients found that these neojunctions are almost entirely derived from tumor cells (n=195) versus immune (n=10) (Figure 16B). Tumor specific neojunctions included experimentally confirmed splicing antigens

(SCL45A2, CDH19, FCRLA) (Figure 16C). Thus, the tumor microenvironment appears to be a significant contributor to splicing neoantigen burden, but not a primary source for shared neoantigens.

[00484] To ensure that such neoantigens are not only a byproduct of a normal cellular proliferative program, thus limiting their therapeutic use, multiple *in vitro* bulk RNA-Seq datasets, in which human epidermal cells were induced to proliferate, were analyzed. This analysis found that proliferation associated junctions can only account for a small percentage (~6%) of all TCGA melanoma identified neojunctions (Figure 17A-C). Neojunctions from shared splicing neoantigens were frequently observed in only one sample from these data, specifically from 16 week fetal fibroblasts, but not from other fetal fibroblasts profiled. Furthermore, these overlapping melanoma neojunctions were only weakly detected (1-50 reads) in any cultured epidermal sample and were associated with cell cycle regulation (Figure 17D).

EXAMPLE 9

SNAF accurately predicts full-length mRNAs and stable proteoforms

[00485] An intriguing set of SNAF-T splicing neoantigens were those that occur in transmembrane proteins, such as GPR143 and SLC45A2, which occurred due to novel in-frame alternative splice-sites (5' or 3'). Such splicing events can result in novel cell-surface expressed proteins. As an alternative source of neoantigens that do not have degradation, MHC presentation, and T-cell receptor recognition, SNAF-B was applied to the same TCGA melanoma cohort. The SNAF-B workflow can be used to identify ExNeoEpitopes that have predicted transmembrane domains, but altered N-terminal or other extracellular sequences that could serve as novel epitopes for specifically designed monoclonal antibodies.

[00486] To predict full-length isoforms, novel exon-exon or exon-intron junctions (not in the Ensembl or UCSC mRNA database), can be inserted into the best matching isoform models based on exon composition, followed by *in silico* translation (Figure 18A, see also the methods described in Example 1). This workflow can optionally exclude predicted mRNAs that are expected to result in NMD and selectively include those that have high-confidence transmembrane domains based on a prior published Hidden Markov Model based topology prediction approach (TMHMM). Putative ExNeoEpitopes that result in deleted or novel extracellular polypeptides can be determined using the SNAF-B interactive viewer.

[00487] This analysis found 378 initial ExNeoEpitopes in the TCGA melanoma cohort, using prior well annotated transcripts as reference models (Ensembl, UCSC mRNAs). To initially assess the validity of such ExNeoEpitopes, long-read RNA isoform sequencing (“Iso-Seq”) was performed in four commonly used melanoma cell lines using the PacBio Sequel II platform. This Iso-Seq analysis found 17 of the predicted ExNeoEpitopes isoforms that perfectly matched the *in silico* predictions, and an additional 20 which partially matched (overlapping neojunction). One example is a novel alternative 3' site in Signal Regulatory Protein Alpha (SIRPA), which results in an in-frame deletion of the 5' end of exon 13, missing 21AA. The predicted full-length isoform was initially confirmed by Iso-Seq (Figure 18B).

[00488] In addition to predicting full-length isoforms, SNAF-B can match short-read identified junctions to other supplied isoform models (e.g., long-read or EST). Matching exon-exon and exon-intron junctions from the TCGA melanoma cohort to a publicly available PacBio Iso-Seq dataset of 10 commonly used cancer cell lines (Universal Human Reference) found 1207 additional full-length neo-isoforms associated with junctions not readily detected in GTEx. This included the same SLC45A2 neojunction, predicted by SNAF-T to be associated with poor survival, shared in 69% of patients, found to be MHC-presented (MS) and only expressed in tumors (Figure 19A-B). This alternative 5' donor site results in the deletion of 80AA which disrupts the 6th transmembrane domain, with the deleted region (AA 215-295) appearing to be composed of a transmembrane segment (AA 215-237) and a cytoplasmic segment (AA 237-295) in the reference protein (Figure 13C-D). As a result of the deletion, a region of polypeptide sequence normally positioned at the cytoplasmic face of the membrane is now predicted to reside in the extracellular domain, representing a potential novel neoepitope for CAR-T therapy.

[00489] Considering all SNAF-B short- and long-read predictions from 10 tumor cell lines (pan-cancer) together, SNAF-B was run to identify a total of 514 unique ExNeoEpitope proteins. These were filtered to 187 predictions in which the neojunction overlapped with the extracellular domain (UniProt) and was not contained within any other Ensembl or UCSC protein isoforms. In addition to proteoforms with missing polypeptide sequences, the study identified 12 initial candidate long-read supported isoforms with high GTEx evidence of tumor-specificity (mean reads/sample <0.1) that result in the inclusion of novel alternative first or cassette exons. These were initially identified through manual inspection of BLAT sequence

matches to the human genome and mRNA transcript databases (UCSC and Ensembl) and biased to ExNeoEpitopes detected in >15% of patients.

[00490] Visualization in the SNAF-B viewer found that 5 out of these 12 ExNeoEpitopes result in new inserted polypeptide sequences that impact a cytoplasmic region of the protein (OCA2, SLC2A10, TMEM9, IL13RA1, ATP13A1), one occurring within a transmembrane domain (ANO10), and five predicted to alter the extracellular region of the protein and result in stable transmembrane predictions (DCBLD2, NALCN, MET, SEMA6A, IGSF11).

[00491] While orthogonal long-read RNA sequencing demonstrates the validity of these transcripts from neojunction inference, it is possible that such isoforms are not properly folded or inserted into the cell membrane. First, to initially evidence the ability of these isoforms to produce functional transmembrane proteins, 2D and 3D protein structures were predicted for the tumor-specific and reference mRNA isoforms using Protter and Alphafold2 (57), respectively. In each of the examples, high-confident 2D and 3D structures of both the tumor and reference isoforms were observed, indicating that the deleted or inserted polypeptide sequence selectively impacts the extracellular portion of these proteins (Figures 18-21).

[00492] Finally, the ability of ExNeoEpitopes, with distinct *in silico* evidence, to traffic to the plasma membrane was evaluated. Specifically, C-terminal fluorescent tagged cDNAs were synthesized and transfected for three ExNeoEpitopes that gain new peptide sequences (SEMA6A, ANO10, IGSF11, DCBLD2) and three isoforms with deletions (SIRPA, MET, SLC45A2), along with the reference isoform for each. The synthesized reference isoforms for 5 out of the 7 ExNeoEpitopes (SEMA6A, SIRPA, MET, SLC45A2, DCBLD2) were expressed and found to co-localize to the transmembrane using a selective cell surface stain in HEK-293 cells. Confirming their expression and trafficking, partial transmembrane localization was also observed in 4 of 7 of the evaluated ExNeoEpitopes predicted by SNAF-B (SEMA6A, SIRPA, MET, SLC45A2) (Figure 18F-G, Figure 21B-F).

[00493] These data illustrate that tumor-specific splice isoforms can be higher precision candidates for CAR-T or monoclonal antibodies over existing targets. This can include conformational epitopes that impact protein structure.

EXAMPLE 10

Neoantigen target-containing peptides and proteins identified via SNAF-T and SNAF-B, respectively

[00494] To identify shared splicing neoantigens that are specific to one cancer or shared among multiple cancers, the SNAF-T and SNAF-B workflows were applied to 5 cancers: 1) melanoma, 2) ovarian cancer, 3) head and neck cancer (HNC), 4) breast cancer, and 5) myelodysplastic syndrome (MDS). All analyses were performed on publicly available datasets comprising: 1) bulk RNA-Seq patient individual tumor biopsies (NCI Genomics Data Commons or NIH Sequence Read Archive), 2) immunopeptidomics mass spectrometry profiles, 3) single-RNA-Seq data, 4) cancer cell lines long-read mRNA sequencing isoform models, and/or 5) associated open-access clinical metadata. The applied processes included the SNAF software in combination with the prioritization/selection/*in silico* confirmation processes of the SNAF-T and SNAF-B workflows, as described herein, for all steps, with the following exceptions: 1) integrated immunopeptidomics analyses were applied in melanoma and ovarian cancer, 2) cell-of-origin single-cell splicing analyses were applied in melanoma and MDS, and/or 3) confirmatory independent cohort analyses were performed in melanoma (Van Allen cohort).

[00495] With the exception of MDS, all analyses were performed on solid cancer tumor biopsies from the TCGA project. In the case of MDS, analyses were performed on bulk RNA-Seq from an independent MDS cohort of 44 patients and matched healthy controls bone marrow liquid biopsies.

[00496] MDS is a heterogeneous disorder of the hematopoietic stem cell (HSC) resulting in bone marrow failure and frequent progression to acute myeloid leukemia (AML). MDS is diagnosed in over 10,000 patients yearly (98). Survival varies greatly by the subtype of MDS, with survival ranging from few months (high risk MDS, t-MDS) to several years (low risk). Despite the introduction of new therapies, such as hypomethylating agents and immunomodulatory drugs, the only curative therapy to date remains allogeneic stem cell transplant. However, marrow transplant is limited to a subset of patients due to its toxicity and limited donor availability (in particular in older individuals). A striking finding in MDS is the frequent recurrence of somatic mutations in core factors that mediate pre-mRNA splicing (>50%). Such splicing factor mutations lead to global alterations in mRNA products shared in

MDS and AML. These mRNA isoforms represent largely unexplored targets for new immunomodulatory therapies for the prevention of cancer in aging adults.

[00497] Over 4,000 dysregulated and coordinated splicing events are induced in MDS patients with evidence of chronic TGF-beta signaling. This splicing was attributed to a loss of HNRPK activity, which is selectively impaired in HSC (patients, engineered mouse models and single cell splicing). Thus, coordinated signaling-mediated splicing changes and splicing factor mutations, are a major contributor to HSC splicing dysfunction in MDS, prior to transformation to AML. MDS clones lose niche associations and mobilize from the bone marrow to blood. Thus, liquid biopsies can be readily collected from blood for RNA- and protein-based diagnostics.

[00498] To evidence the existence of splicing antigens in MDS for the creation of new immunoprevention strategies, SNAF was applied to 44 patients and 23 controls bone marrow biopsies (RNA-Seq). The analyses of this MDS cohort demonstrate the existence of 5,473 splicing antigens, including 78 shared in 10-75% of patients. These include targets for antigen vaccines (predicted SNAF-T immunogenic) and those that alter the sequence composition of extracellular protein domains of transmembrane encoding proteins (ExNeoEpitopes, SNAF-B) (Figures 22 and 23). While the latter predictions are derived from short-read sequencing, they are secondarily evidenced from long-read single-cell RNA-Seq profiling of MDS bone marrow.

[00499] In head and neck cancer tumor biopsies from individual patients profiled in TCGA, SNAF-T splicing neoantigens corresponding to 155 neojunctions were found to be predicted to be presented in > 15% of patients. Further, 84 prioritized ExNeoEpitope full-length isoforms present in >15% of patients were identified. In breast cancer, 155 SNAF-T shared splicing neoantigens present in >10% of patients and 118 SNAF-B ExNeoEpitopes present in >10% of patients were identified.

[00500] The definition of shared splicing neoantigens was set at either >10% or >15%, depending on the size of the cohort and previous extent of tumor heterogeneity. Taken together, 2241 unique shared predicted immunogenic splicing neoantigens were found in these cohorts (SNAF-T) (SEQ ID NOs: 1-2241, included in the sequence listing submitted concurrently herewith; see also Table 3 of U.S. Provisional Application No. 63/580,300, IDENTIFICATION OF TARGETS FOR IMMUNOTHERAPY IN MELANOMA USING SPLICING-DERIVED NEOANTIGENS, filed on September 1, 2023, which is currently co-pending herewith and

which is incorporated by reference in its entirety). Among these 2241 predictions, 124 shared splicing neoantigens were present in 2 or more cancers. Of these 2241 shared splicing neoantigens identified by SNAF-T, four have been experimentally validated for immunogenicity, as shown in Table 3, namely SEQ ID NOs: 1-4.

[00501] Using SNAF-B, 304 unique shared predicted ExNeoEpitope protein isoforms were identified (SEQ ID NOs: 2242-2545, included in the sequence listing submitted concurrently herewith; see also Table 4 of U.S. Provisional Application No. 63/580,300, IDENTIFICATION OF TARGETS FOR IMMUNOTHERAPY IN MELANOMA USING SPLICING-DERIVED NEOANTIGENS, filed on filed September 1, 2023, which is currently co-pending herewith and which is incorporated by reference in its entirety). Among these 304 predictions, 54 were found to be present in 2 or more cancers, including the *in vitro* experimentally evidenced ExNeoEpitope in the gene MET (Melanoma and Head and Neck cancer).

Table 3. Shared splicing neoantigens identified by SNAF-T which have been experimentally validated for immunogenicity.

SEQ ID No.	Splicing Neoantigen	Junction Coordinate	Gene
1	IPDSQGNDI ₁ RLLGTEFQ ₂ TRRAMTLK	chr5:33954504-33963931(-)	SLC45A2
2	IIDNQEPVF	chr18:66509195-66511568(-)	CDH19
3	STESITATRL	chr12:55956189-55956949(-)	PMEL
4	HAAASFETL ₁ QLDVMPLP ₂ VLRR	chr1:161710492-161710760(+)	FCRLA

EXAMPLE 11

Exemplary Splicing-Derived Neoantigens (RLLGTEFQTRRAMTLK, IPDSQGNDI, QRETDFKMKF, TVTAGRQGIY, VILQDAGIGFI)

[00502] The processes described in Examples 1-10 above have resulted in the identification of various splicing-derived neoantigens, some candidates of which are described below.

[00503] Candidate 1:

[00504] Two splicing-derived neoantigens (RLLGTEFQTRRAMTLK and IPDSQGNDI), encoded by the gene *SLC45A2* (ENSG00000164175:E3.2_33963931-E4.2), were identified and found to be associated with poor overall survival (Wald test $p < 0.05$, z-score > 2) in a large diagnostic melanoma RNA-Seq cohort (TCGA) and good overall survival (Wald test $p < 0.05$, z-score < -2.0) in a melanoma immunotherapy RNA-Seq cohort (phs000452.v3). This neoantigen is not readily detected in normal healthy tissues (GTEx, see Figure 6B) but is frequently detected in melanoma (328/472 patients) by SNAF and by immunopeptidomics (4 independent melanoma patients).

[00505] Accordingly, this neoantigen is predictive of good response with standard anti-PD-1 checkpoint blockade therapy. As such, this candidate can be used: A) as a diagnostic (quantitative PCR for exon-exon junction chr5:33954504-33963931) to identify patients that are likely to benefit from standard therapy.

[00506] Candidate 2:

[00507] A splicing-derived neoantigen (QRETDFKMKF), encoded by the gene *TBC1D16* (ENSG00000167291:E38.6-E39.1, chr17:79950594-79950756), was identified and found to be associated with poor overall survival (Wald test $p < 0.0005$, z-score > 3) in a large diagnostic melanoma RNA-Seq cohort (TCGA) and decent overall survival (Wald z-score < -1.5) in a melanoma immunotherapy RNA-Seq cohort (phs000452.v3). This neoantigen is not readily detected in normal healthy tissues (GTEx, see Figure 7 and Figure 8) but is frequently detected in Melanoma (158/472 patients) by SNAF.

[00508] Accordingly, this neoantigen is a predictive of poor survival. As such, this candidate can be used: A) as a diagnostic (quantitative PCR for exon-exon junction chr17:79950594-79950756) to identify patients that may benefit (mixed outcome) from standard therapy, and/or B) targeted immunotherapy using this neoantigen.

[00509] Candidate 3:

[00510] A splicing-derived neoantigen (TVTAGRQGIY), encoded by the gene *GPRI43* (ENSG00000101850:E8.1-E9.1_9743655, chrX:9743655-9746044), was found to be associated with poor overall survival (Wald test $p < 0.05$, z-score > 2) in a large diagnostic melanoma RNA-Seq cohort (TCGA). This neoantigen is not readily detected in normal healthy tissues (GTEx, see Figure 6B) but is frequently detected in melanoma (192/472 patients) by SNAF.

[00511] Accordingly, this neoantigen is a predictive of poor response with standard anti-PD-1 checkpoint blockade therapy. As such, this candidate can be used: A) as a diagnostic (quantitative PCR for exon-exon junction chrX:9743655-9746044) to identify patients that are candidates for new therapeutics, and/or B) targeted immunotherapy using this neoantigen.

[00512] Candidate 4

[00513] A splicing-derived neoantigen (VILQDAGIGFI), encoded by the gene *MCF2L* (ENSG00000126217:E18.4-E19.1, chr13:113044888-113045271), were found to be associated with poor overall survival (Wald test $p = 0.05$, z-score = 2) in a large diagnostic melanoma RNA-Seq cohort (TCGA) and poor overall survival (Wald test $p = 0.05$, z-score = 1.9) in a melanoma immunotherapy RNA-Seq cohort (phs000452.v3). This neoantigen is not readily detected in normal healthy tissues (GTEx) but is frequently detected in melanoma (136/472 patients) by SNAF.

[00514] Accordingly, this neoantigen is a predictive of poor response with standard anti-PD-1 checkpoint blockade therapy. As such, this candidate can be used: A) as a diagnostic (quantitative PCR for exon-exon junction chr13:113044888-113045271) to identify patients that are unlikely to benefit from standard therapy but are candidates for new therapeutics, and/or B) targeted immunotherapy using this neoantigen.

[00515] Other exemplary neoantigens identified by the processes described herein include RETDFKMKF, YQDWQQNLY, FNEDGLRQV, as well as the immunogenic targets listed in SEQ ID NOs: 1-2545 submitted concurrently herewith.

EXAMPLE 12

Interactive Neoantigen Web Explorer

[00516] To facilitate the exploration and prioritization of the predicted neoantigens from SNAF, two interactive web applications were developed to visualize both T-cell and B-cell

neoantigens (Figure 24). The SNAF-T viewer allows users to explore different global neoantigen features (i.e., amino acid length, and frequency within a cohort) in a projected 2D UMAP space. Here, each neoantigen is embedded based on its physicochemical properties. Users can manually select clusters to identify enriched web logo AA motifs and explore individual neojunctions and neoantigens. The SNAF-B viewer can be used to interactively visualize GTEx and tumor counts for neojunctions, generate protein sequence alignments between a putative ExNeoEpitope and selected reference proteins, compare protein feature composition, secondary structure and solvability prediction, and perform topology modeling. Hence, these tools can be used to select optimal targets for experimental validation.

[00517] The various methods and techniques described above provide a number of ways to carry out the disclosure . Of course, it is to be understood that not necessarily all objectives or advantages described can be achieved in accordance with any particular embodiment described herein. Thus, for example, those skilled in the art will recognize that the methods can be performed in a manner that achieves or optimizes one advantage or group of advantages as taught herein without necessarily achieving other objectives or advantages as taught or suggested herein. A variety of alternatives are mentioned herein. It is to be understood that some preferred embodiments specifically include one, another, or several features, while others specifically exclude one, another, or several features, while still others mitigate a particular feature by inclusion of one, another, or several advantageous features.

[0100] Furthermore, the skilled artisan will recognize the applicability of various features from different embodiments. Similarly, the various elements, features and steps discussed above, as well as other known equivalents for each such element, feature or step, can be employed in various combinations by one of ordinary skill in this art to perform methods in accordance with the principles described herein. Among the various elements, features, and steps some will be specifically included and others specifically excluded in diverse embodiments.

[0101] Although the application has been disclosed in the context of certain embodiments and examples, it will be understood by those skilled in the art that the embodiments of the disclosure extend beyond the specifically disclosed embodiments to other alternative embodiments and/or uses and modifications and equivalents thereof.

[0102] In some embodiments, the numbers expressing quantities of ingredients, properties such as molecular weight, reaction conditions, and so forth, used to describe and claim certain embodiments of the application are to be understood as being modified in some instances by the term “about.” Accordingly, in some embodiments, the numerical parameters set forth in the written description and attached claims are approximations that can vary depending upon the desired properties sought to be obtained by a particular embodiment. In some embodiments, the numerical parameters should be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. Notwithstanding that the numerical ranges and parameters setting forth the broad scope of some embodiments of the application are approximations, the numerical values set forth in the specific examples are reported as precisely as practicable.

[0103] In some embodiments, the terms “a” and “an” and “the” and similar references used in the context of describing a particular embodiment of the application (especially in the context of certain of the following claims) can be construed to cover both the singular and the plural. The recitation of ranges of values herein is merely intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (for example, “such as”) provided with respect to certain embodiments herein is intended merely to better illuminate the application and does not pose a limitation on the scope of the application otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the application.

[0104] Preferred embodiments of this application are described herein. Variations on those preferred embodiments will become apparent to those of ordinary skill in the art upon reading the foregoing description. It is contemplated that skilled artisans can employ such variations as appropriate, and the application can be practiced otherwise than specifically described herein. Accordingly, many embodiments of this application include all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible

variations thereof is encompassed by the application unless otherwise indicated herein or otherwise clearly contradicted by context.

[0105] All patents, patent applications, publications of patent applications, and other material, such as articles, books, specifications, publications, documents, things, and/or the like, referenced herein are hereby incorporated herein by this reference in their entirety for all purposes, excepting any prosecution file history associated with same, any of same that is inconsistent with or in conflict with the present document, or any of same that may have a limiting affect as to the broadest scope of the claims now or later associated with the present document. By way of example, should there be any inconsistency or conflict between the description, definition, and/or the use of a term associated with any of the incorporated material and that associated with the present document, the description, definition, and/or the use of the term in the present document shall prevail.

[0106] In closing, it is to be understood that the embodiments of the application disclosed herein are illustrative of the principles of the embodiments of the disclosure . Other modifications that can be employed can be within the scope of the application. Thus, by way of example, but not of limitation, alternative configurations of the embodiments of the application can be utilized in accordance with the teachings herein. Accordingly, embodiments of the present application are not limited to that precisely as shown and described.

Claims

What is claimed is:

1. An immunogenic target associated with one or more types of cancer, wherein the immunogenic target comprises a shared splicing amino acid-based neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor.
2. The immunogenic target of claim 1, wherein the target comprises a shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2241.
3. The immunogenic target of claim 2, wherein the target comprises a shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-4.
4. The immunogenic target of claim 1, wherein the target comprises a shared splicing neoantigen present in a protein having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2545.
5. The immunogenic target of claim 4, wherein the target comprises a shared splicing neoantigen present in a protein having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2245.
6. The immunogenic target of claim 1, wherein the target comprises a nucleotide coding for a shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2545.
7. The immunogenic target of claim 1, wherein the target comprises a nucleotide coding for a shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-4 or SEQ ID NOs 2242-2245.

8. The immunogenic target of any one of claims 1-7, wherein the immunogenic target comprises at least 90% sequence homology to any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor.

9. The immunogenic target of claim 1, wherein the target comprises a shared splicing neoantigen selected from SEQ ID NOs: 1-2545.

10. The immunogenic target of claim 9, wherein the target comprises a shared splicing neoantigen selected from SEQ ID NOs: 1-2241.

11. The immunogenic target of claim 9, wherein the target comprises a shared splicing neoantigen selected from SEQ ID NOs: 1-4.

12. The immunogenic target of claim 9, wherein the target comprises a shared splicing neoantigen selected from SEQ ID NOs: 2242-2545.

13. The immunogenic target of claim 1, wherein the target consists of a shared splicing neoantigen selected from SEQ ID NOs: 1-2545.

14. The immunogenic target of claim 13, wherein the target consists of a shared splicing neoantigen selected from SEQ ID NOs: 1-2241.

15. The immunogenic target of claim 13, wherein the target consists of a shared splicing neoantigen selected from SEQ ID NOs: 1-4.

16. The immunogenic target of claim 13, wherein the target consists of a shared splicing neoantigen selected from SEQ ID NOs: 2242-2545.

17. The immunogenic target of any one of the preceding claims, wherein the target is *in vitro*.

18. A cell comprising the immunogenic target of any one of the preceding claims.

19. An immunotherapy for treating one or more type of cancer, wherein the immunotherapy is specific for, or comprises, one or more immunogenic target of any preceding claim.

20. The immunotherapy of claim 19, wherein the immunotherapy is specific for, or comprises, one or more shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2241, or one or more nucleotide coding therefor, and wherein the type of cancer comprises skin cancer, breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS).

21. The immunotherapy of claim 20, wherein the immunotherapy is specific for, or comprises, one or more shared splicing neoantigen selected from SEQ ID NOs: 1-2241, or one or more nucleotide coding therefor, and wherein the type of cancer comprises skin cancer, breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS).

22. The immunotherapy of claim 19, wherein the immunotherapy is specific for, or comprises, one or more shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-4, or one or more nucleotide coding therefor, and wherein the type of cancer comprises skin cancer.

23. The immunotherapy of claim 22, wherein the immunotherapy is specific for, or comprises, one or more shared splicing neoantigen selected from SEQ ID NOs: 1-4, or one or more nucleotide coding therefor, and wherein the type of cancer comprises skin cancer.

24. The immunotherapy of claim 19, wherein the immunotherapy is specific for, or comprises, one or more shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2545, or one or more nucleotide coding therefor, and wherein the type of cancer comprises skin cancer, breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS).

25. The immunotherapy of claim 24, wherein the immunotherapy is specific for, or comprises, one or more shared splicing neoantigen selected from SEQ ID NOs: 2242-2545, or one or more nucleotide coding therefor, and wherein the type of cancer comprises skin cancer, breast cancer, head and neck cancer, ovarian cancer, and/or myelodysplastic syndrome (MDS).

26. The immunotherapy of claim 19, wherein the immunotherapy is specific for, or comprises, one or more shared splicing neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2245, or one or more nucleotide coding therefor, and wherein the type of cancer comprises skin cancer and/or head and neck cancer.

27. The immunotherapy of claim 26, wherein the immunotherapy is specific for, or comprises, one or more shared splicing neoantigen selected from SEQ ID NOs: 2242-2245, or one or more nucleotide coding therefor, and wherein the type of cancer comprises skin cancer and/or head and neck cancer.

28. The immunotherapy of any one of claims 19-27, wherein the immunotherapy comprises an RNA-, DNA-, or peptide-based vaccine, dendritic cell vaccine, chimeric antigen receptor T cell (CAR-T) therapy, T cell receptor-engineered T cell (TCR-T) therapy, antibody or antibody derivative, cell-based immunotherapy, immune checkpoint inhibitor, immunomodulator, topical immunotherapy, injection immunotherapy, adoptive cell transfer, oncolytic virus therapy, immunosuppressive drug, helminthic therapy, and/or other non-specific immunotherapy.

29. The immunotherapy of any one of claims 19-28, wherein the immunotherapy comprises an RNA-, DNA-, or peptide-based vaccine, dendritic cell vaccine, CAR-T therapy, and/or antibody or antibody derivative.

30. A composition comprising the immunogenic target of any one of claims 1-18.

31. A composition comprising the immunotherapy of any one of claims 19-29.

32. The composition of any one of claims 30 or 31, wherein the composition is formulated as a vaccine.

33. The composition of any one of claims 30-32, wherein the composition further comprises one or more cytokines, growth factors, or adjuvants.

34. An *in vitro* isolated dendritic cell comprising the immunogenic target of any one of claims 1-18.

35. The dendritic cell of claim 34, wherein the dendritic cell comprises a mature dendritic cell.

36. The dendritic cell of any one of claims 34 or 35, wherein the cell is a cell with an HLA type selected from HLA-A, HLA-B, or HLA-C.

37. An engineered T-cell receptor (TCR) or chimeric antigen receptor (CAR) that specifically recognizes the immunogenic target of any one of claims 1-18.

38. A cell comprising one or more TCR or CAR of claim 37.

39. An antibody, or antigen binding fragment thereof, that specifically recognizes the immunogenic target of any one of claims 1-18.

40. A method of treating one or more type of cancer, the method comprising:
administering, to a patient having a type of cancer, a treatment comprising therapeutically effective amount of a composition comprising the immunotherapy according to any one of claims 8-14.

41. A method of stimulating an immune response in a subject, the method comprising:

administering, to a subject having a type of cancer, a treatment comprising therapeutically effective amount of a composition comprising the immunotherapy according to any one of claims 8-14

42. The method of claim 40 or 41, wherein the immunotherapy is administered in combination with one or more additional therapies.

43. The method of claim 42, wherein the immunotherapy and the one or more additional therapies are administered together in one administration or composition.

44. The method of claim 42, wherein the immunotherapy and the one or more additional therapies are administered separately in more than one administration or more than one composition.

45. The method of claim 42, wherein the one or more additional therapies comprise surgical intervention, chemotherapy, radiation therapy, hormone therapies, immunotherapy, and/or adjuvant systematic therapies.

46. The method of any one of claims 40-45, wherein the subject has a tumor resistant to chemotherapy and/or radiation therapy.

47. The method of any one of claims 40-46, wherein the subject has been determined to be resistant to the previous treatment.

48. The method of any one of claims 40-47, wherein the cancer comprises stage I, II, III, or IV cancer.

49. The method of any one of claims 40-48, wherein the cancer comprises metastatic and/or recurrent cancer.

50. The method of any one of claims 40-49, wherein the subject is enrolled in a clinical trial.

51. A method of identifying one or more immunogenic targets for treating a type of cancer, wherein the method comprises:

receiving a sample dataset, wherein the sample dataset comprises mRNA from a cancer cell line, from a tumor cohort, or from a tumor sample from an individual patient;

analyzing the sample dataset by a machine learning model comprising:

identifying and quantifying post transcriptional regulation to detect one or more neojunctions and/or alternative splicing and/or alternative promoter regulatory events;

determining one or more candidate neoantigen targets based on the neojunctions and/or alternative splicing and/or alternative promoter regulatory events;

filtering the candidate neoantigen targets by:

a) setting a minimum threshold for antigen abundance in the dataset;

and/or

b) assessing the specificity of exon-exon or exon-intron junctions for tumor samples versus a healthy reference cell line or cohort dataset based on junction counts and isoform structure association; and

identifying a validated tumor- or cancer-specific immunogenic target when the candidate neoantigen target meets the filtering conditions.

52. The method of claim 51, the method further comprising identifying candidate neoantigen targets by one or more of the following:

a. immunopectidomics validation;

b. long-read RNA isoform sequence validation;

c. single-cell short-read and/or long-read RNA-Seq validation;

d. identification of unique tumor-specific extracellular protein sequences;

e. splicing event type filtering;

f. BayesTS probability;

g. survival association

- h. SashimiPlot visualization;
- i. BLAT searching;
- j. 3D topology evaluation; and/or
- i. filtering the candidate neoantigen targets based on one or more of the above steps and/or one or more additional filtering step.

53. The method of claim 52, wherein:

immunopeptidomics validation comprises filtering candidate neoantigen targets using one or more immunopeptidomics datasets comprising LC/MS data, with or without matching bulk RNA-Seq data;

long-read RNA isoform sequence validation comprises confirming candidate neoantigen targets in long-read isoform sequencing data from the cancer cell line, tumor cohort, or tumor sample and/or identifying tumor-specific full-length transmembrane protein isoforms;

single-cell short-read and/or long-read RNA-Seq validation comprises identifying common candidate neoantigen targets which are detected in one or more patients in a tumor sample versus microenvironment in a matching cancer and/or identifying tumor-specific full-length transmembrane protein isoforms;

identification of unique extracellular tumor-specific protein sequences comprises sequence comparison and alignment with a prior database, determination of genomic correspondence, and prediction of secondary protein structures and protein features;

splicing event type filtering comprises detecting splice-junctions not present in a reference database;

Bayes TS probability comprises evaluating BayesTS for the healthy reference cell line or cohort dataset and/or the sample dataset;

survival association comprises using a univariate Cox regression analysis to identify candidate neoantigen targets with a parental PSI value that is positively or negatively associated with patient outcome;

SashimiPlot visualization comprises determining the specificity of exon-exon or exon-intron junctions for the healthy reference cell line or cohort dataset vs the sample dataset via junction counts and isoform structure association relative to known isoforms in a reference database;

BLAT searching comprises evaluating full-length predicted mRNA sequences to confirm exon-region isoform specificity relative to prior annotated mRNA and ESTs and evaluating transposable element associations; and/or

3D topology evaluation comprises predicting 3D conformational differences in protein folding and domain architecture each candidate neoantigen target relative to a reference isoform.

54. The method of any one of claims 51-53, wherein filtering the candidate neoantigen targets comprises one or more of the following:

- a. determining if the splicing event is previously unknown;
- b. setting a threshold percentage of the sample dataset having the candidate neoantigen target;
- c. predicting 2D and/or 3D folding;
- d. predicting extracellular domain impact;
- e. SashimiPlot validation;
- f. evaluating secondary evidence from long-read sequencing;
- g. exclusion of protein isoform sequence data from a reference database;
- h. setting a threshold BayesTS probability score;
- i. setting a threshold average control tissue count;
- j. setting a threshold MHCNetPan binding; and/or
- k. setting a threshold DeepImmuno score > 0.5 .

55. The method of claim 54, wherein:

- a. the candidate neoantigen target is present in greater than 5%, 10%, 15%, 20%, 25%, 30%, or higher, of the sample dataset;
- b. the BayesTS probability score is less than 10, less than 5, less than 4, less than 3, less than 2, less than 1, or lower;
- c. the average control tissue count mean is less than 10, less than 5, less than 3, less than 2, less than 1, less than 0.5, or lower; and/or
- d. the MHCNetPan binding is less than 5%, less than 4%, less than 3%, less than 2%, less than 1%, less than 0.5%, less than 0.1%, or lower.

56. The method of claim 55, wherein:

- a. the candidate neoantigen target is present in greater than 15% of patients in the cohort;
- b. the BayesTS probability score is less than 3;
- c. the average control tissue count mean is less than 1; and/or
- d. the MHCNetPan binding is $< 2\%$ as a weak binder and $< 0.5\%$ as a strong binder.

57. The method of any one of claims 51-56, wherein filtering the candidate neoantigen targets comprises:

- a. long-read single-cell genomics data;
- b. patient survival association data;
- c. immunopeptidomics data; and/or
- d. patient survival association with antigen presence.

58. The method of any one of claims 51-57, the method further comprising validating candidate neoantigen targets by LC/MS, peptide-MHC binding, proteomics, and/or immunogenicity.

59. The method of claim 58, wherein:

- a. validation via LC/MS comprises spectral matching;
- b. validation via peptide-MHC binding comprises performing an MHC stabilization assay; and/or
- c. validation via immunogenicity comprises testing pMHCs on HLA genotyped healthy blood PMBC.

60. The method of any one of claims 51-59, wherein a ranking is derived for the candidate neoantigen target.

61. The method of claim 60, wherein the ranking further comprises immunogenicity score, HLA-binding score, neoantigen tumor specificity scores relative to healthy control samples

(maximum likelihood Estimation, Bayesian hierarchical model, control cohort mean counts), alternative splicing PSI value, mRNA junction occurrence in a reference transcriptome, patient cohort occurrence frequency, recurrence in independent cancers and cohorts, splice-event type, full-length peptide prediction length versus the reference, peptide sequence similarity to reference peptides, the presence of signal peptide sequence, secondary structure, 3D protein structure prediction, insertion versus deletion of amino acids, absolute junction read counts, topology predictions, and/or other programmatic outputs.

62. The method of any one of claims 51-61, wherein alternative splicing and alternative promoter regulatory events comprises cassette exon, 3'/5' splice site exon, intron retention, alternative terminal exon, trans-splicing, and/or other event which can produce unique exon-exon or exon-intron junctions for *in silico* translation.

63. The method of any one of claims 51-62, wherein the method identifies intron retention associated antigens, and/or predicts immunogenicity of the candidate neoantigen target.

64. The method of any one of claims 51-63, wherein the neojunctions comprise gene-associated exon-exon and exon-intron junctions.

65. The method of any one of claims 51-64, wherein the machine learning model comprises a deep learning model, probabilistic algorithms (Maximum Likelihood Estimation (MLE), hierarchical Bayesian model), and/or long-read sequencing concordance analysis.

66. The method of any one of claims 51-65, wherein the machine learning model comprises a Maximum Likelihood Estimation (MLE), hierarchical Bayesian model, and long-read sequencing concordance analysis.

67. The method of any one of claims 51-66, wherein the immunogenic target is located on a transmembrane protein.

68. The method of any one of claims 51-67, wherein the sample dataset comprises mRNA from a cancer cell line.

69. The method of claim 68, wherein the immunogenic target is shared among a patient population.

70. The method of any one of claims 51-69, wherein the sample dataset comprises mRNA from a tumor sample.

71. The method of claim 70, wherein the immunogenic target comprises a patient-specific neoantigen.

72. The method of any one of claims 51-71, wherein the method further comprises one or more additional reference datasets and/or algorithms.

73. The method of any one of claims 51-72, wherein the method is automated.

74. The method of any one of claims 51-73, wherein the immunogenic target comprises a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2545, or a nucleotide coding therefor.

75. The method of claim 74, wherein the immunogenic target comprises a neoantigen selected from SEQ ID NOs: 1-2545, or a nucleotide coding therefor.

76. The method of any one of claims 51-73, wherein the immunogenic target comprises a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-2241.

77. The method of claim 76, wherein the immunogenic target comprises a neoantigen selected from SEQ ID NOs: 1-2241.

78. The method of any one of claims 51-73, wherein the immunogenic target comprises a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 1-4.

79. The method of claim 78, wherein the immunogenic target comprises a neoantigen selected from SEQ ID NOs: 1-4.

80. The method of any one of claims 51-73, wherein the immunogenic target comprises a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID NOs: 2242-2545.

81. The method of claim 80, wherein the immunogenic target comprises a neoantigen present in a protein selected from SEQ ID NOs: 2242-2545.

82. The method of any one of claims 51-73, wherein the immunogenic target comprises a neoantigen having at least 80% sequence homology to the sequence in any of SEQ ID Nos: 2242-2245.

83. The method of claim 82, wherein the immunogenic target comprises a neoantigen present in a protein selected from SEQ ID Nos: 2242-2245.

84. The method of any one of claims 51-73, wherein the immunogenic target comprises a nucleotide coding for a shared splicing neoantigen selected from SEQ ID NOs: 1-2545.

85. The method of any one of claims 51-73, wherein the immunogenic target comprises a nucleotide coding for a neoantigen selected from SEQ ID NOs: 1-4 or SEQ ID NOs 2242-2245.

86. The method of any one of claims 51-85, wherein the method further comprises treating the type of cancer by administering, to a subject with the type of cancer, a therapeutically effective amount of a treatment, wherein the treatment targets the tumor or cancer-specific immunogenic target.

87. The method of any one of claims 51-86, wherein the sample is a tumor sample from the subject, and wherein the treatment targets the tumor-specific immunogenic target.

88. The method of any one of claims 51-87, wherein the type of cancer comprises melanoma, ovarian, head and neck, breast, skin, lung, adult and pediatric acute myeloid leukemia, myelodysplastic syndrome (MDS), bladder, cervical, colorectal, kidney, low grade glioma, glioblastoma, liver, prostate, pancreatic, esophageal, rectal, sarcomas, stomach, thymus, thyroid, testicular, and/or uterine cancer.

89. The method of claim 88, wherein the type of cancer comprises melanoma, ovarian, head and neck, breast, skin, and/or MDS.

90. The method of any one of claims 51-89, wherein the treatment comprises an immunotherapy.

91. The method of claim 90, wherein the immunotherapy comprises an RNA-, DNA-, or peptide-based vaccine, dendritic cell vaccine, chimeric antigen receptor T cell (CAR-T) therapy, T cell receptor-engineered T cell (TCR-T) therapy, antibody or antibody derivative, cell-based immunotherapy, immune checkpoint inhibitor, immunomodulator, topical immunotherapy, injection immunotherapy, adoptive cell transfer, oncolytic virus therapy, immunosuppressive drug, helminthic therapy, and/or other non-specific immunotherapy.

92. The method of claim 91, wherein the immunotherapy comprises an RNA vaccine, chimeric antigen receptor T cell (CAR-T) therapy, and/or antibody or antibody derivative.

93. The method of any one of claims 51-92, wherein the immunotherapy comprises the immunotherapy of any one of claims xx-xx.

94. The method of any one of claims 51-93, wherein the type of cancer is responsive to targeting of the tumor- or cancer-specific immunogenic target.

95. A method of treating a type of cancer in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of an immunotherapy comprising an immunogenic target selected from the peptides and proteins listed in SEQ ID NOs: 1-2545, or a nucleotide coding therefor.

96. A method of treating a type of cancer in a subject in need thereof, the method comprising administering to the subject a therapeutically effective amount of a compound, or a composition comprising a compound, directed to an immunogenic target selected from the peptides and proteins listed in SEQ ID NOs: 1-2545, or a nucleotide coding therefor.

97. The method of claim 96, wherein the compound binds to an immunogenic target selected from the peptides and proteins listed in SEQ ID NOs: 1-2545.

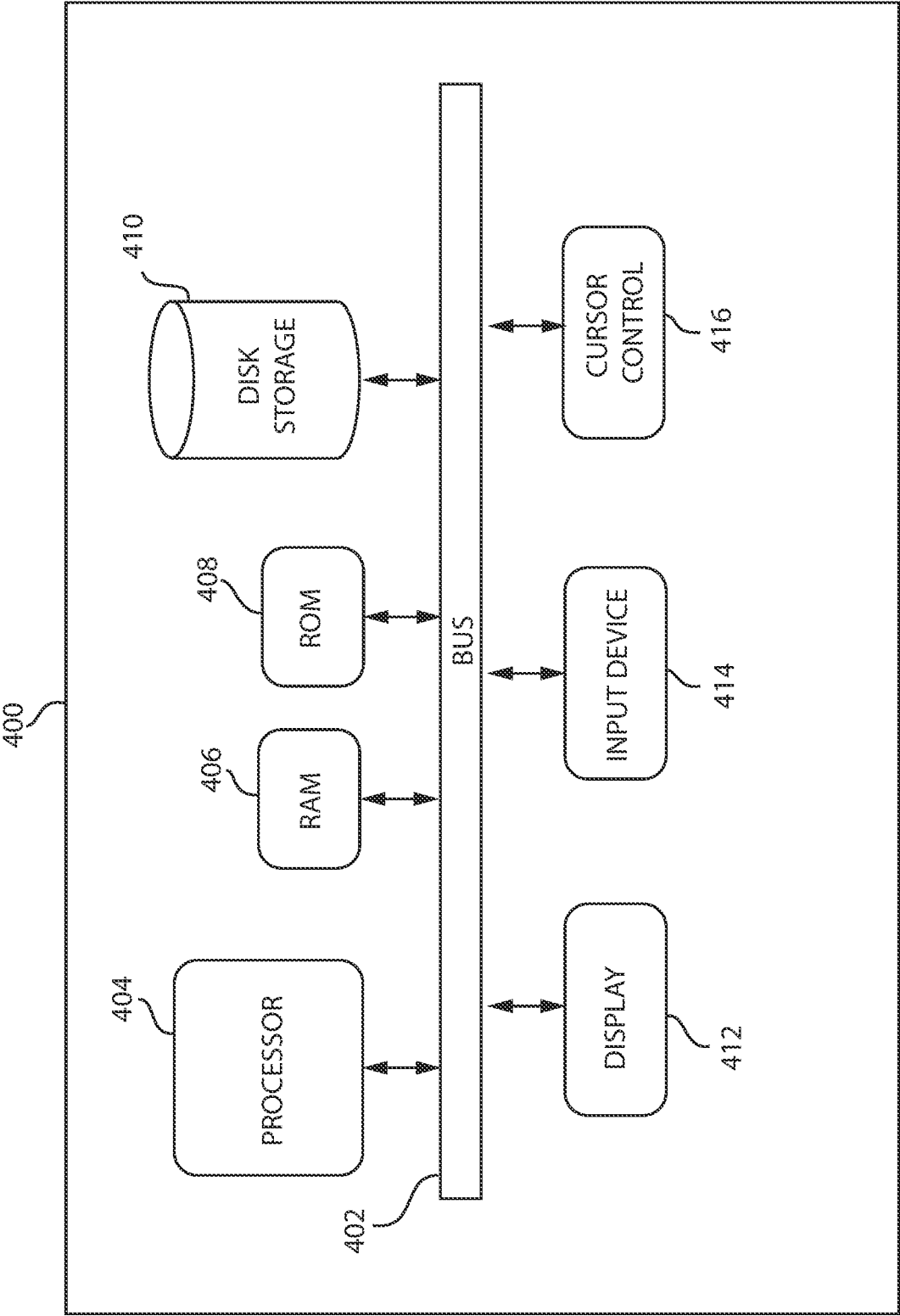


FIG. 1

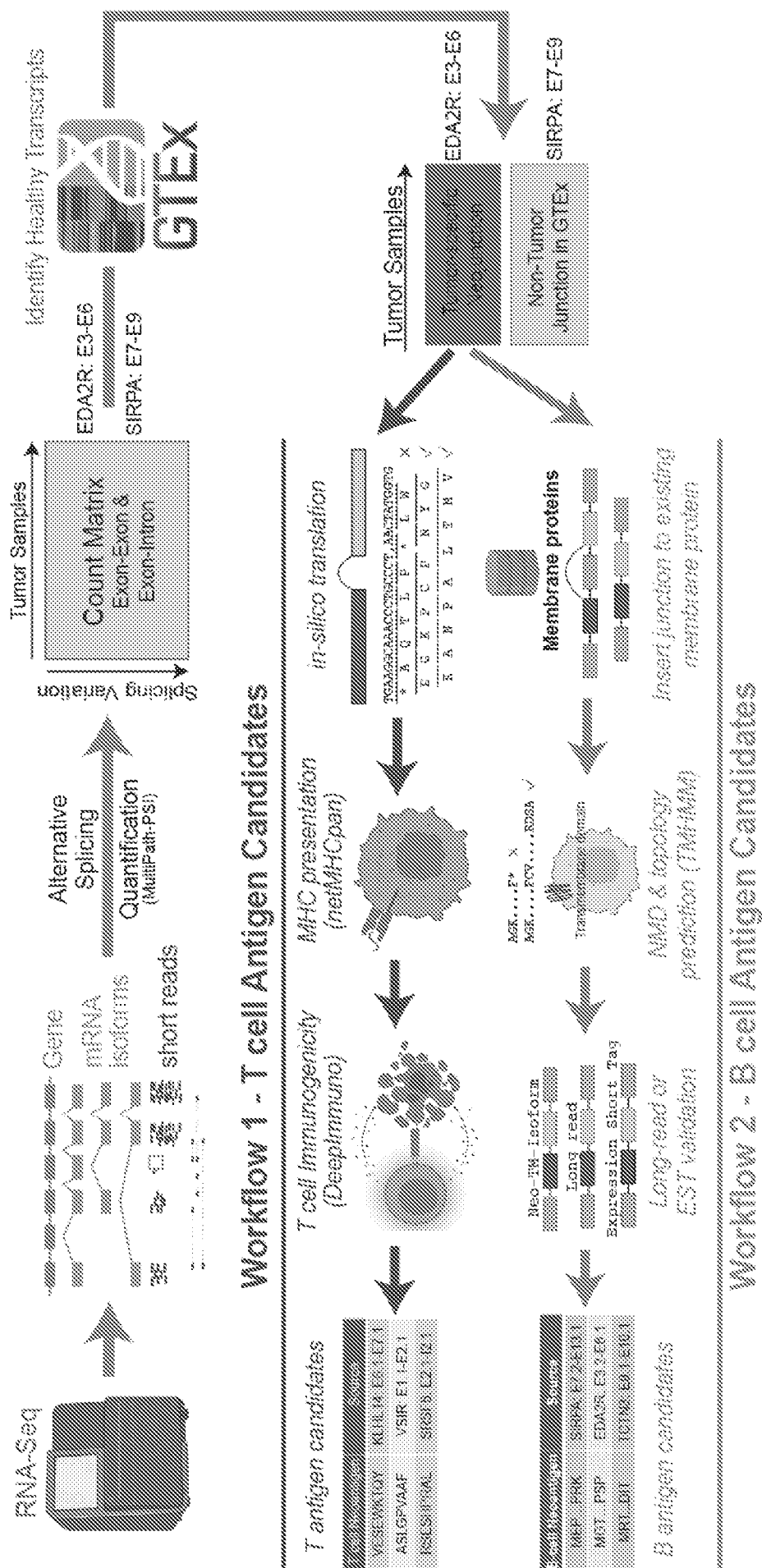


FIG. 2

Junction Detection

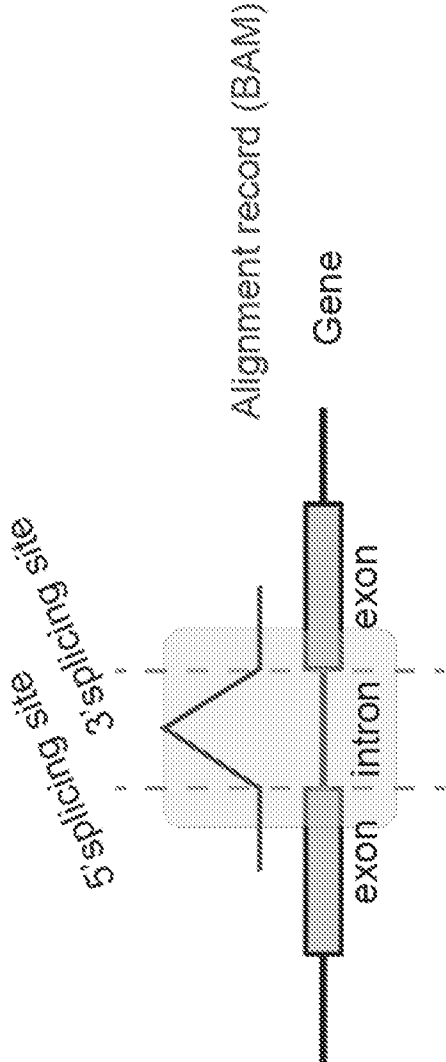
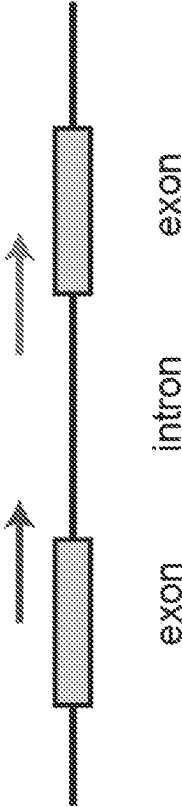


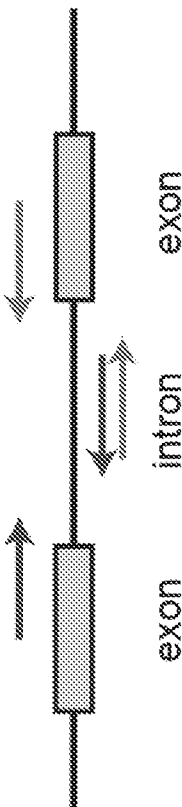
FIG. 3A

Intron Retention (single-end)



- Both intron-exon boundaries supported by >4 single-end reads

Intron Retention (pair-end)



- Both intron-exon boundaries supported by >4 single-end reads
- The mate reads for the above mentioned boundary reads should entirely lie in the intron region

FIG. 3B

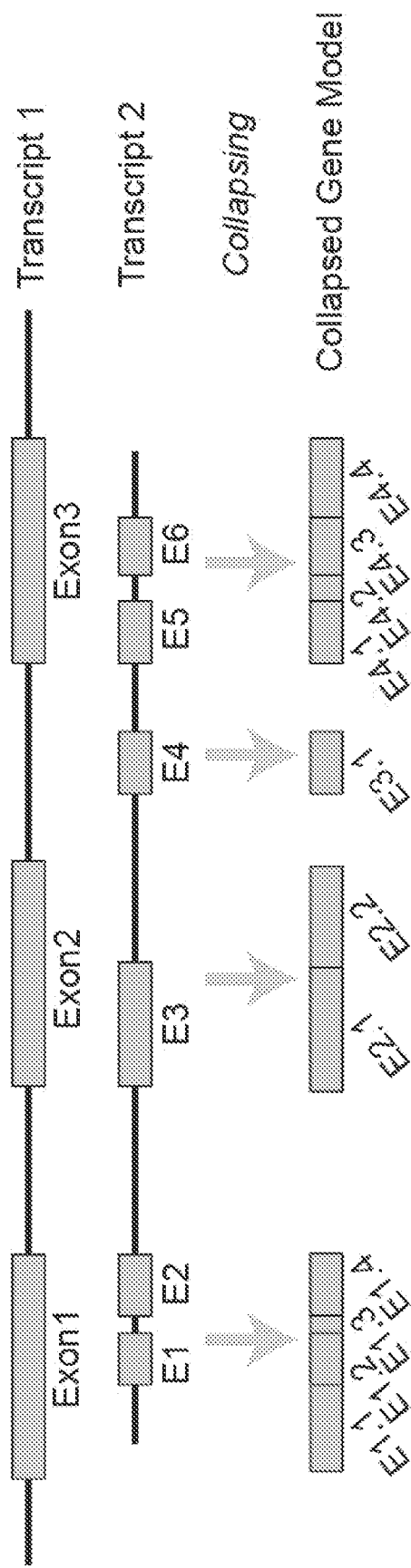


FIG. 3C

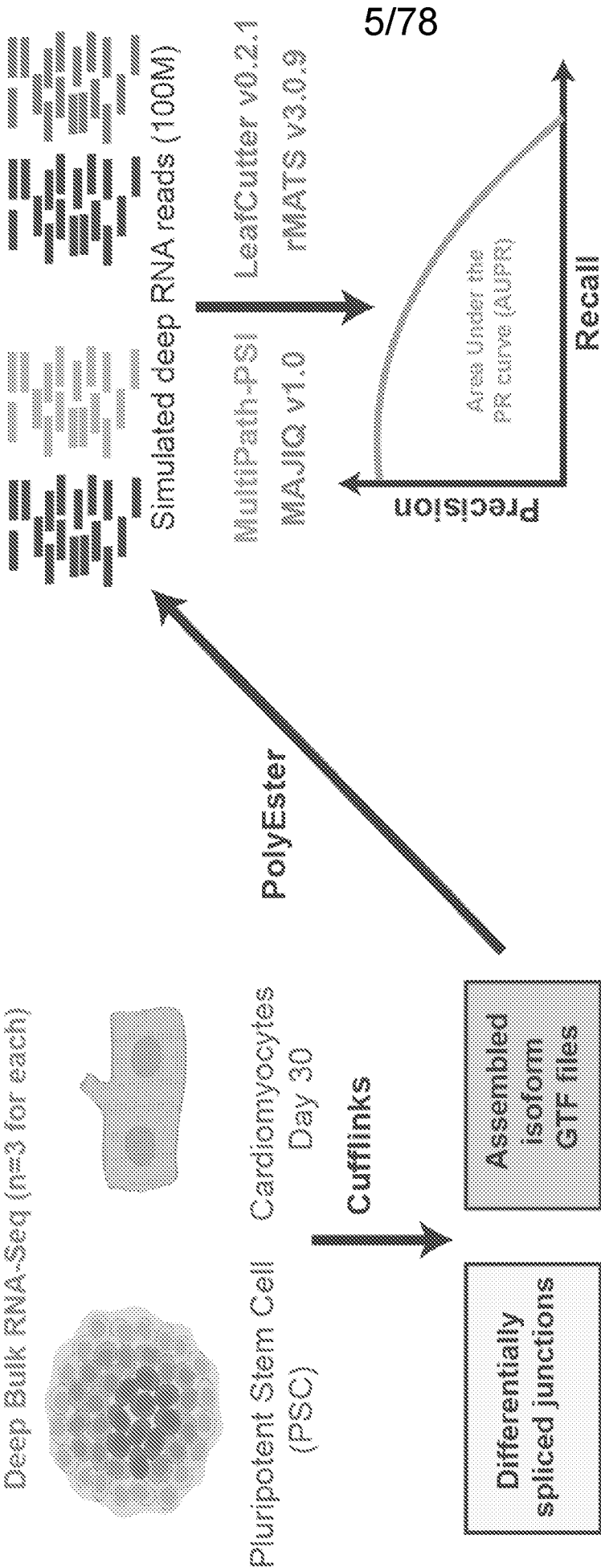
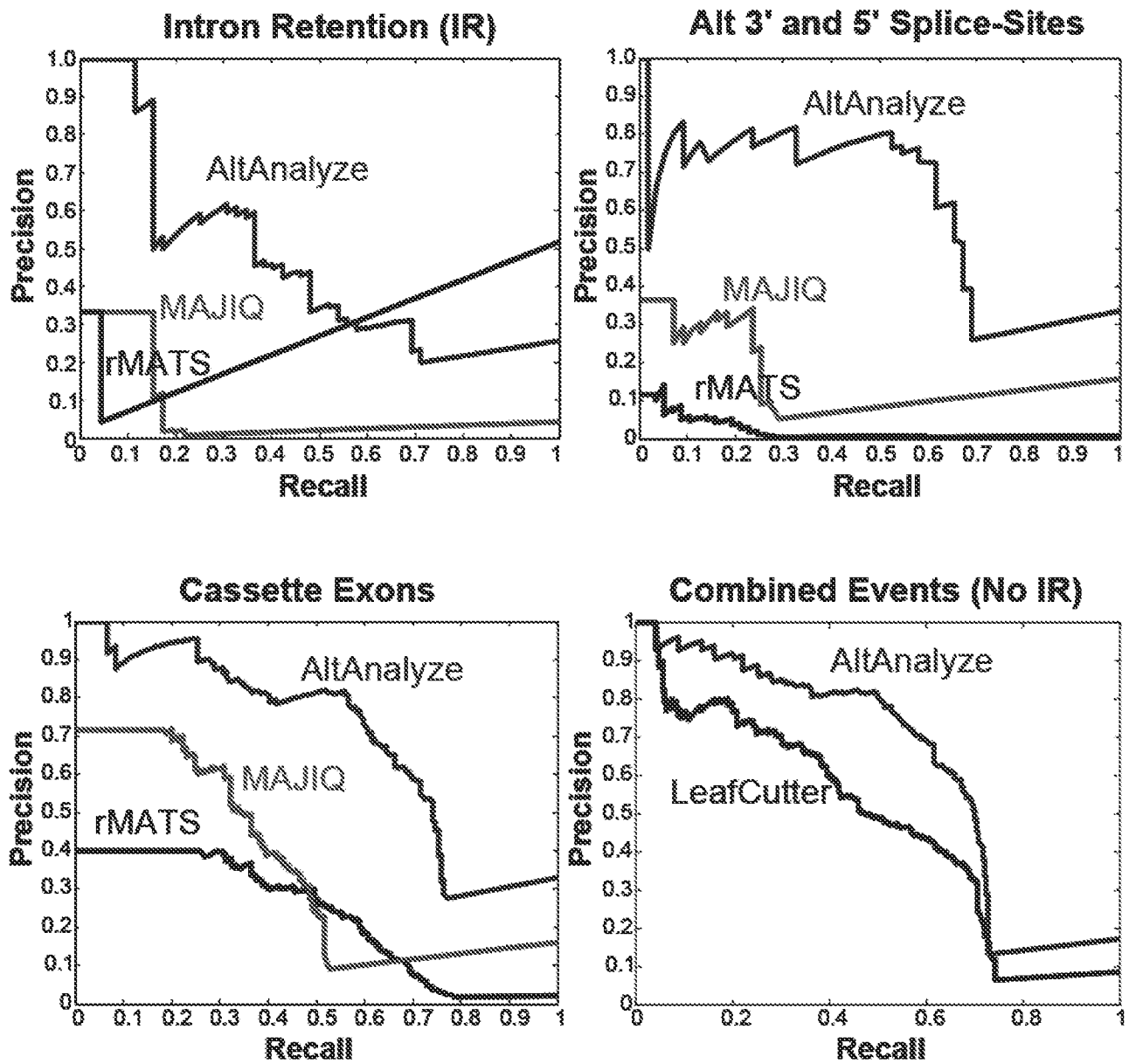


FIG. 4A

6/78

**FIG. 4B**

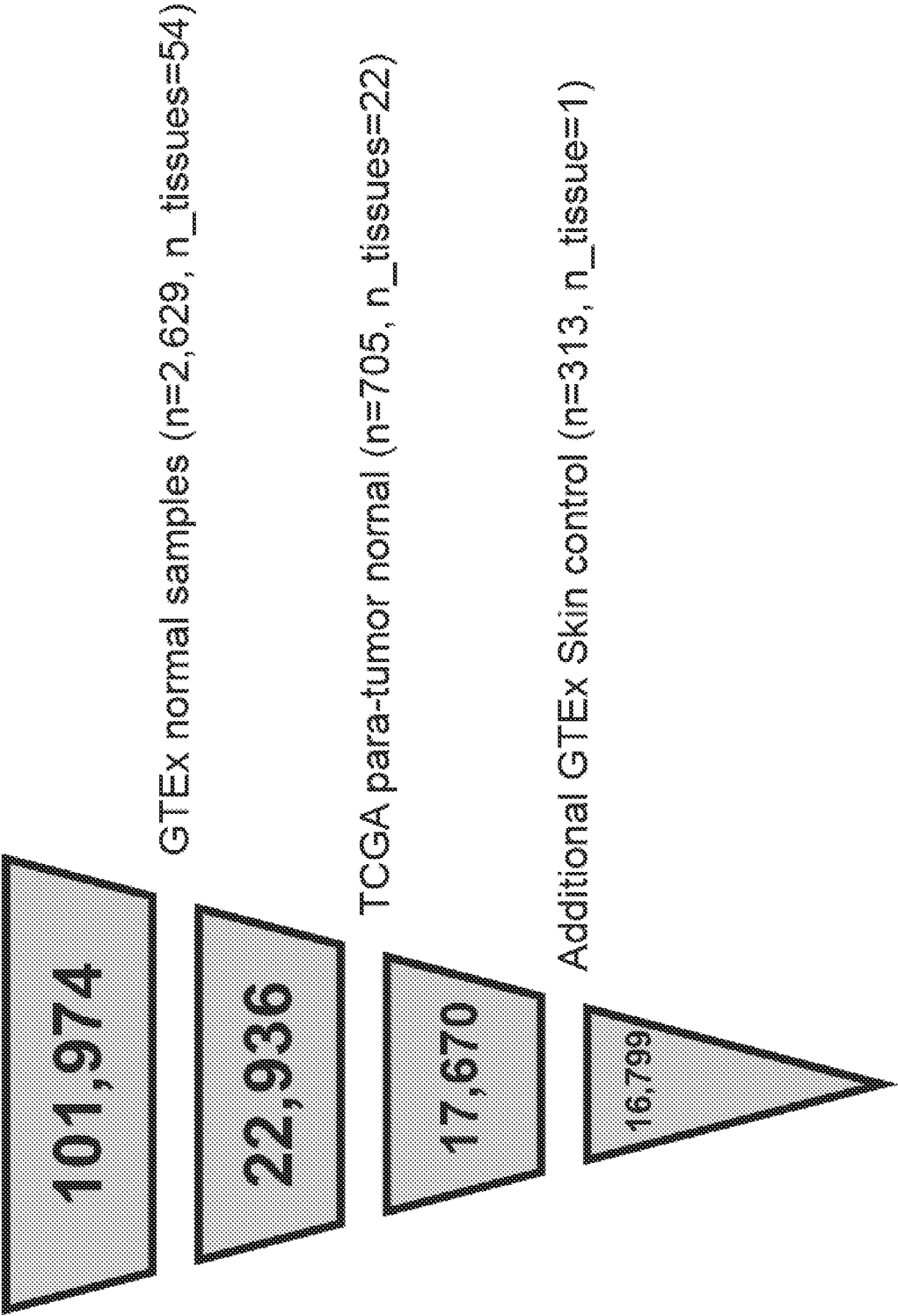


FIG. 5

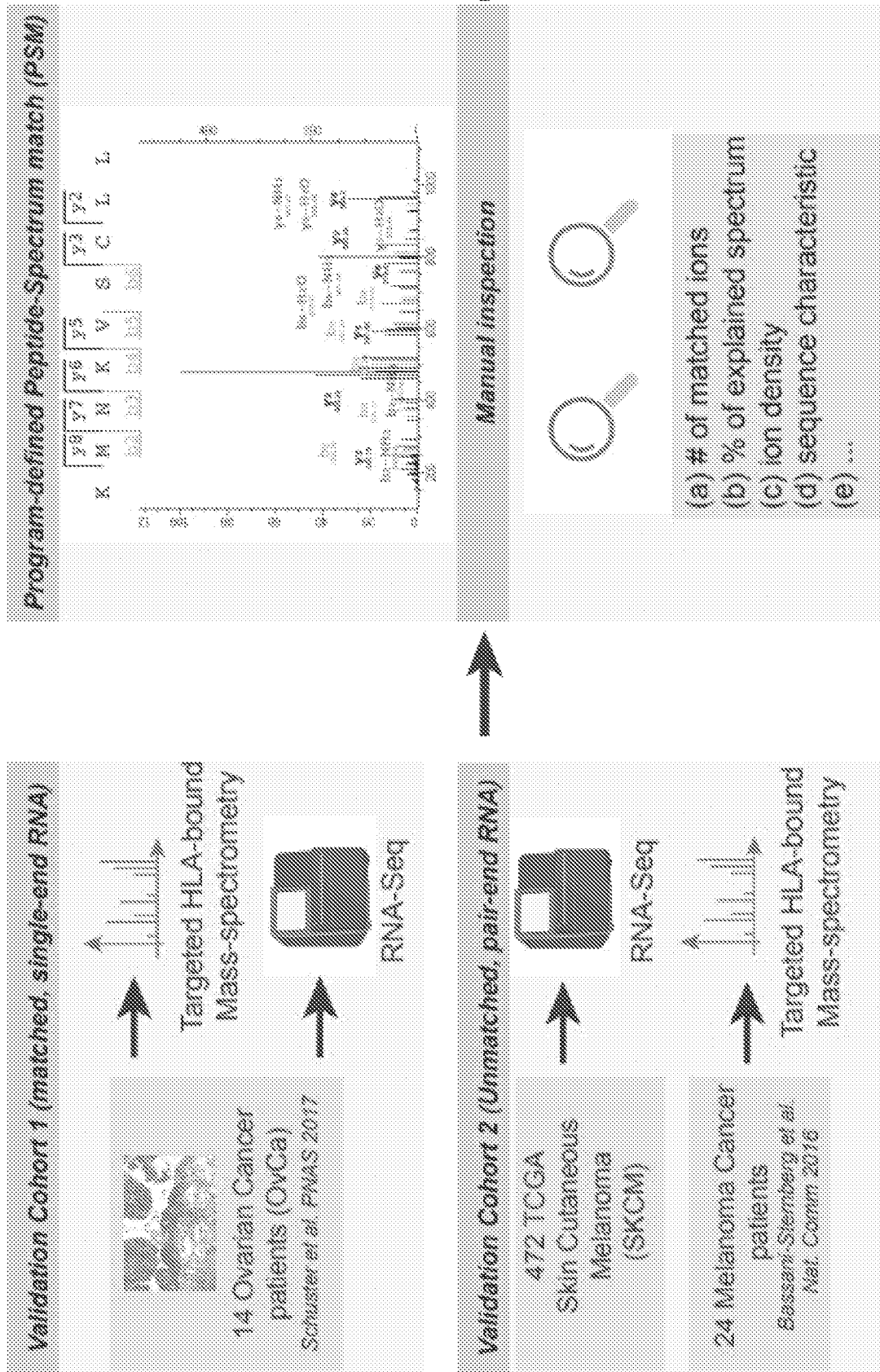


FIG. 6A

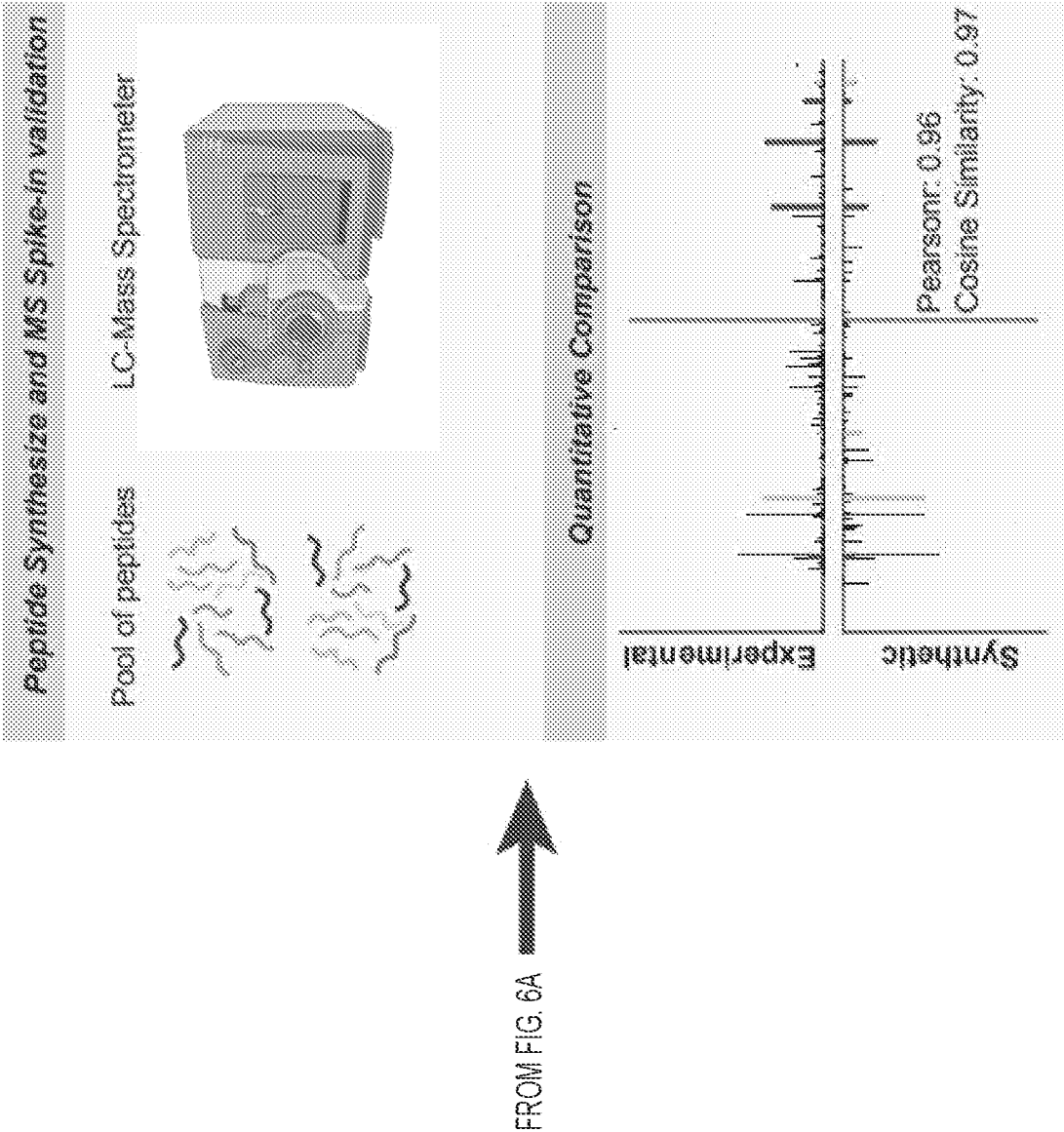


FIG. 6A (CONT.)

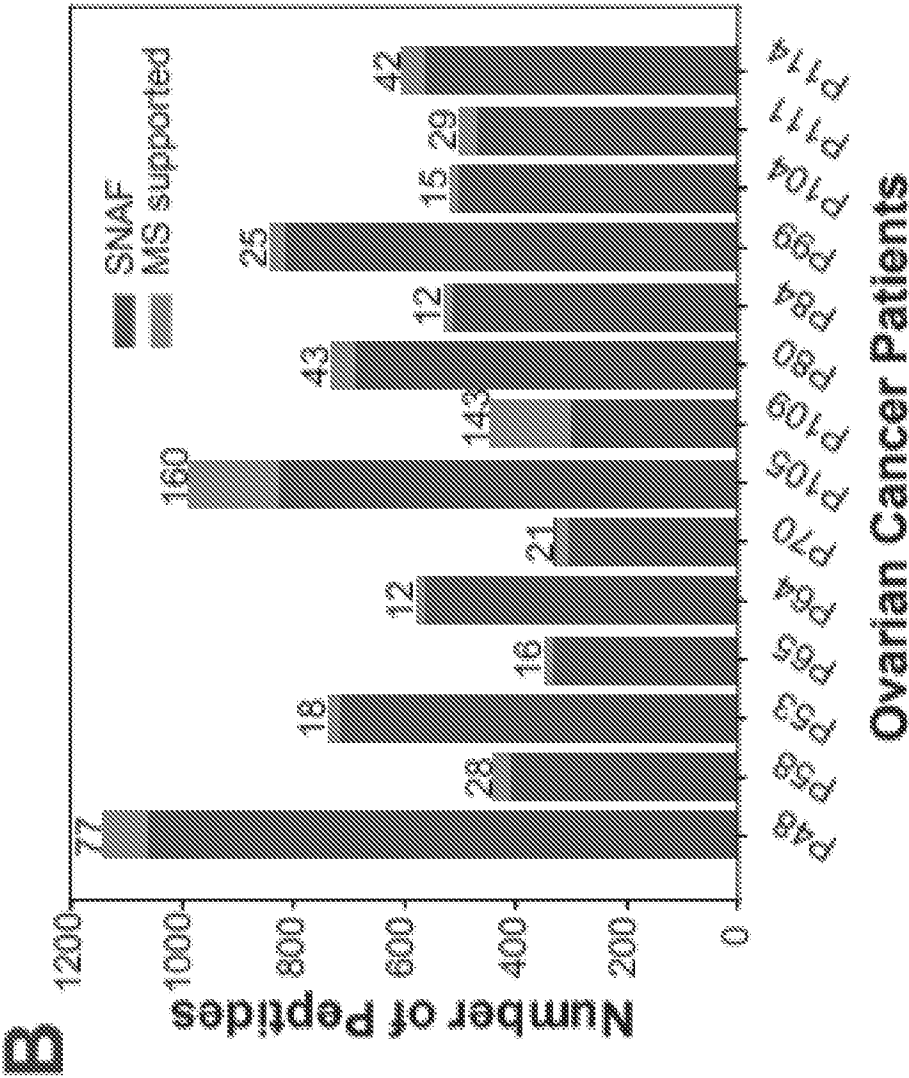


FIG. 6B

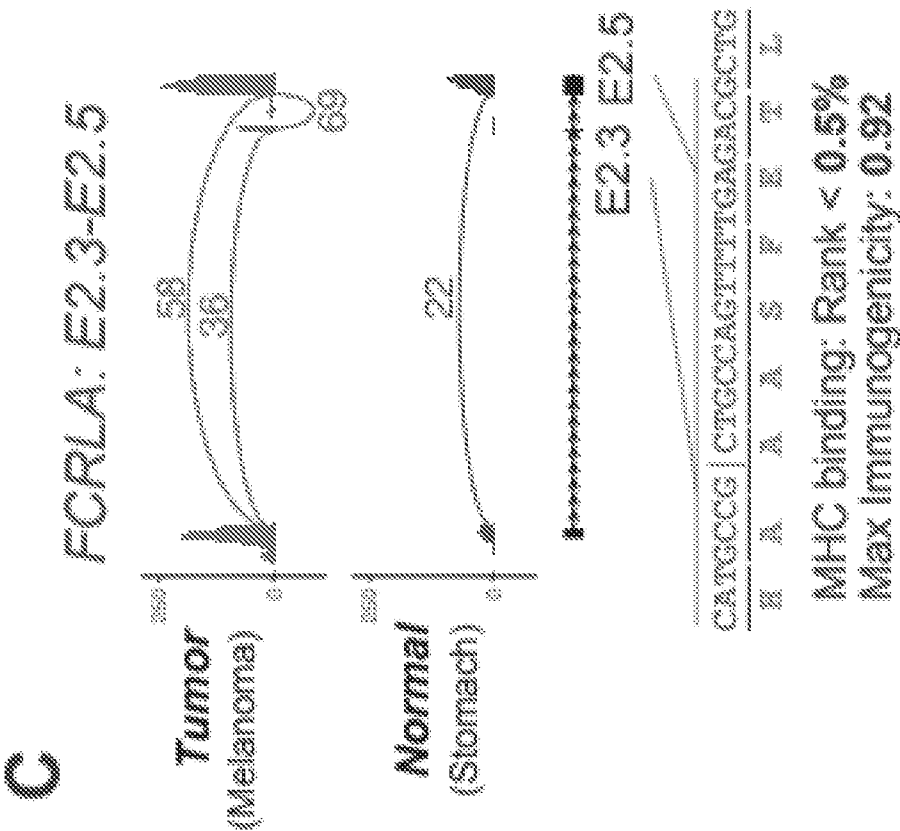


FIG. 6C

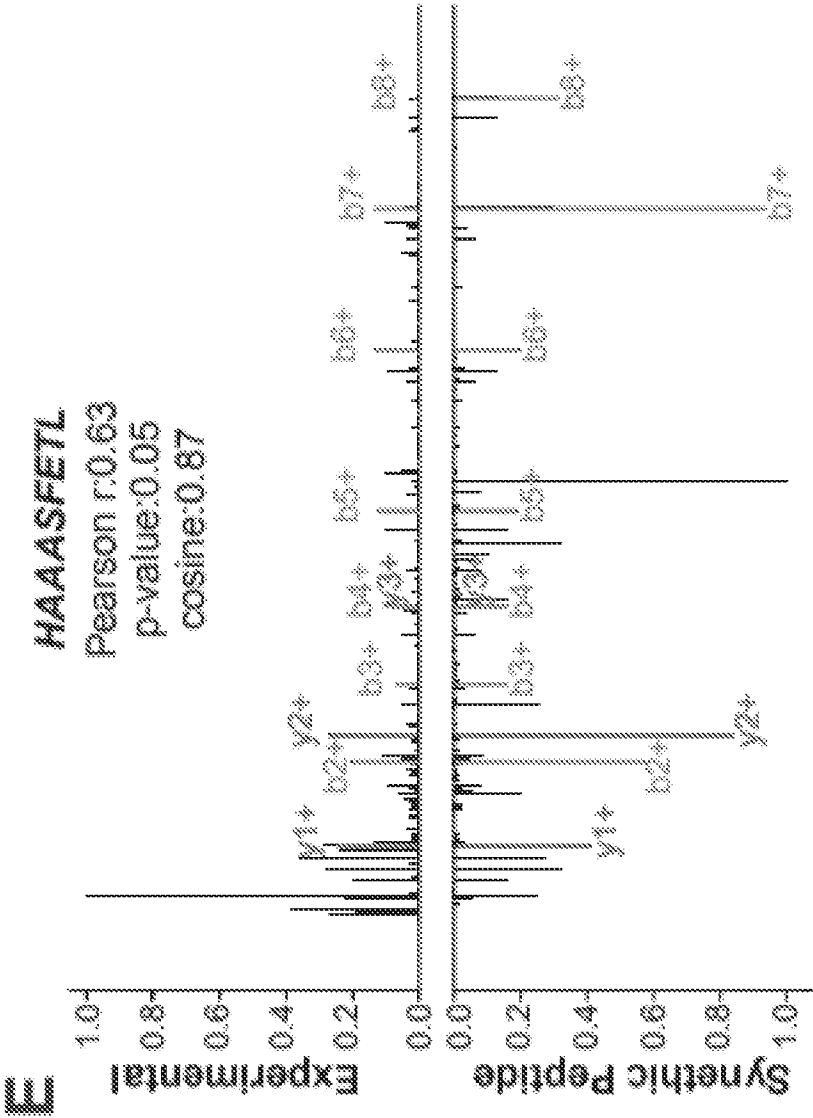


FIG. 6E

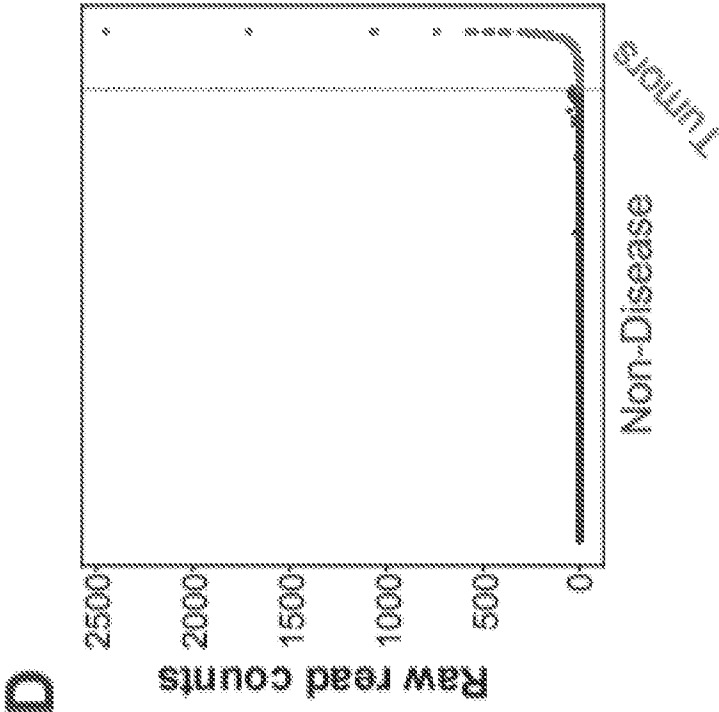


FIG. 6D

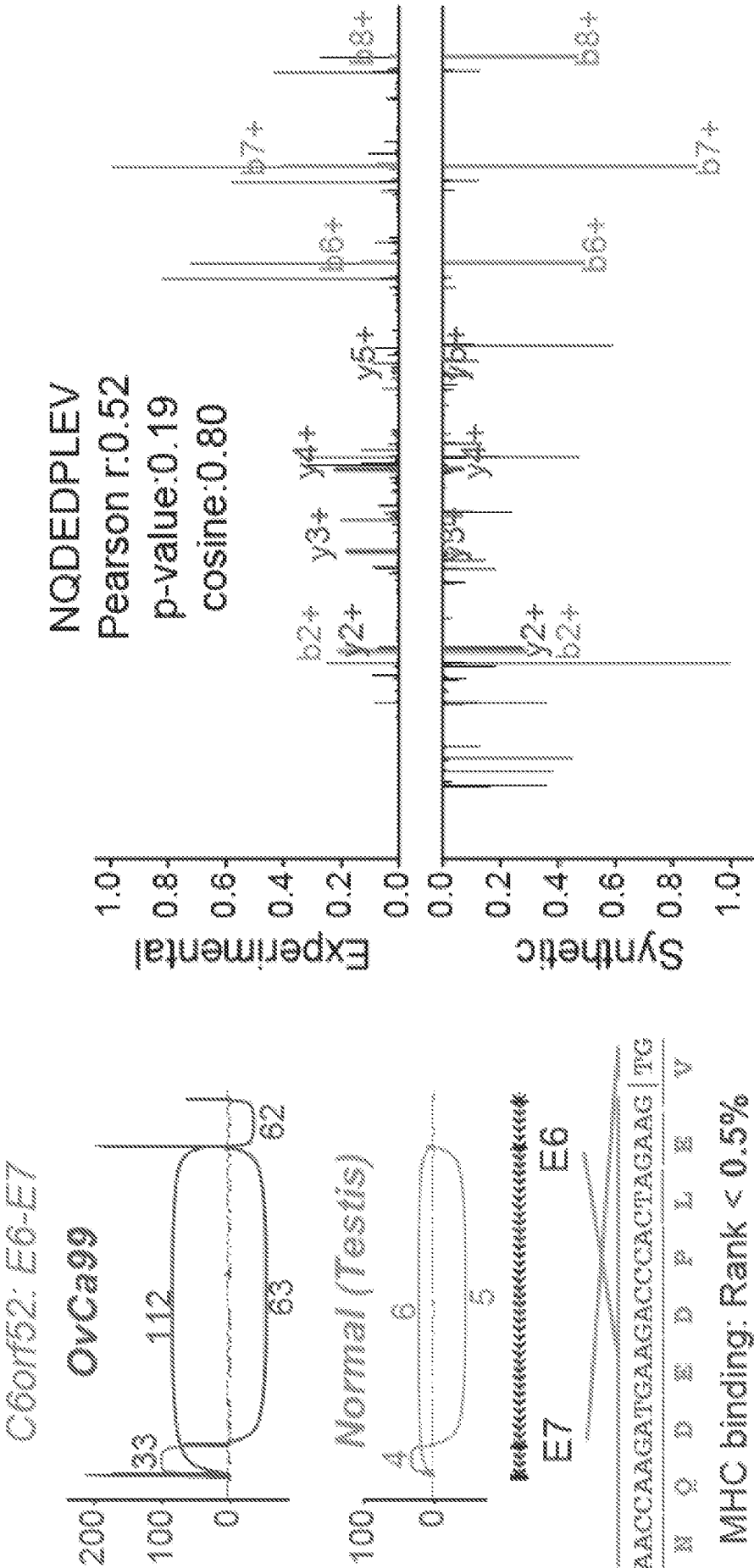


FIG. 7A

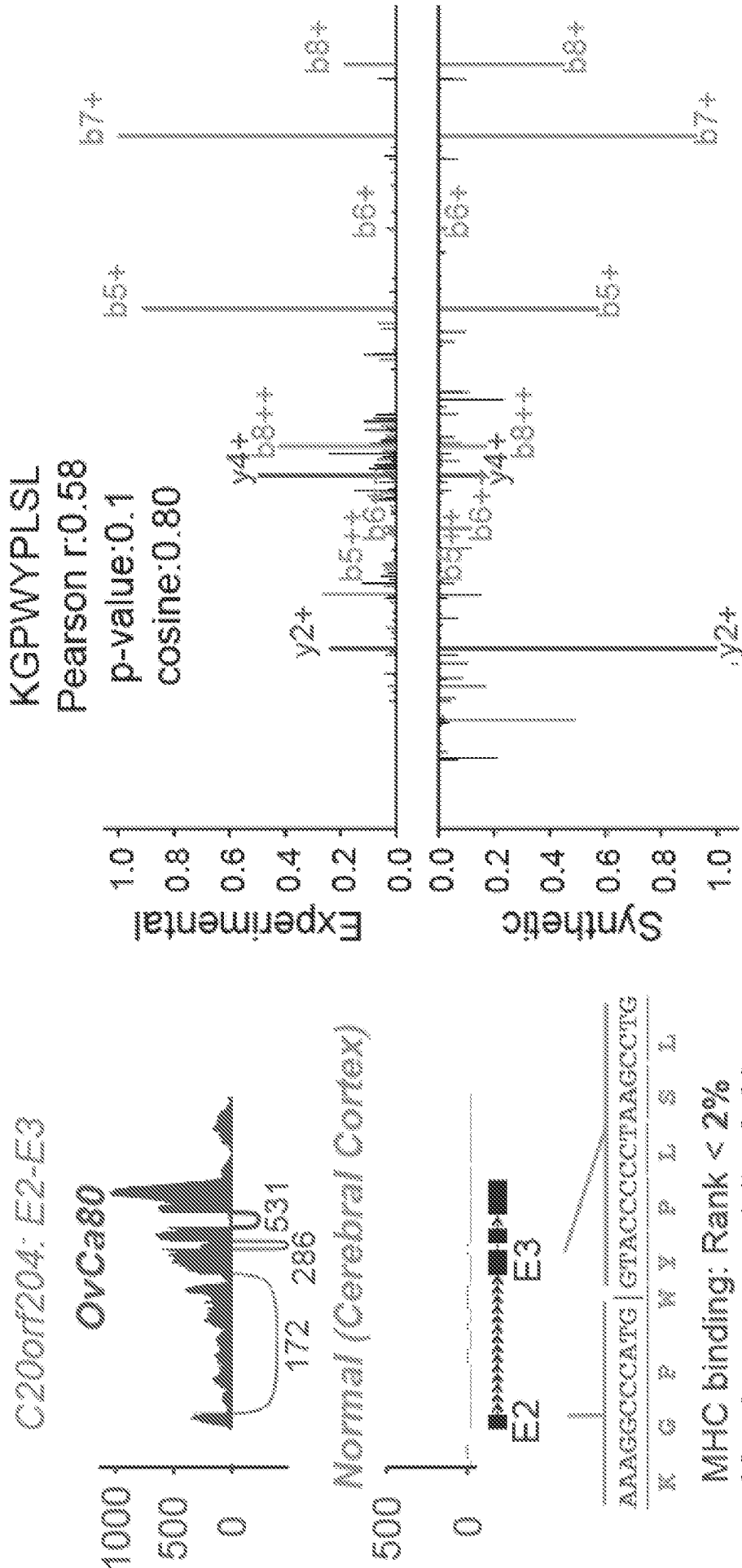


FIG. 7B

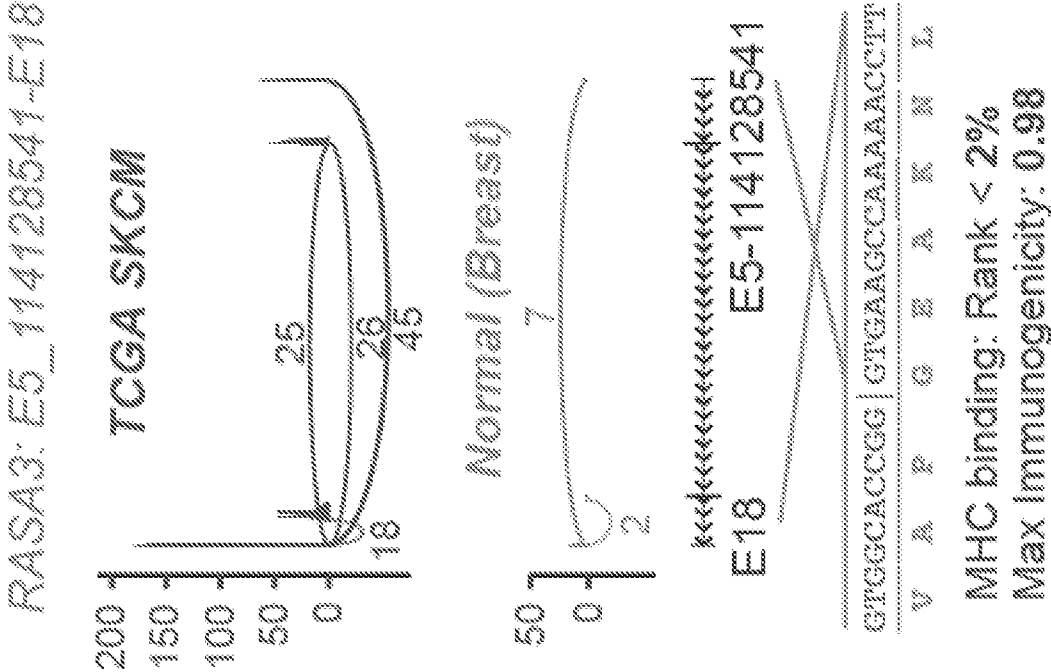
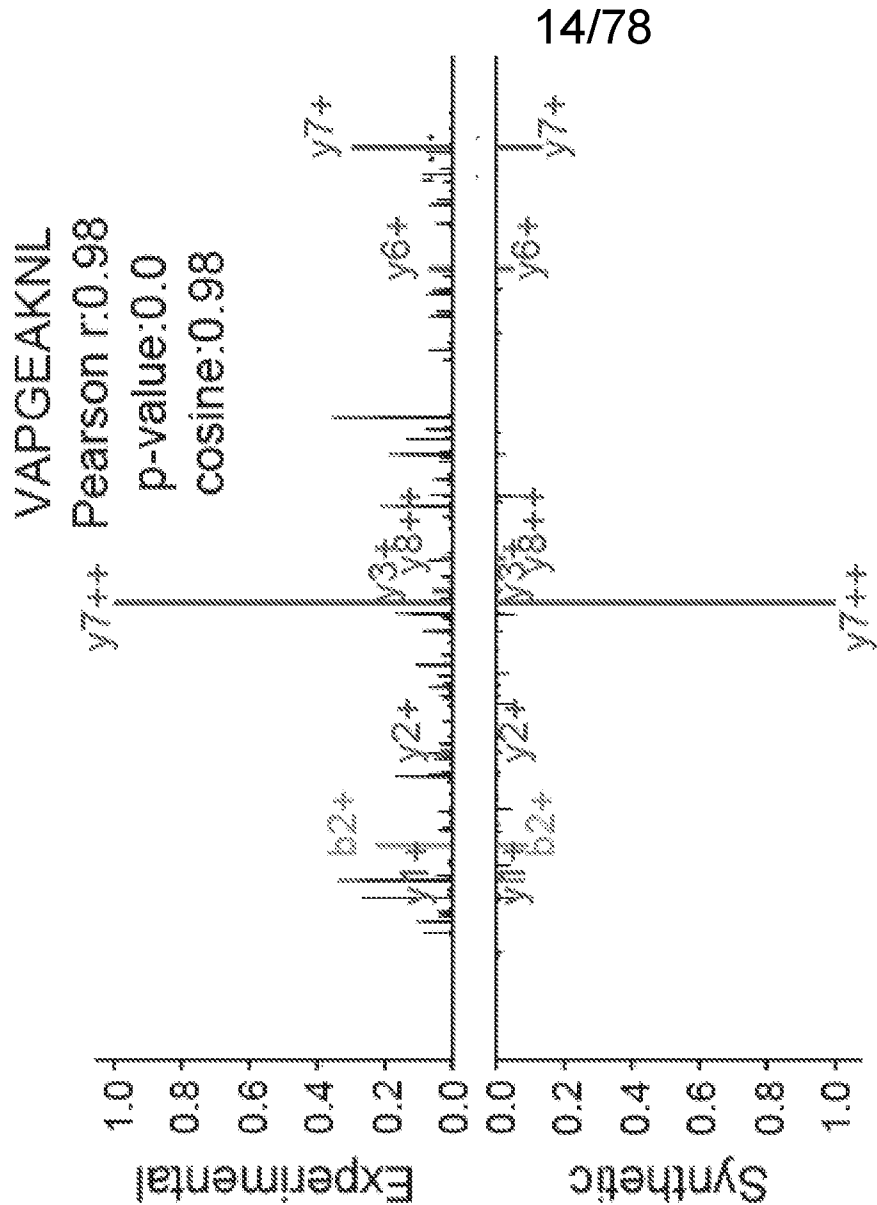


FIG. 7C

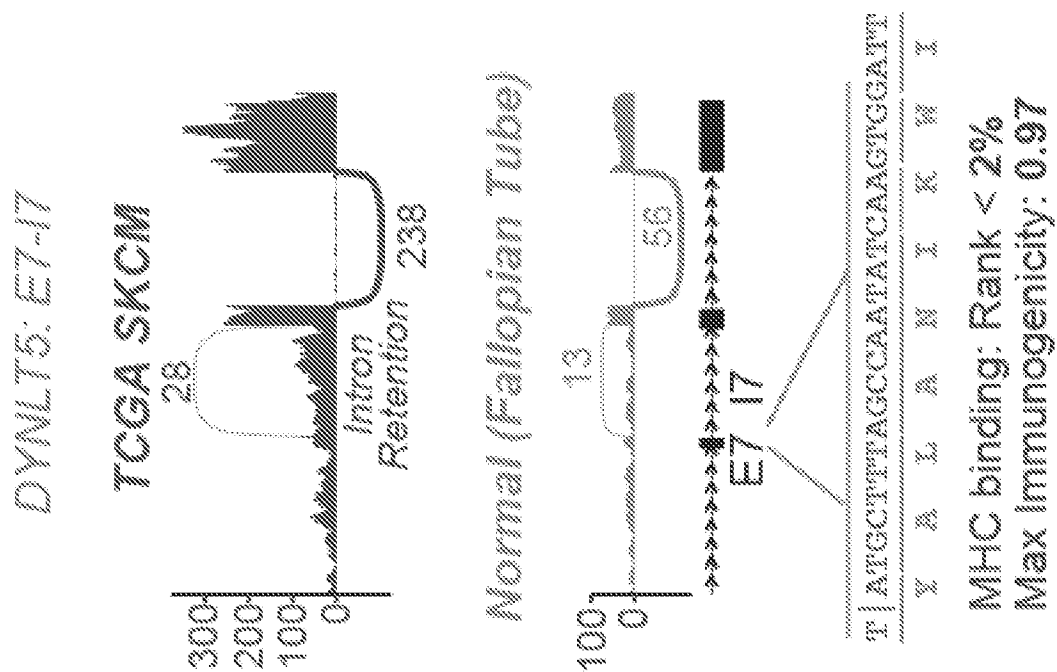
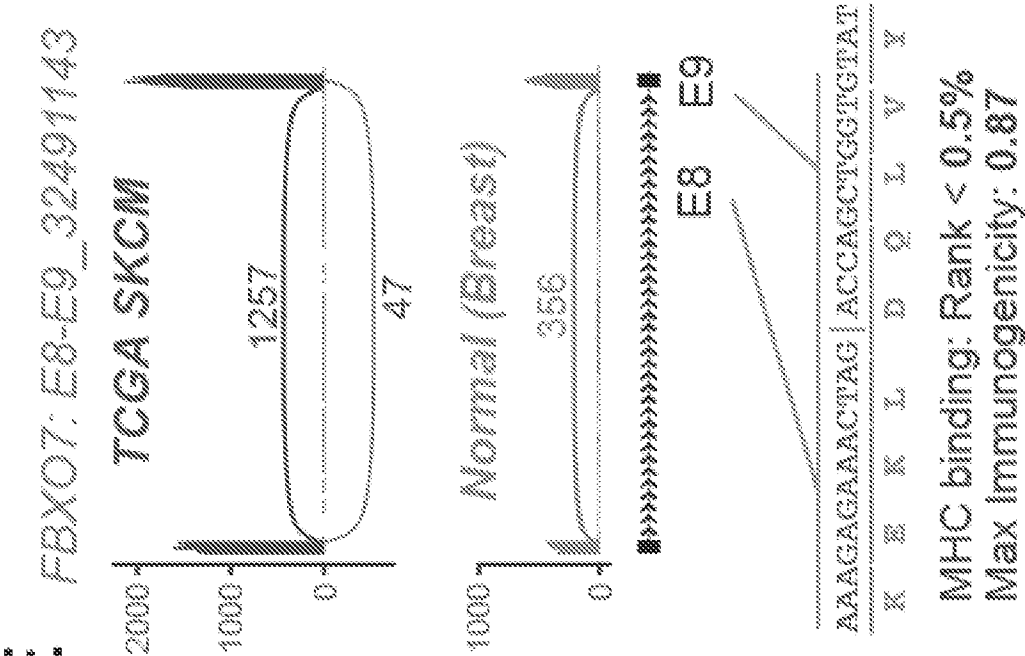


FIG. 7D



KEKLDQLVY

Pearson r:0.78

p-value:0.02

cosine:0.91

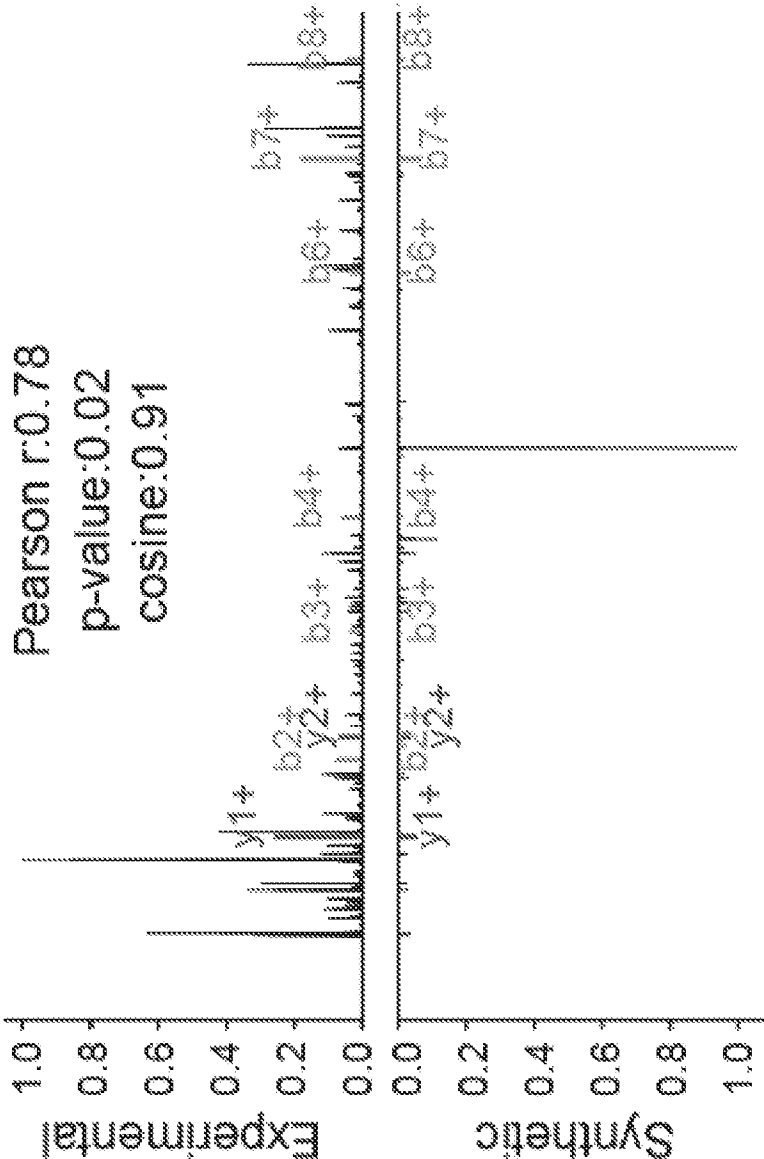


FIG. 7E

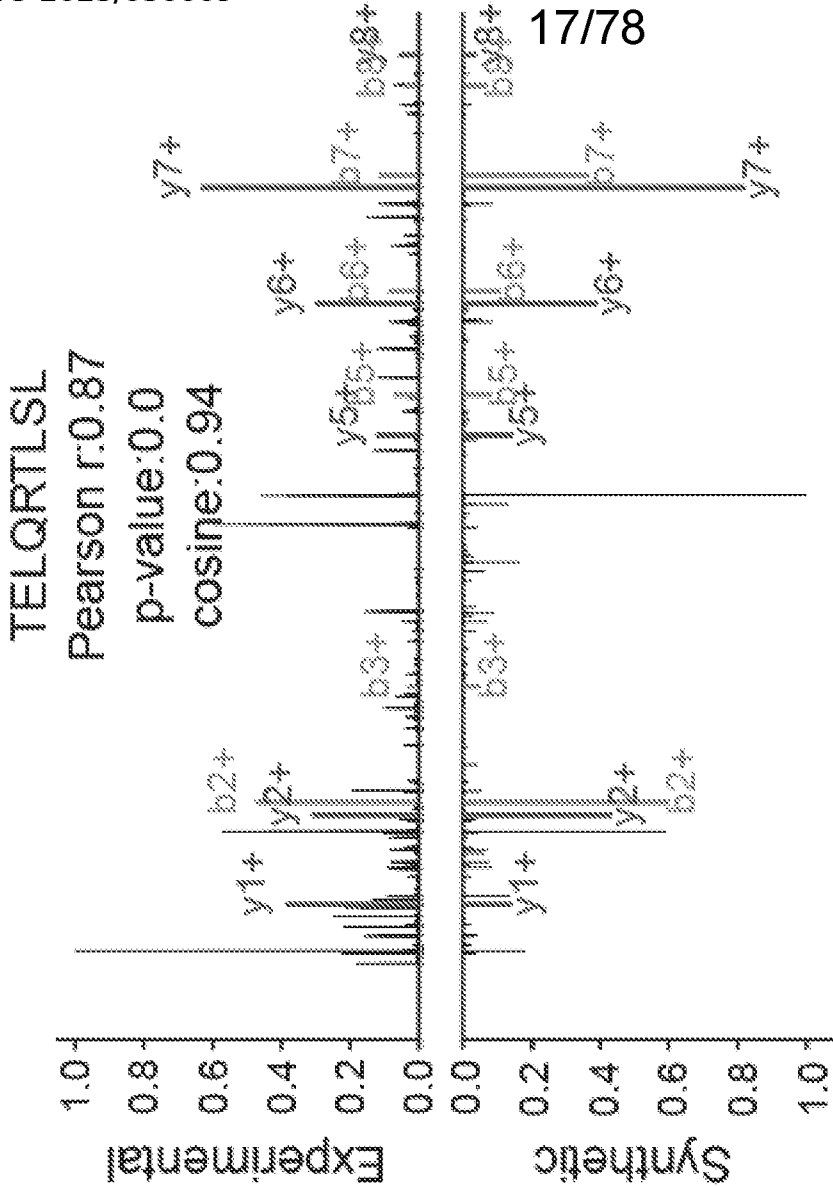
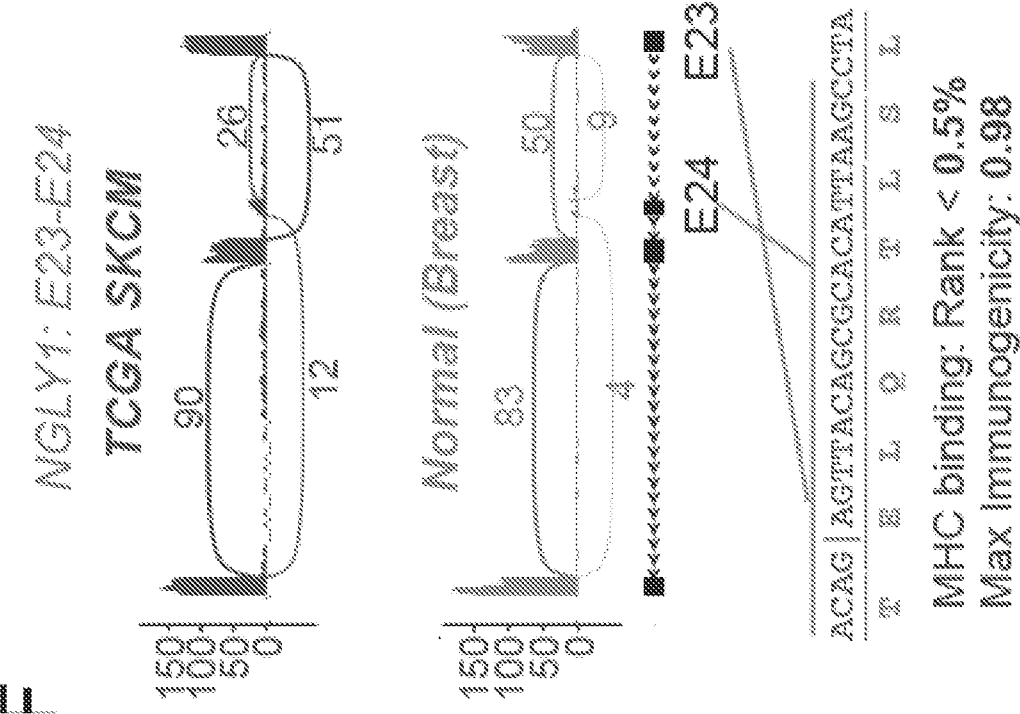
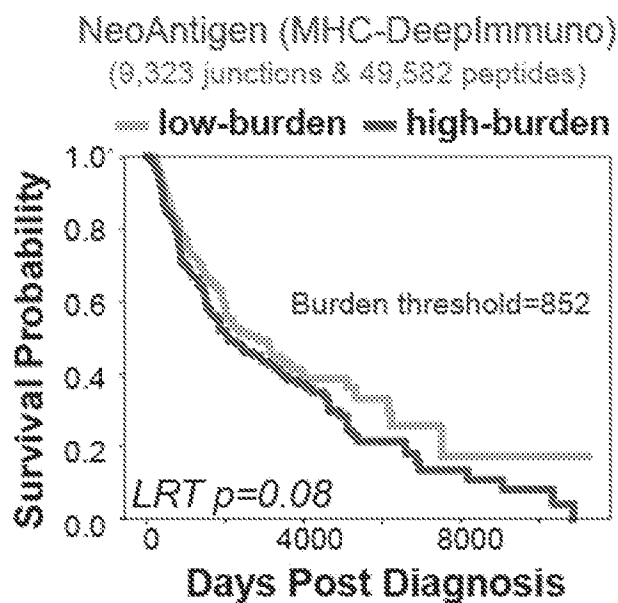
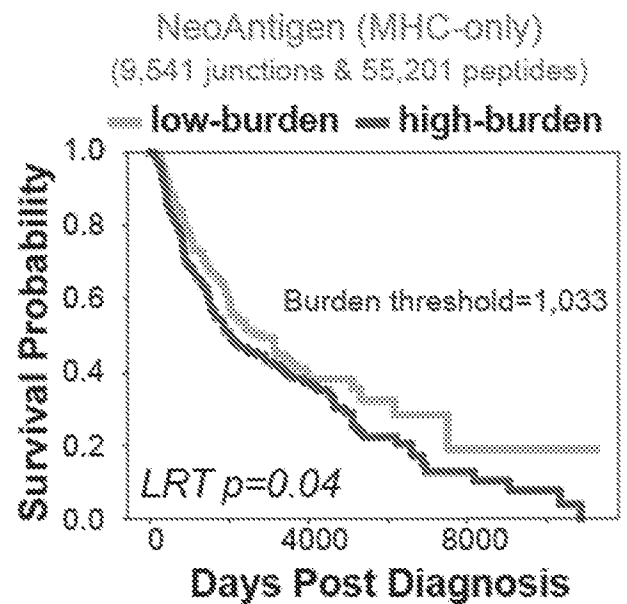
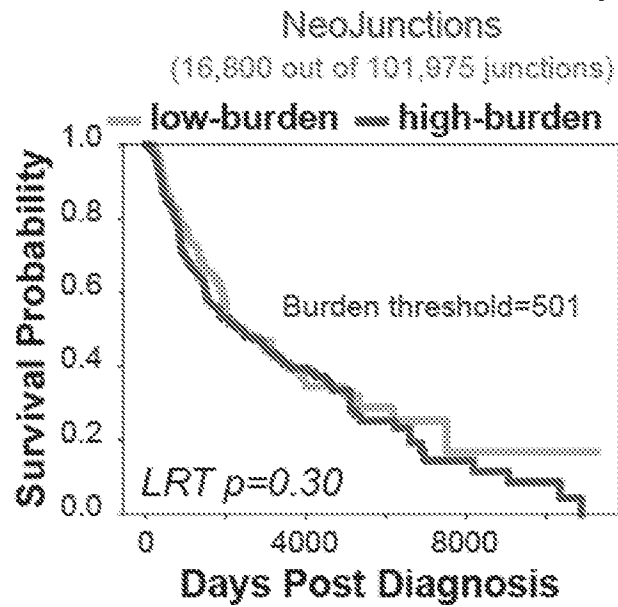
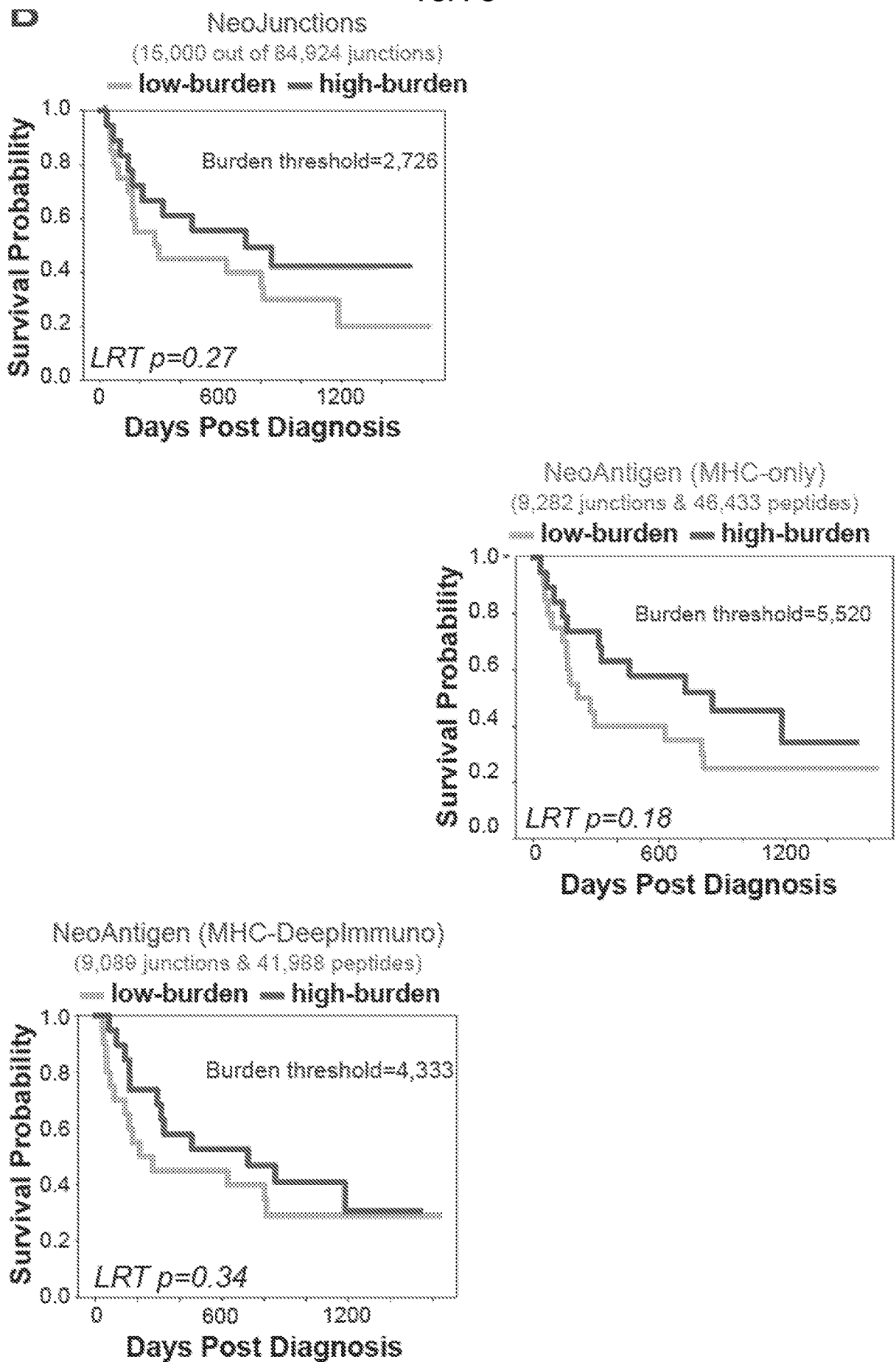


FIG. 7F

**FIG. 8A**

**FIG. 8B**

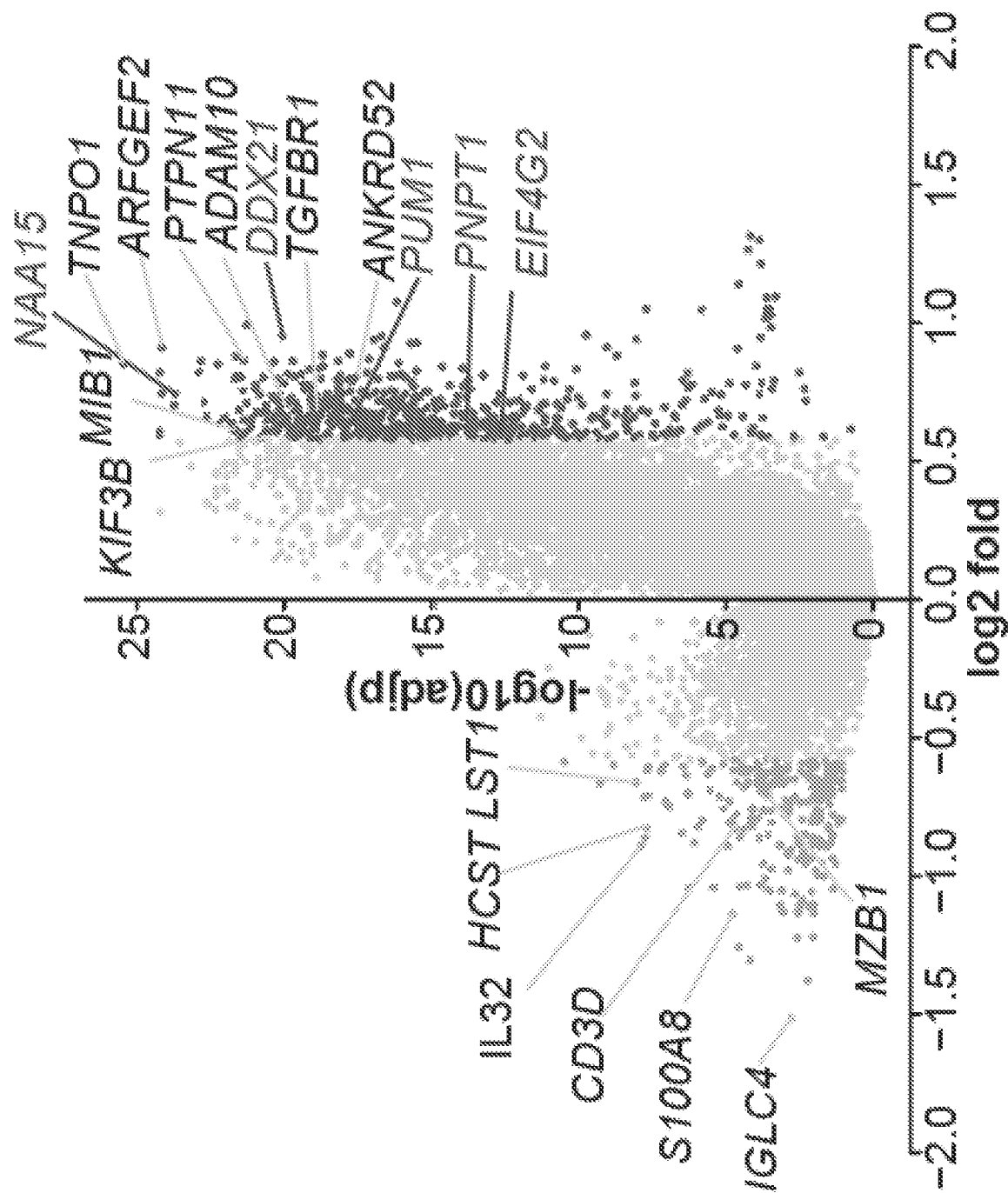


FIG. 8C

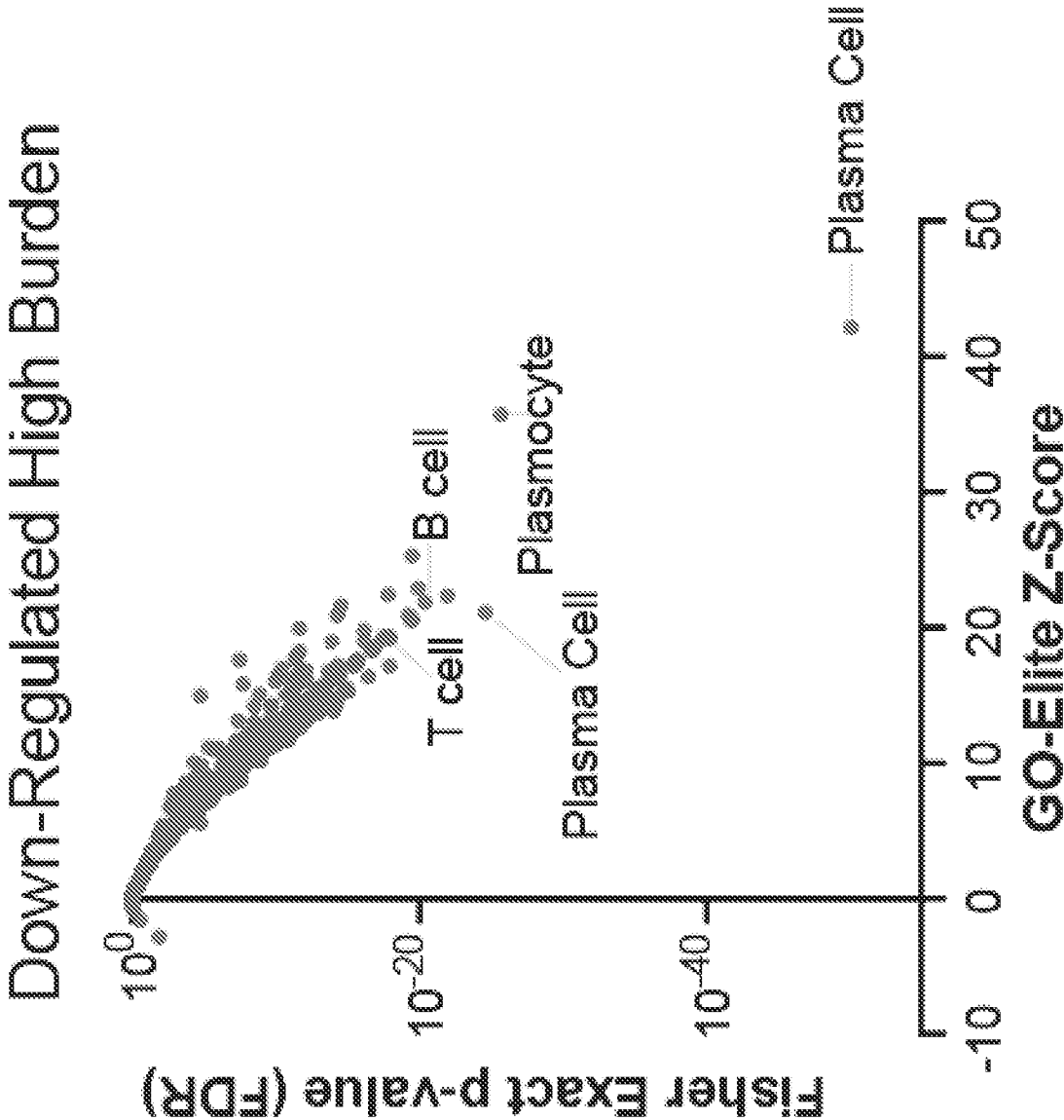


FIG. 8D

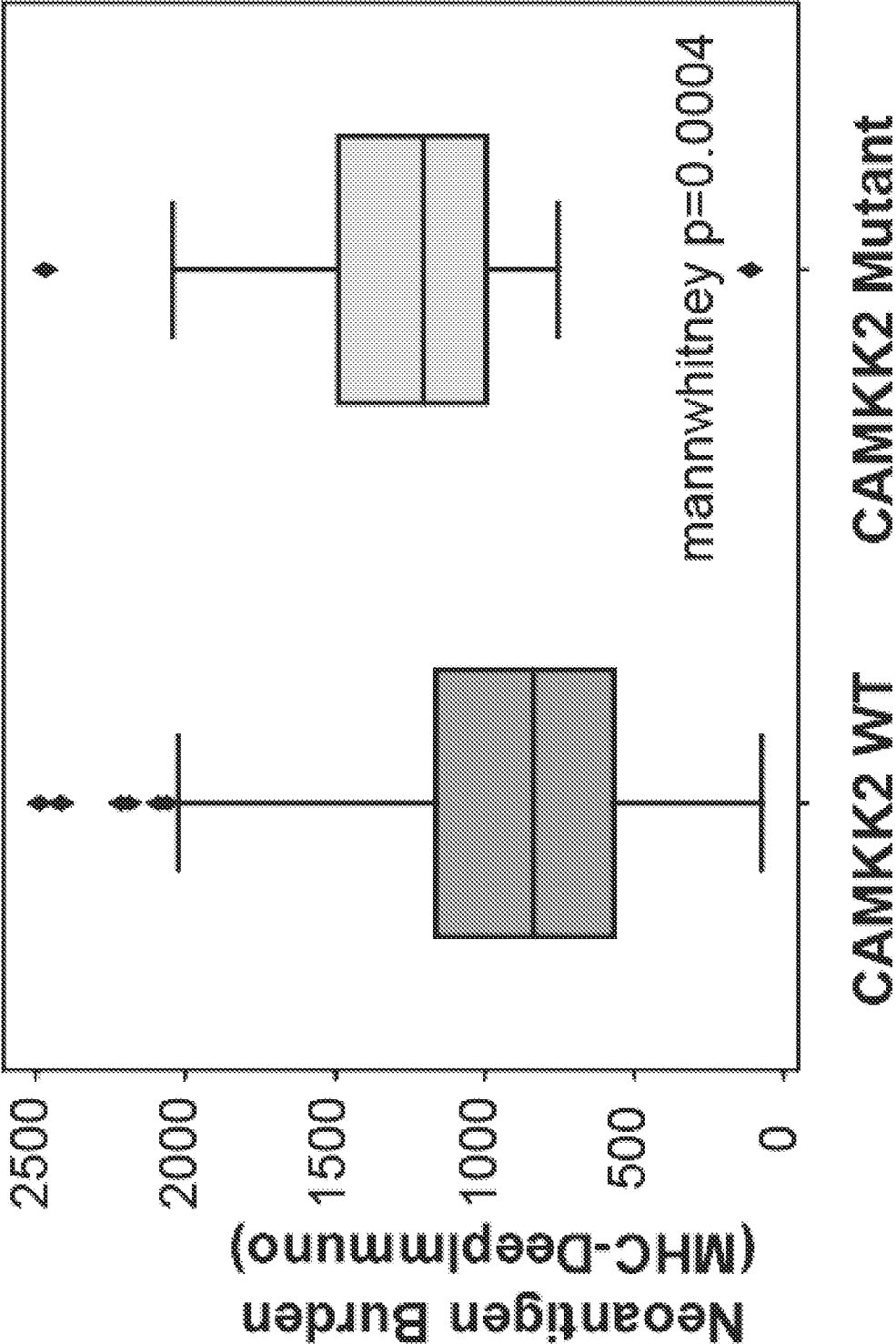


FIG. 8E

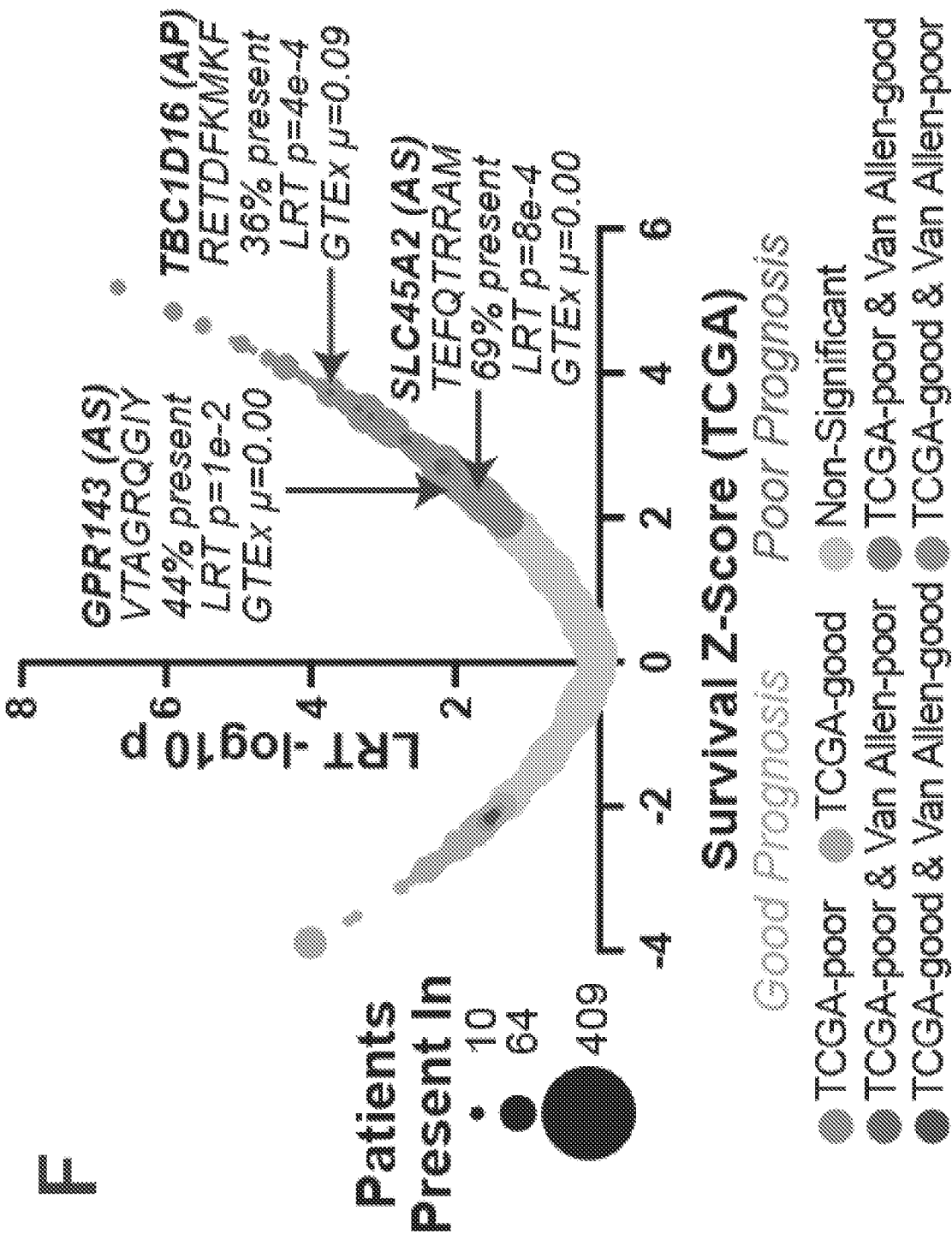


FIG. 8F

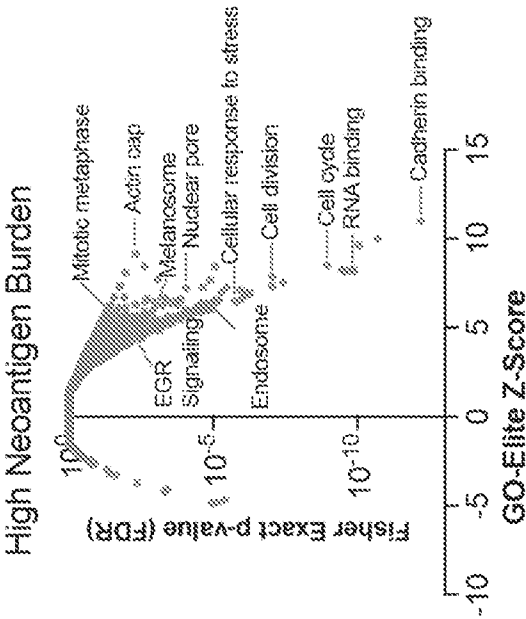


FIG. 9A

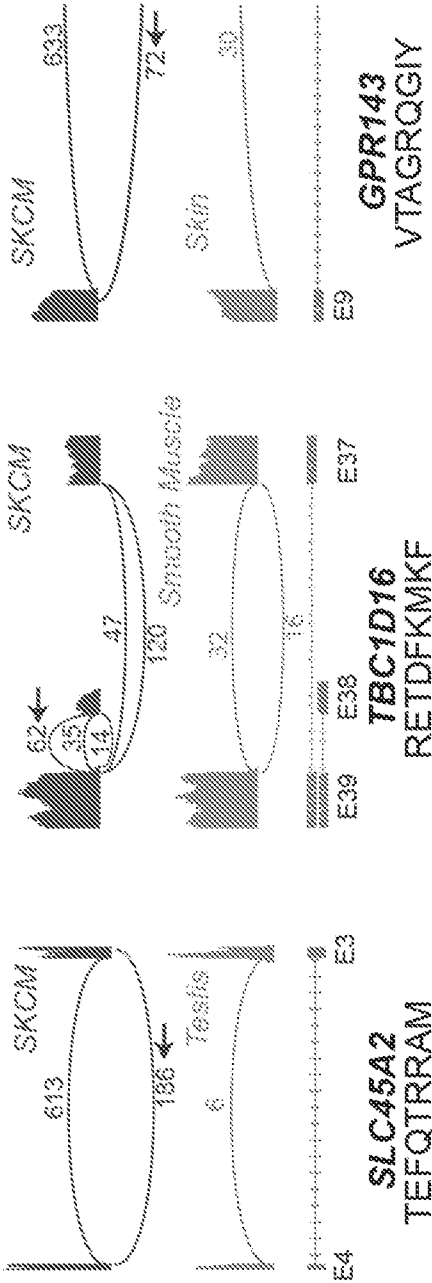
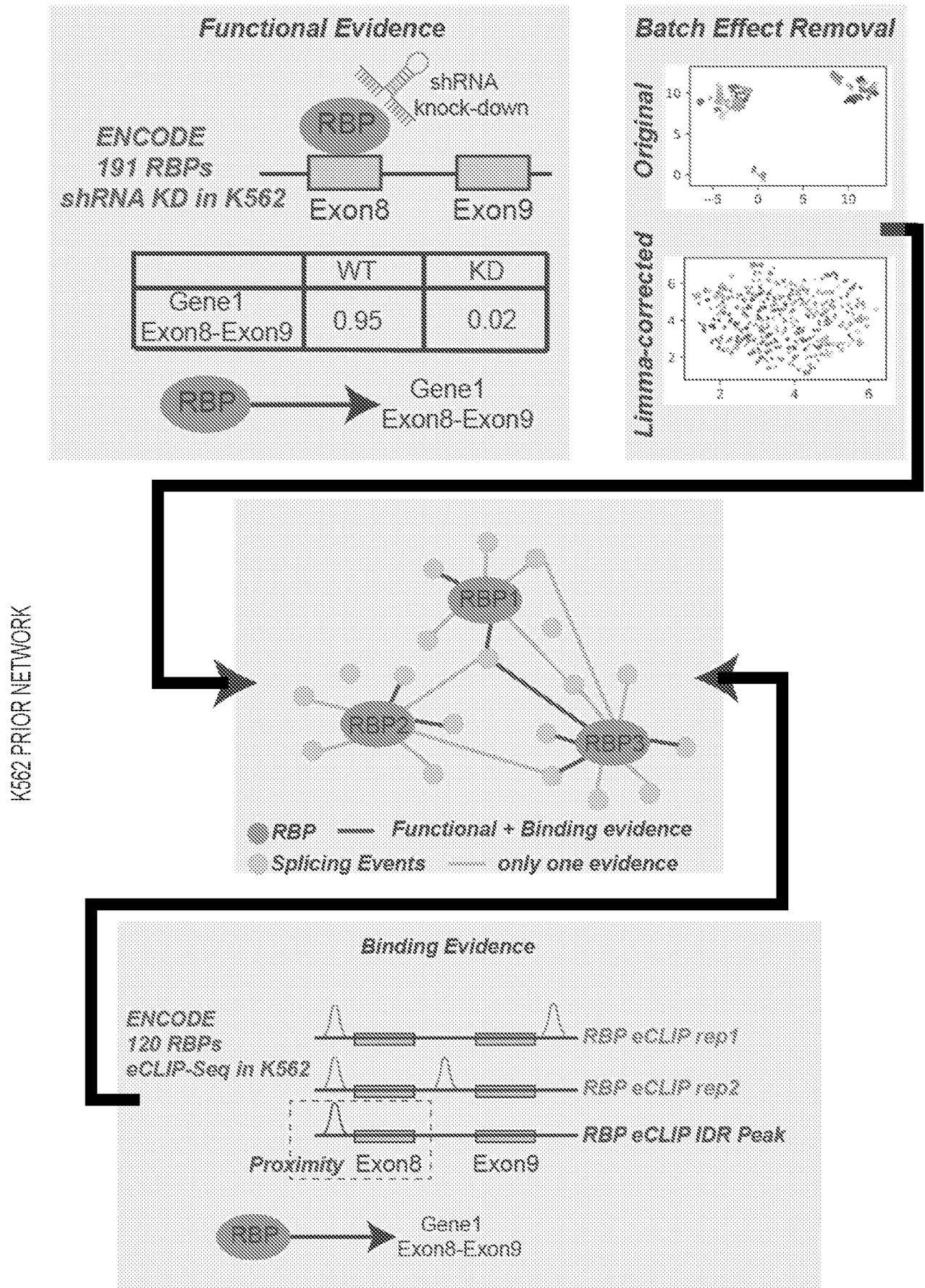


FIG. 9B

25/78

RNA-SPRINT - RNA-BASED SPICING PROTEIN ACTIVITY INFERENCE FROM MULTIVARIATE DECISION TREES

**FIG. 10A**

RNA-SPRINT - RNA-BASED SPICING PROTEIN ACTIVITY INFERENCE FROM MULTIVARIATE DCISION TREES

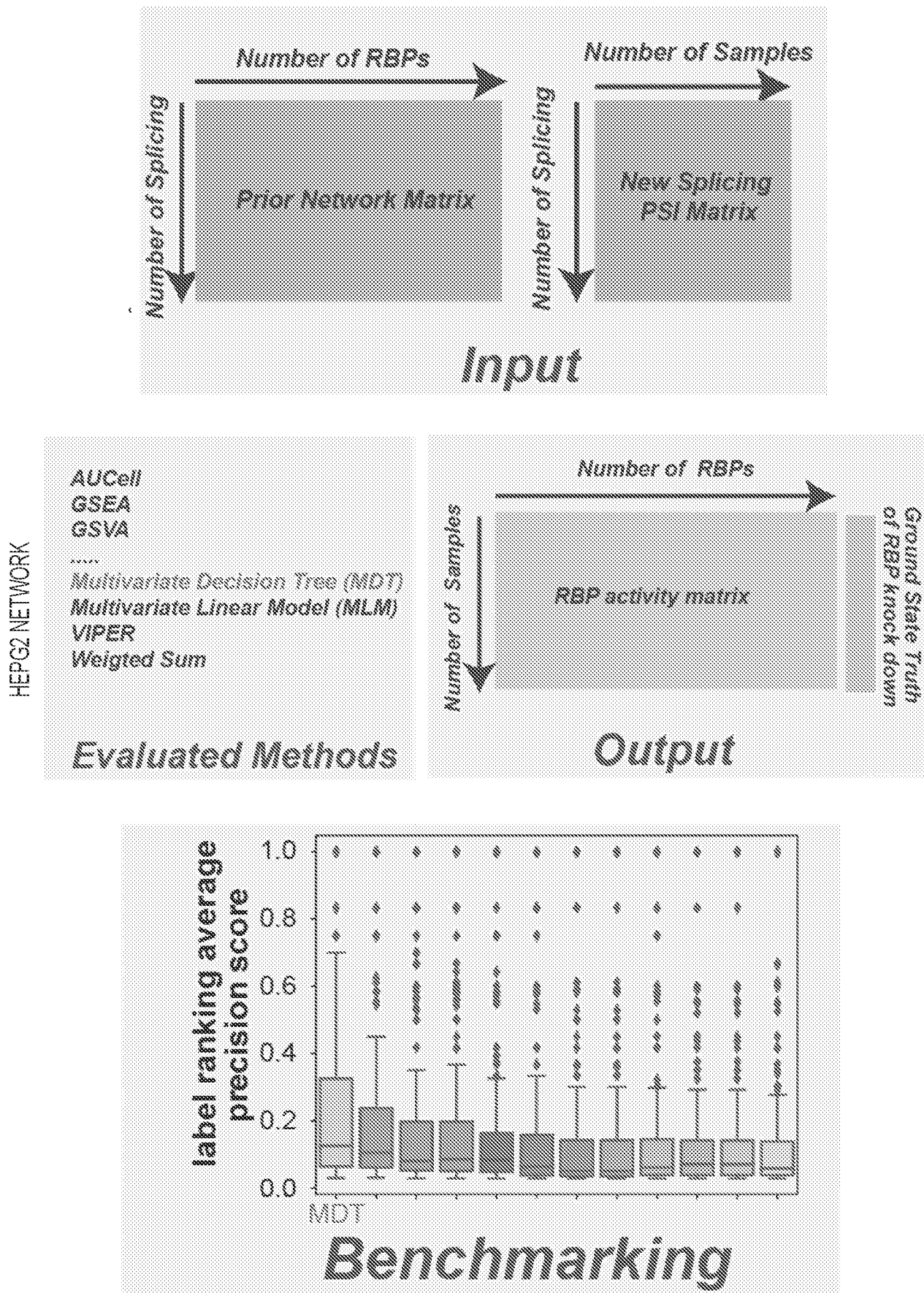


FIG. 10A (CONT.)

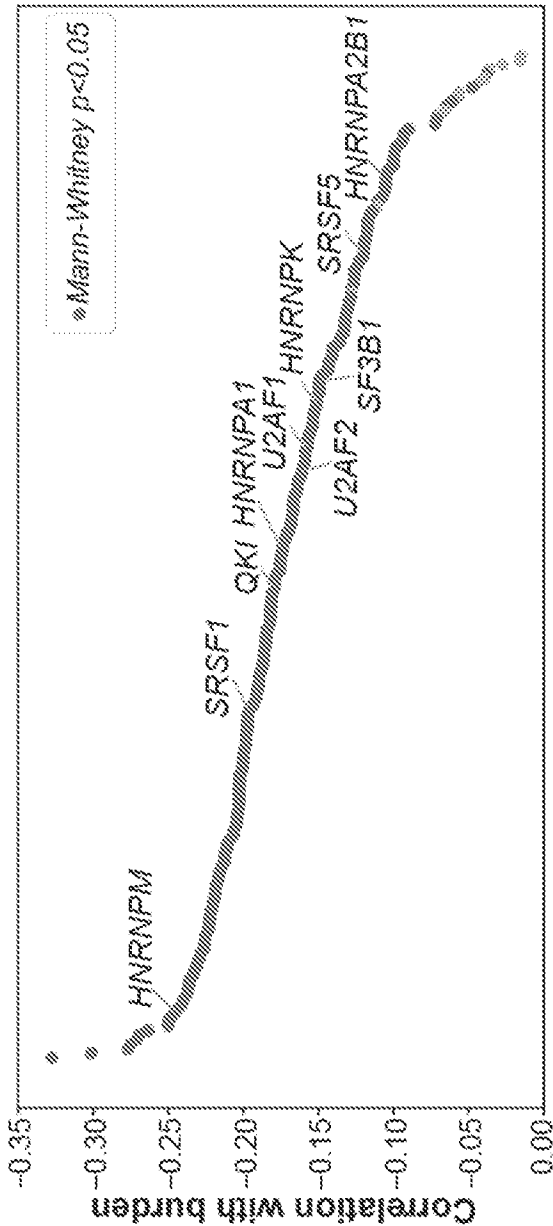


FIG. 10B

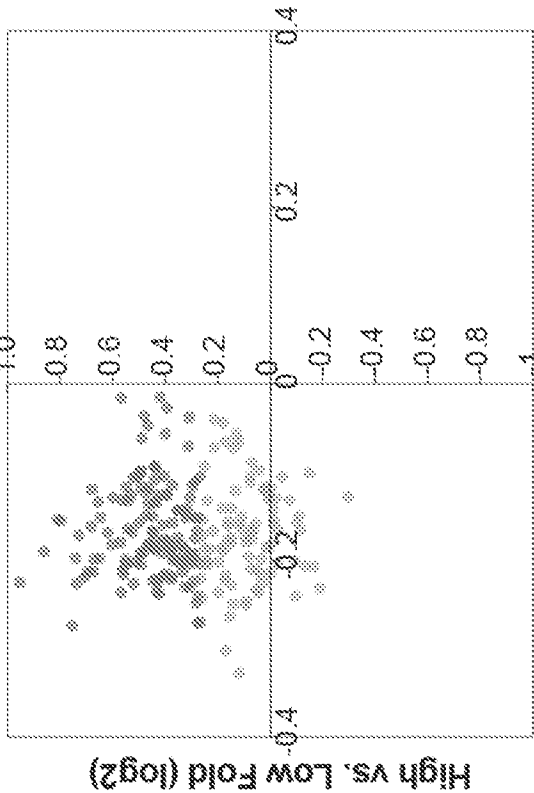


FIG. 10C

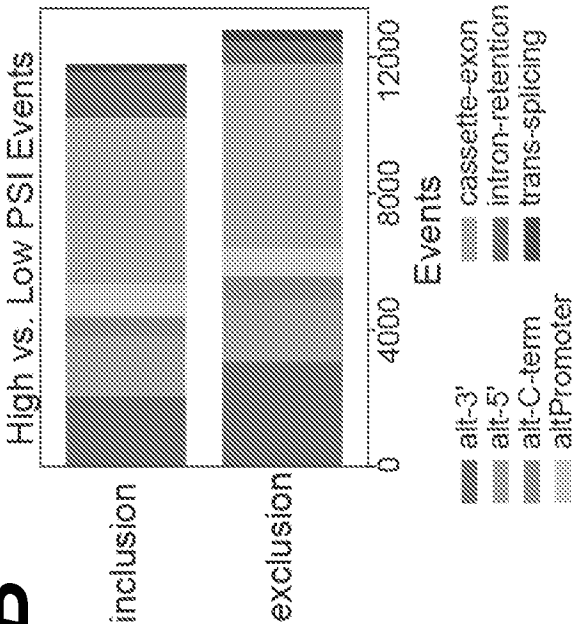


FIG. 10D

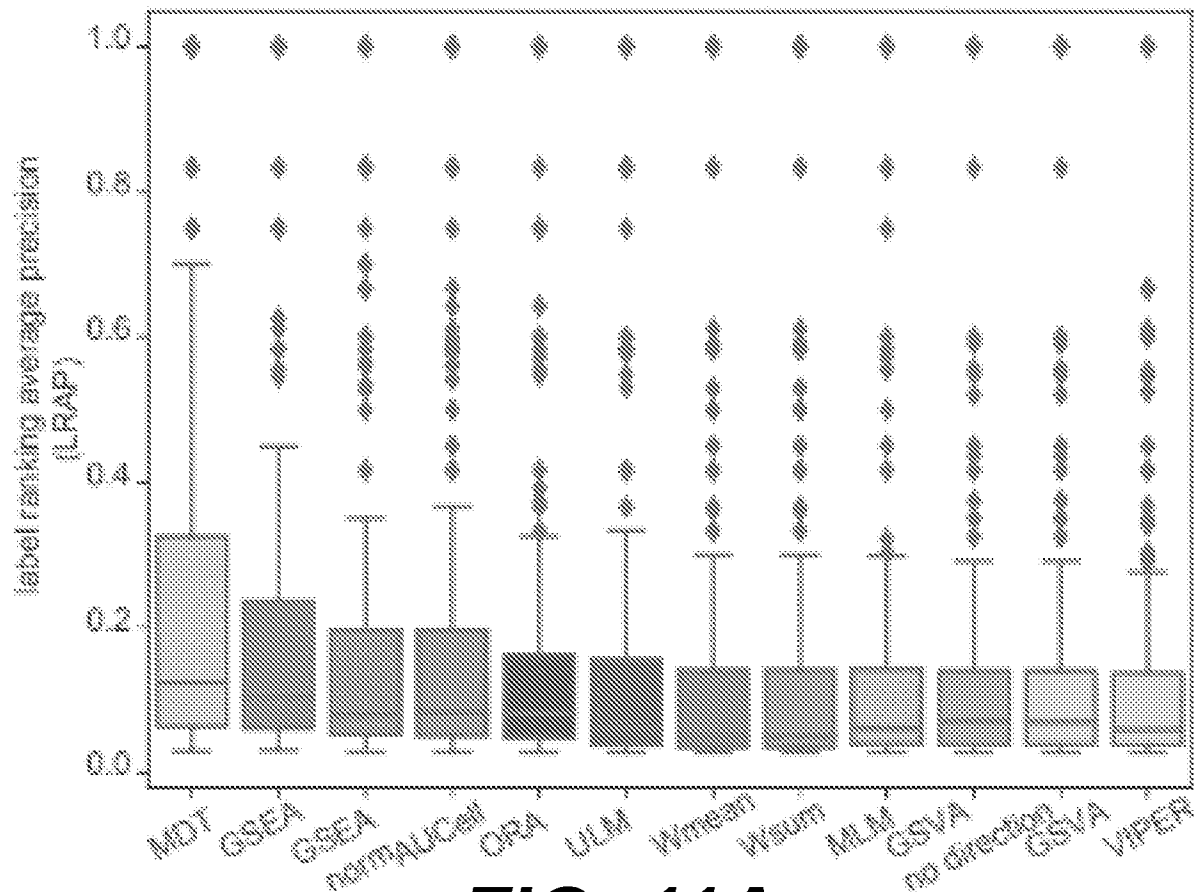


FIG. 11A

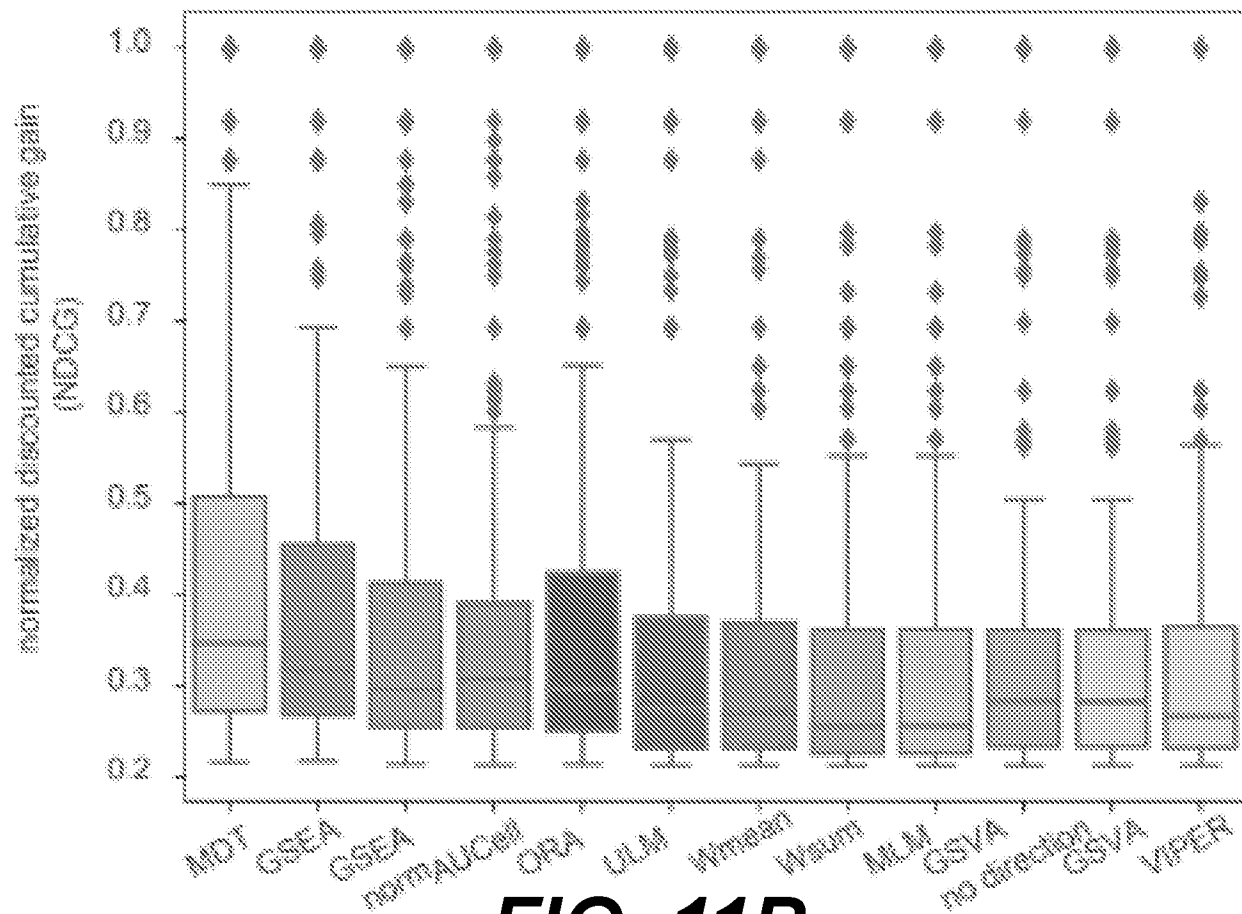


FIG. 11B

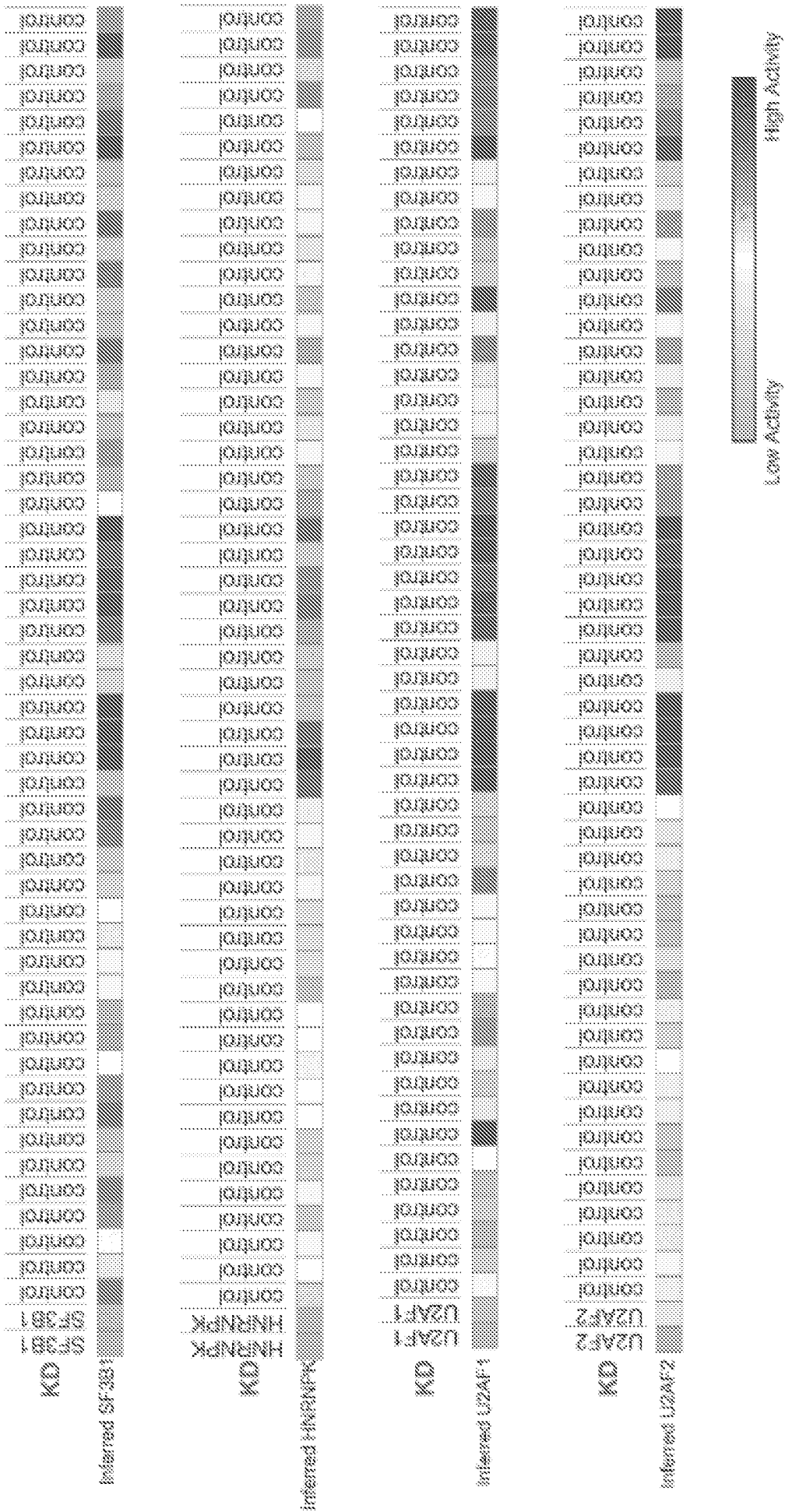


FIG. 11C

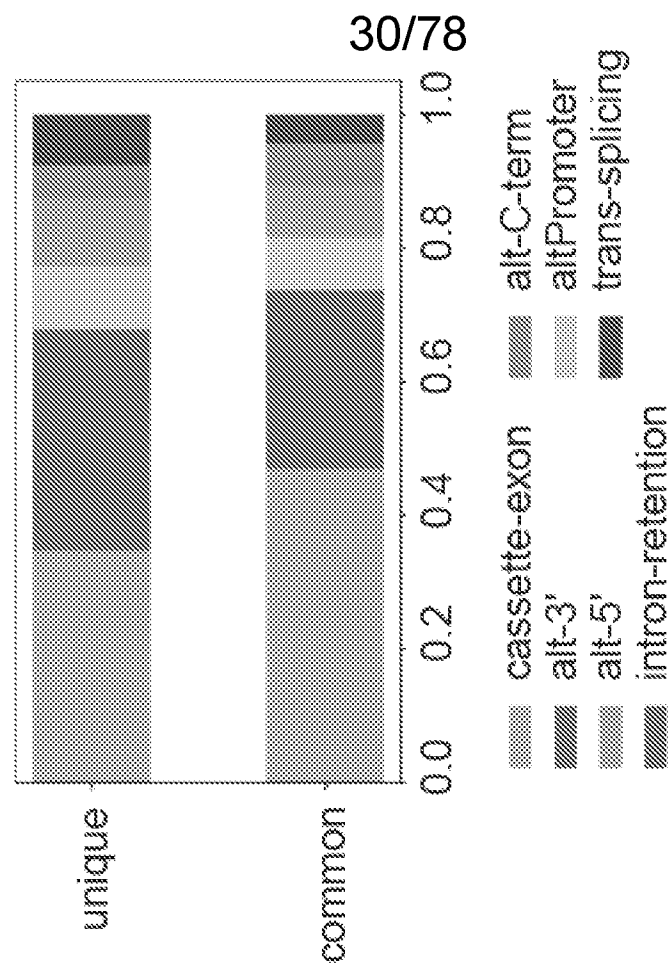
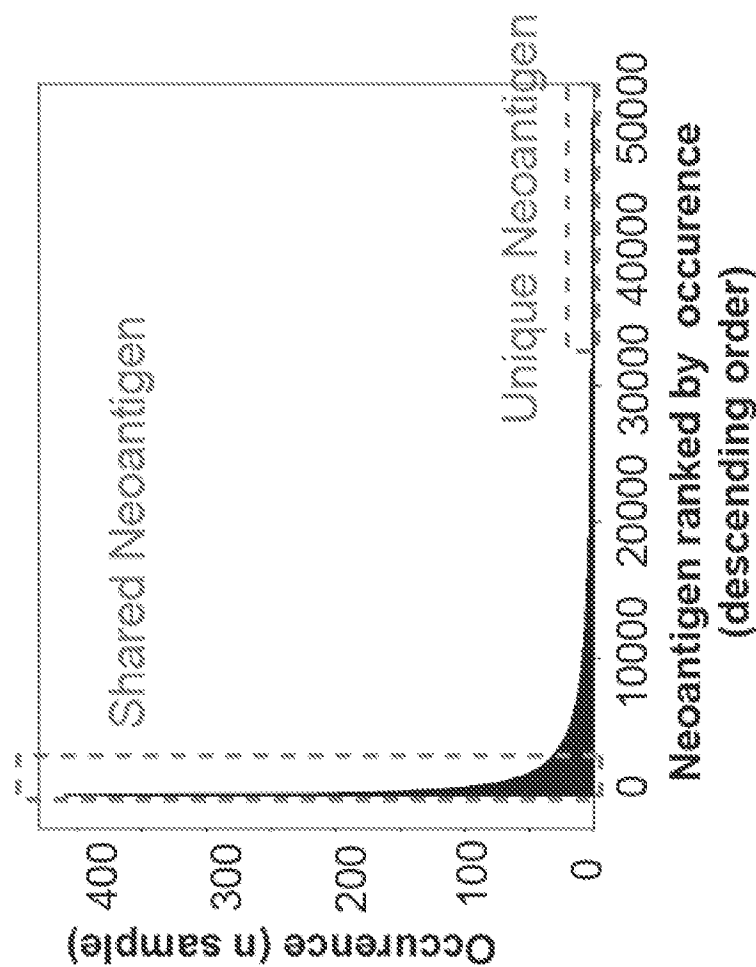


FIG. 12A

FIG. 12B

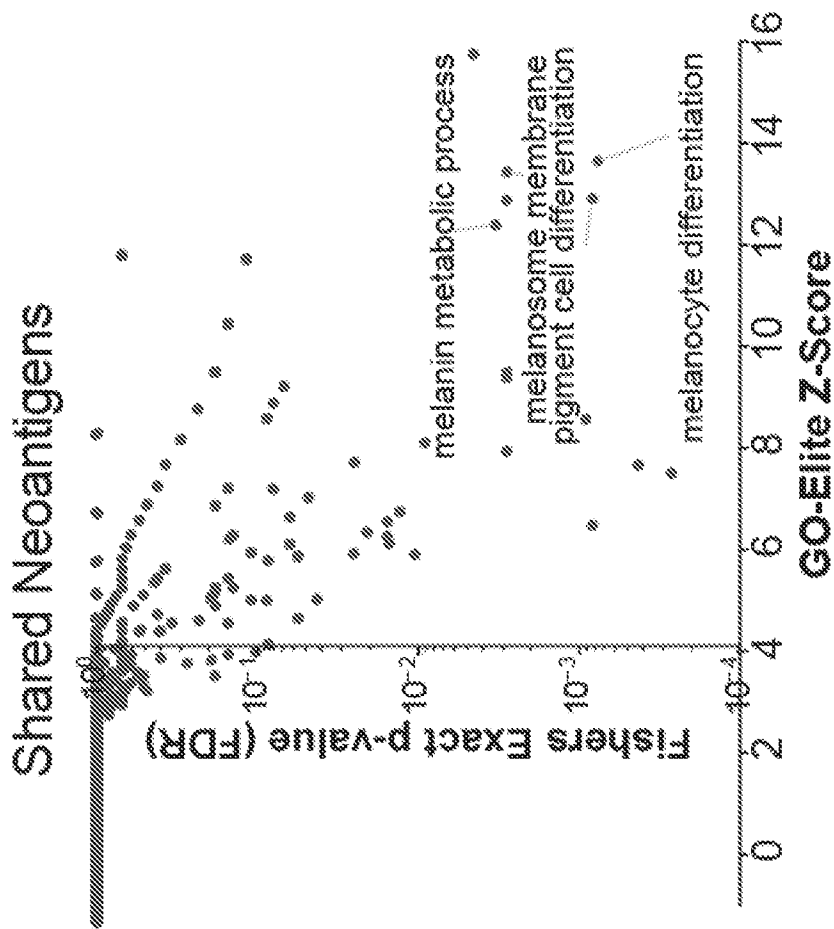


FIG. 12C

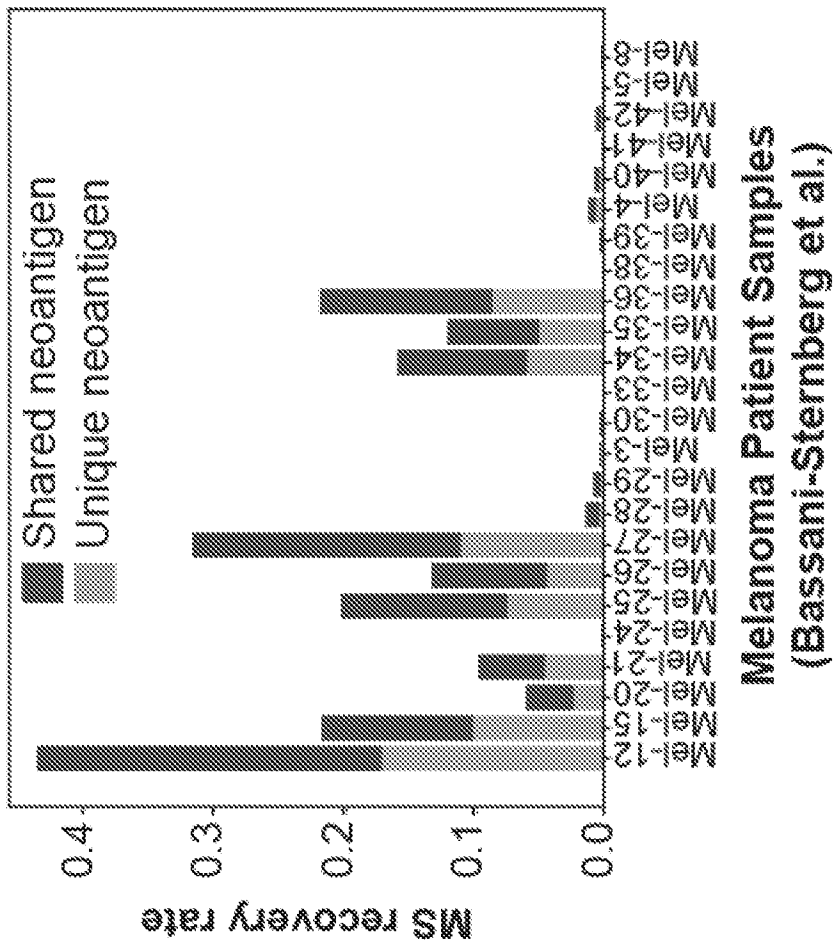


FIG. 12D

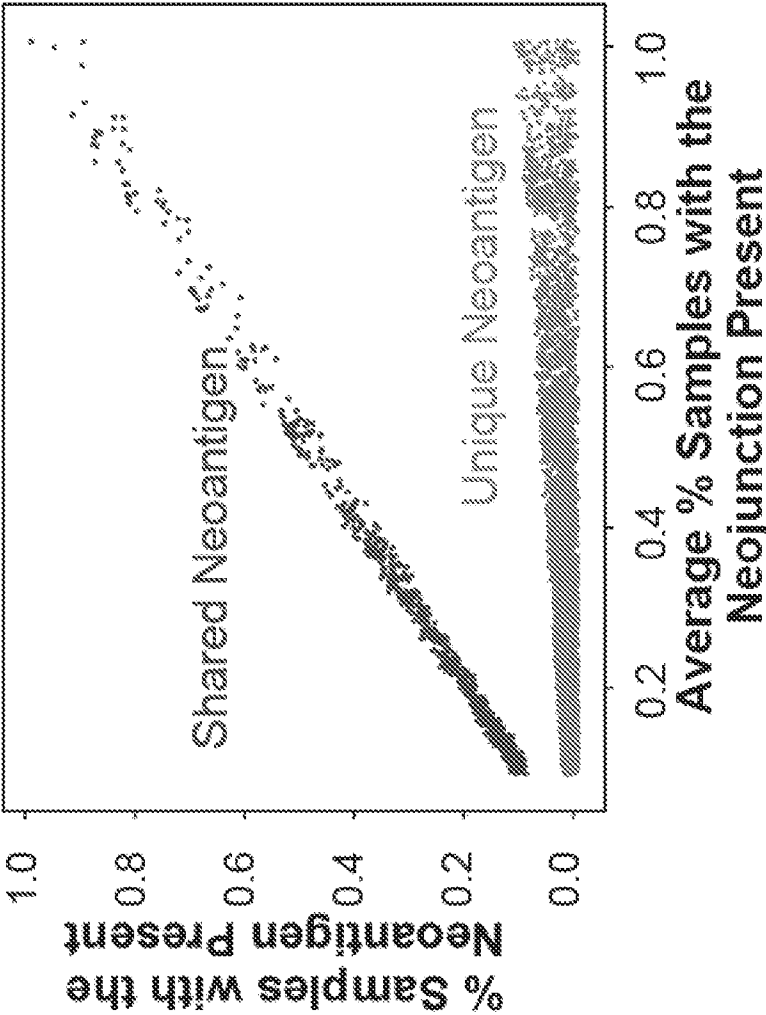


FIG. 12F

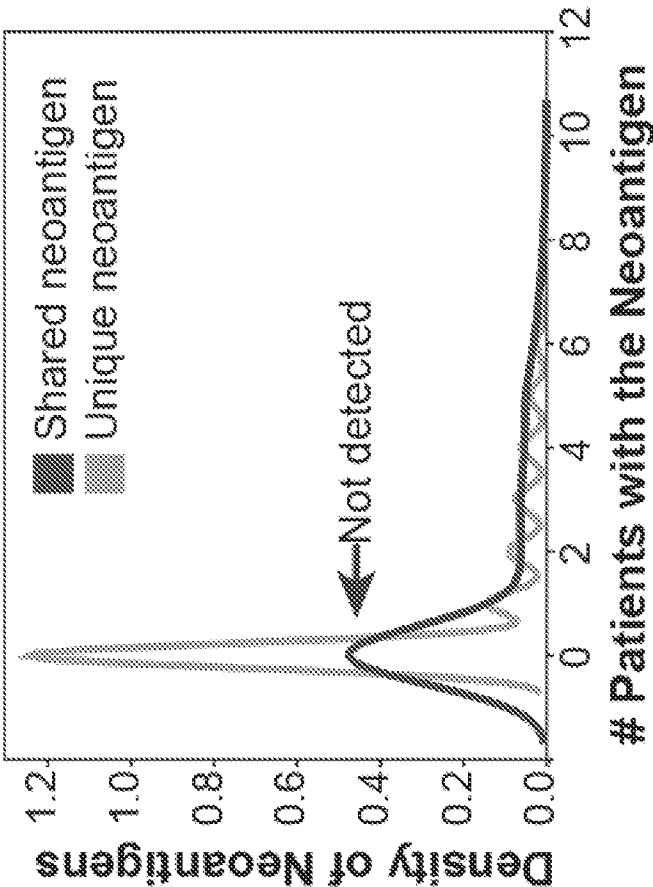
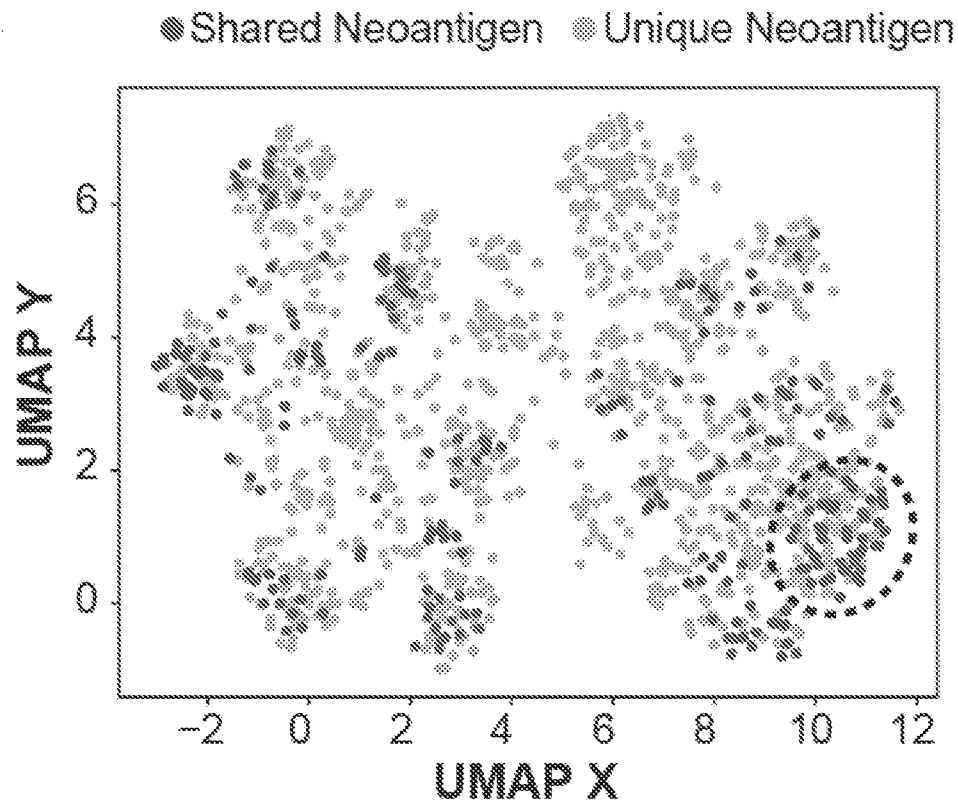
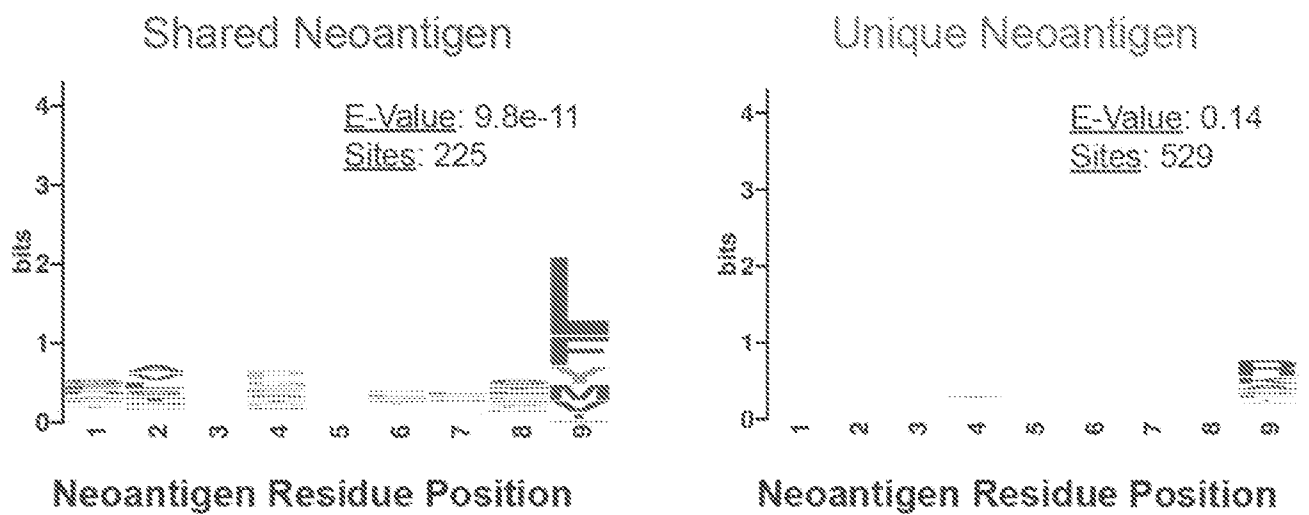
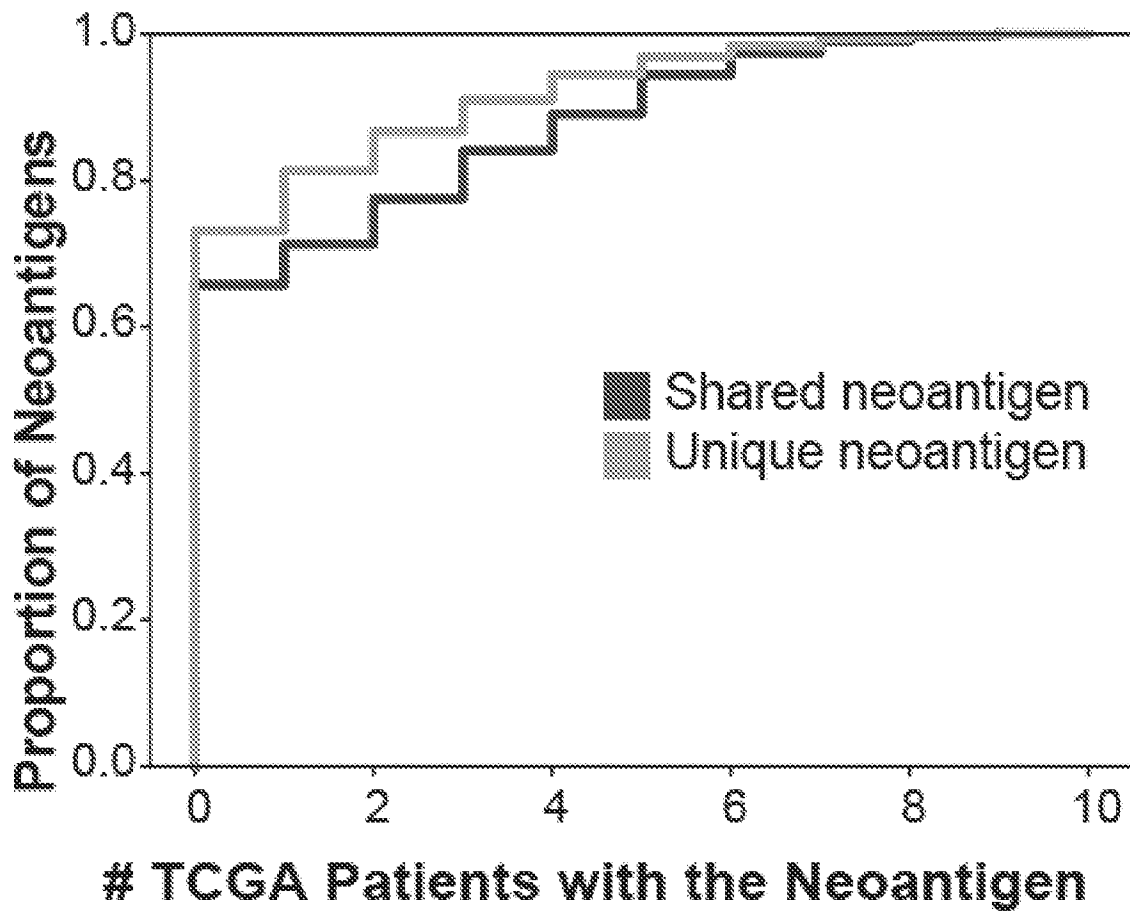


FIG. 12E

33/78

**FIG. 12G****FIG. 12H**

34/78

**FIG. 13A**

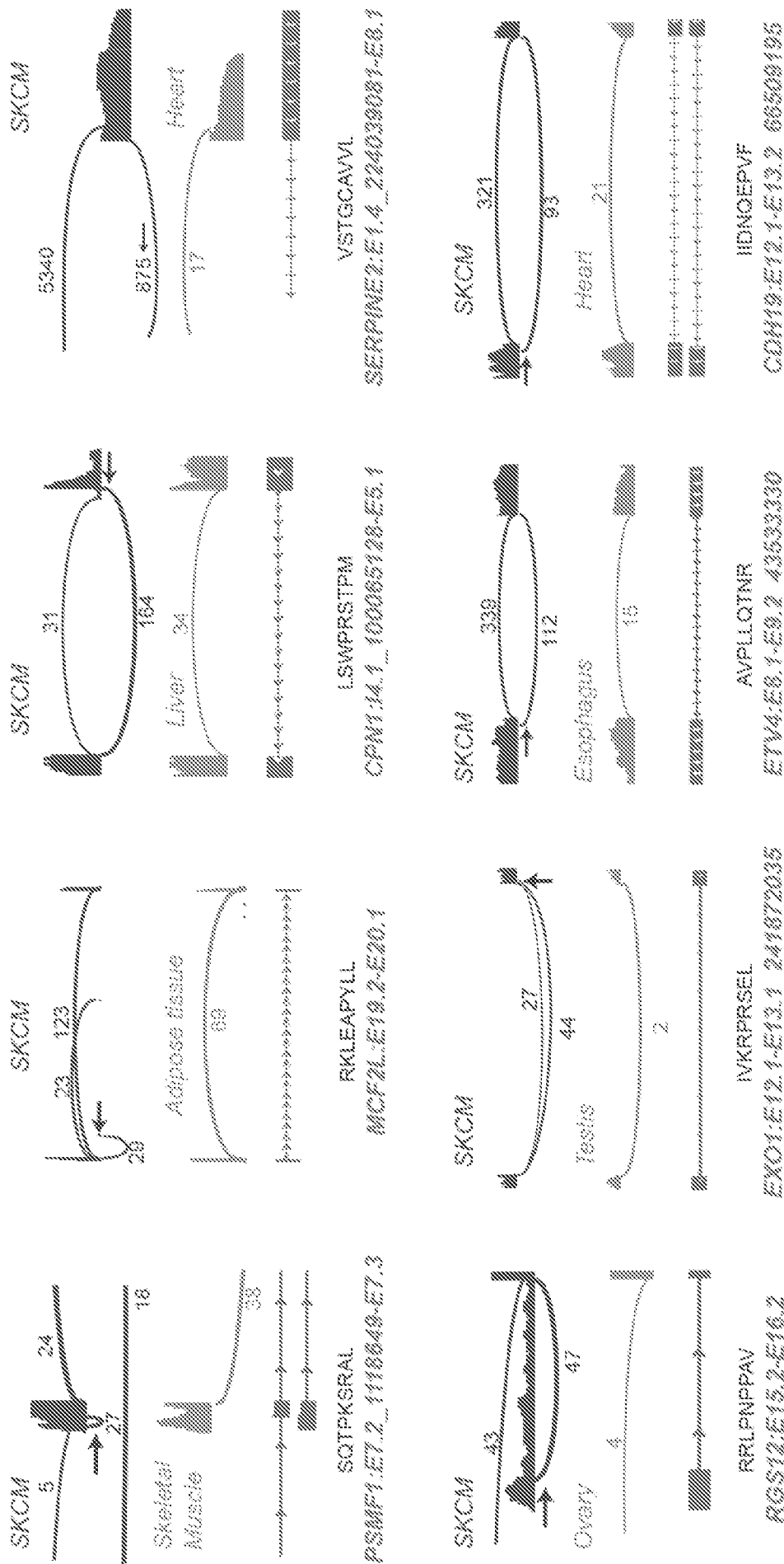


FIG. 13B

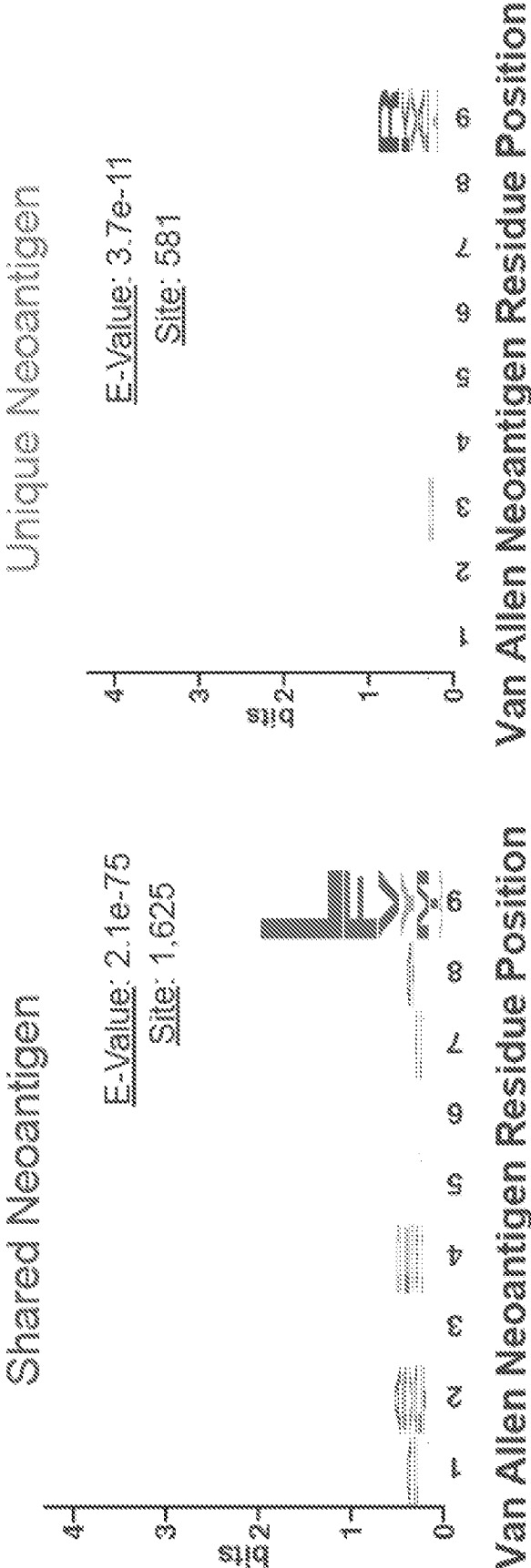


FIG. 13C

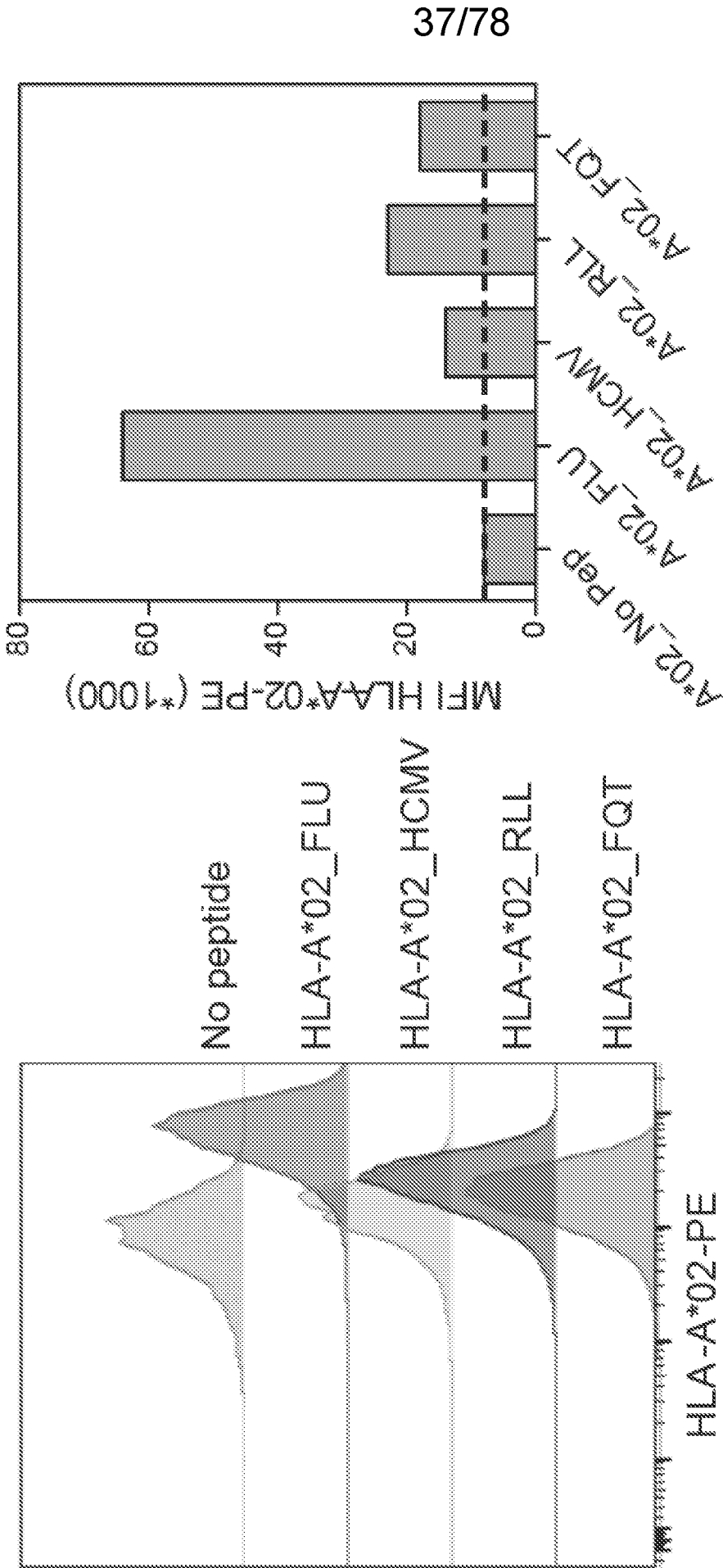


FIG. 14B

FIG. 14A

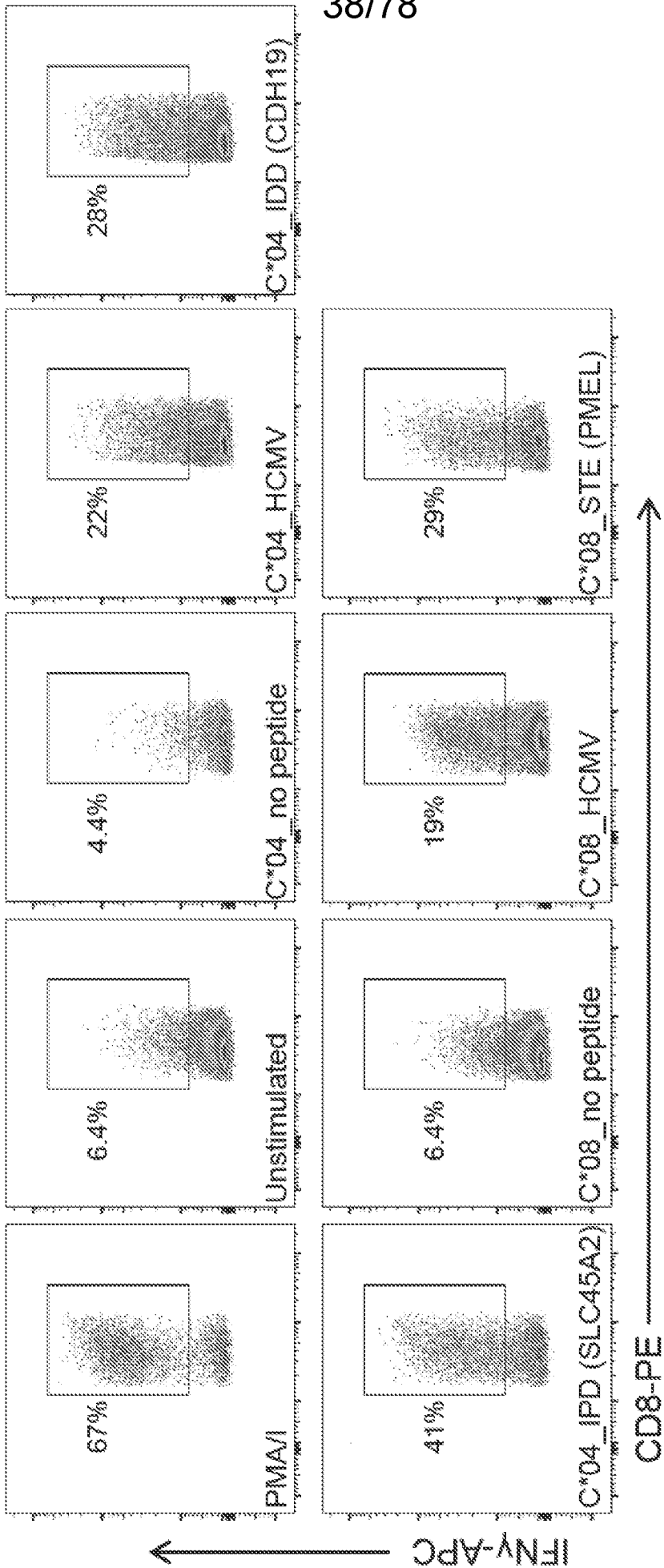


FIG. 14C

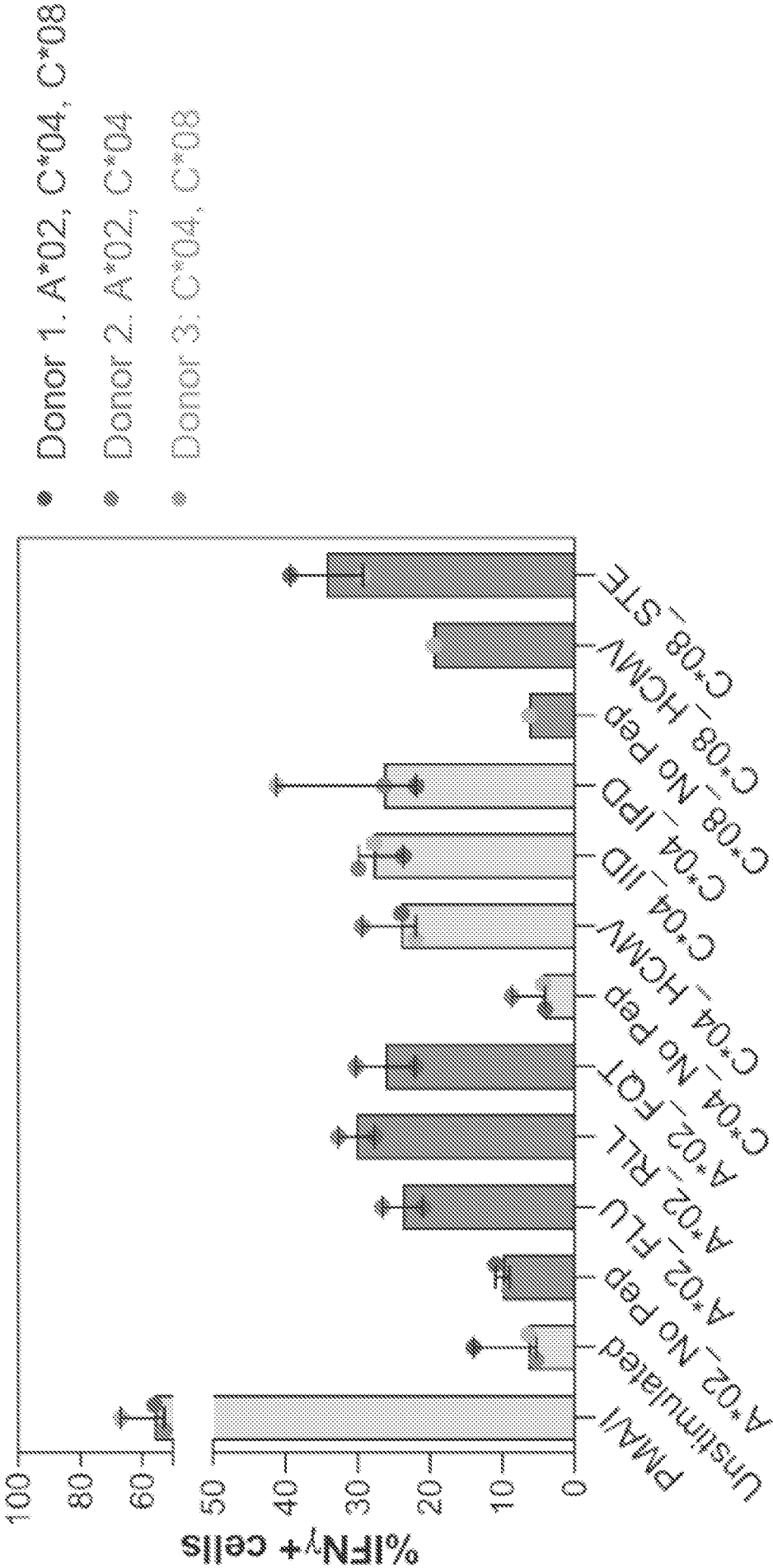
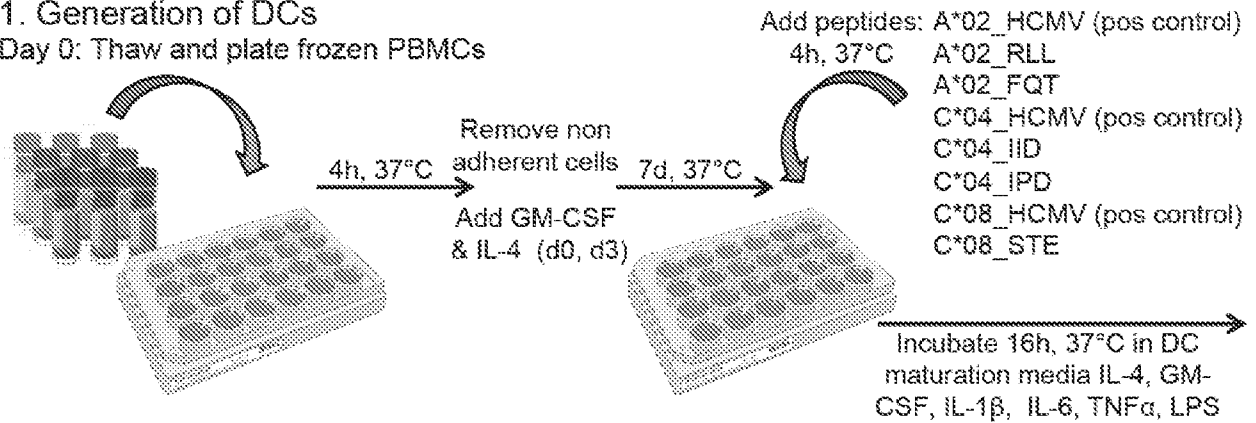


FIG. 14D

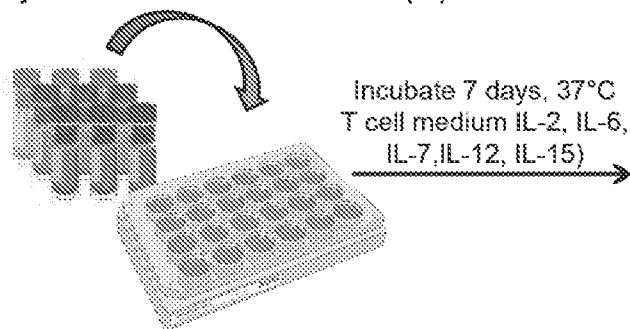
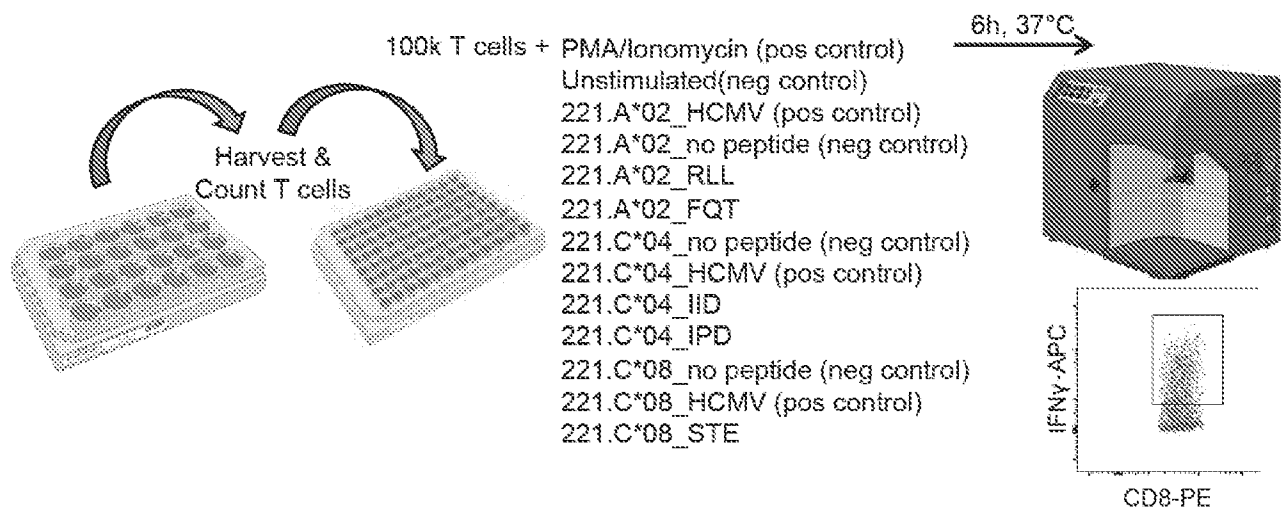
40/78

1. Generation of DCs

Day 0: Thaw and plate frozen PBMCs

**2. T cell stimulation (3x with fresh DCs)**

Day 7: Thaw PBMCs and add to peptide loaded mature DC

**3. Test immunogenicity of expanded T cells**Day 28: Harvest primed T cells and test IFN γ production in response to peptide loaded single antigen cell lines (221).**FIG. 15**

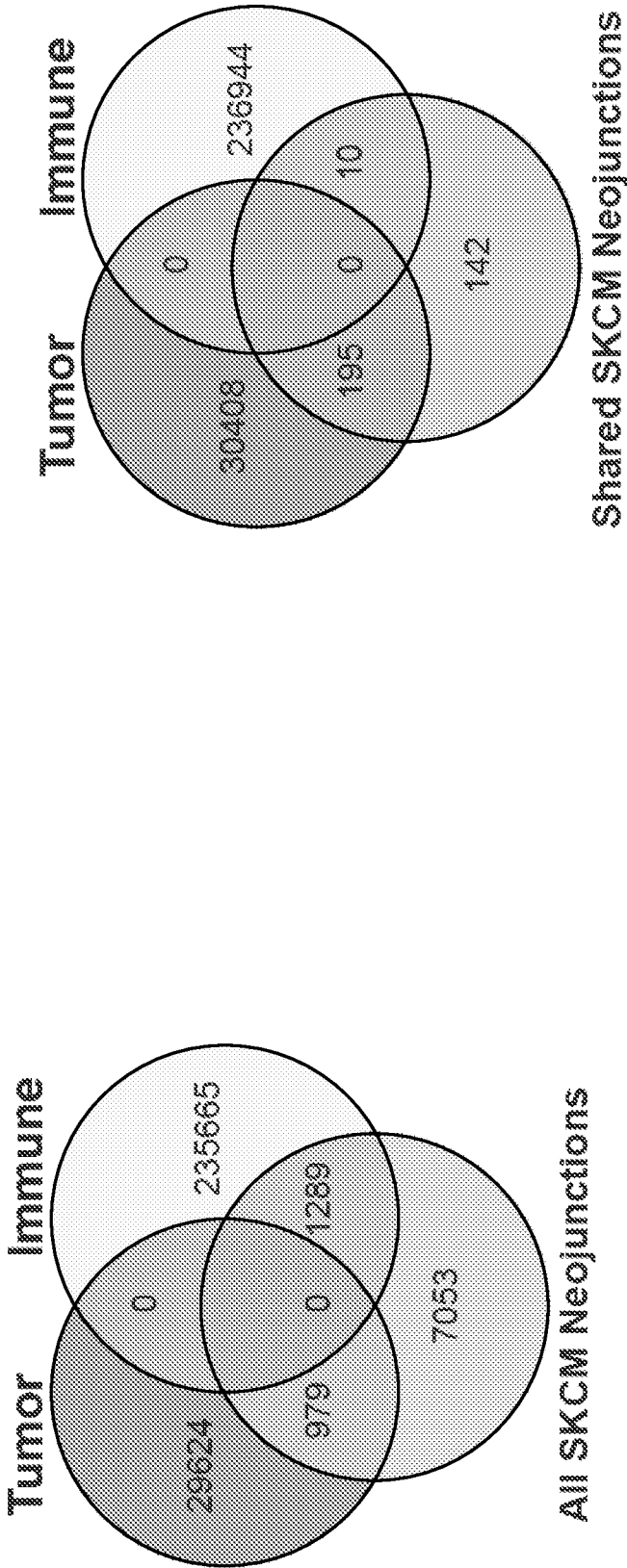


FIG. 16A

FIG. 16B

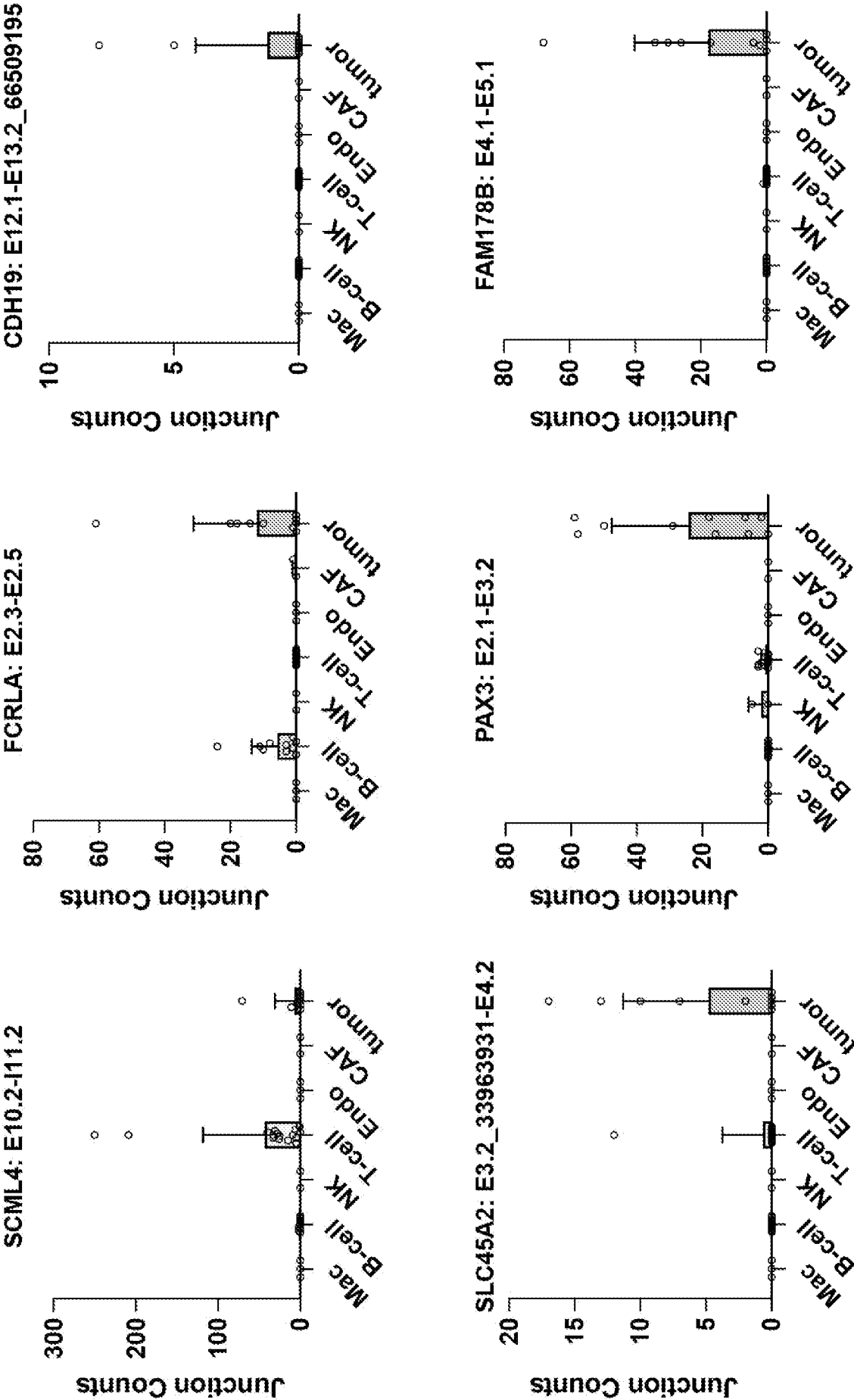


FIG. 16C

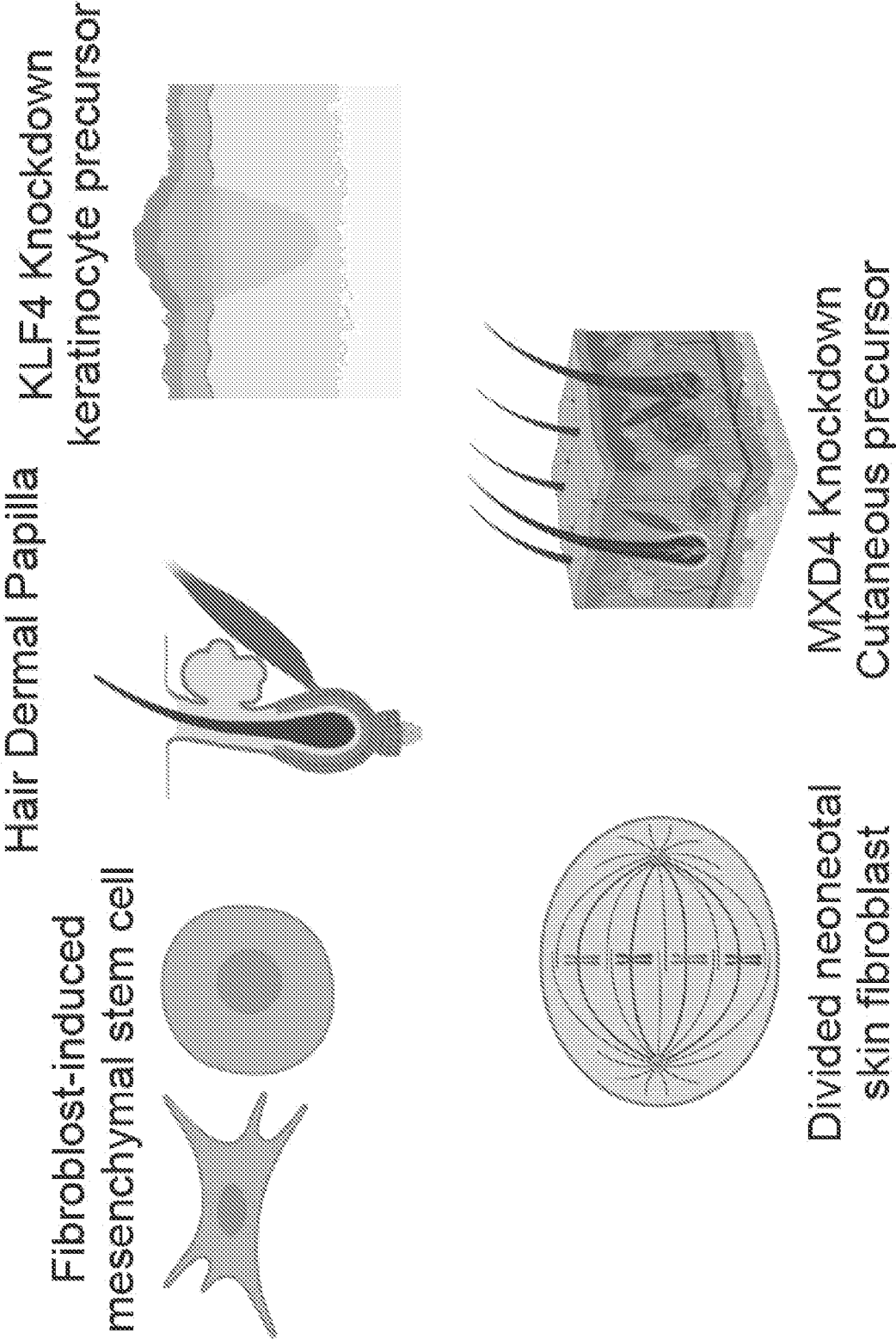


FIG. 17A

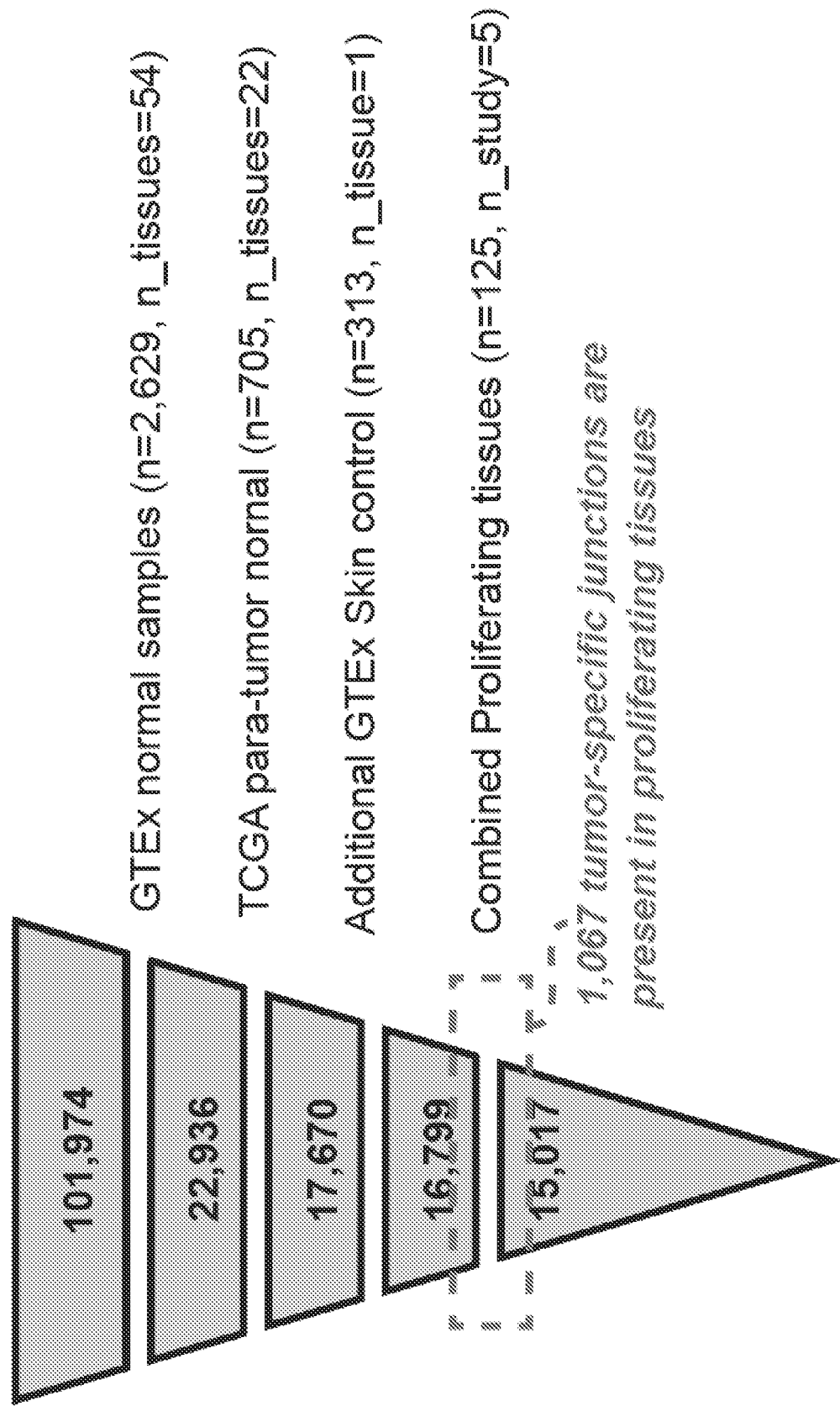


FIG. 17B

45/78

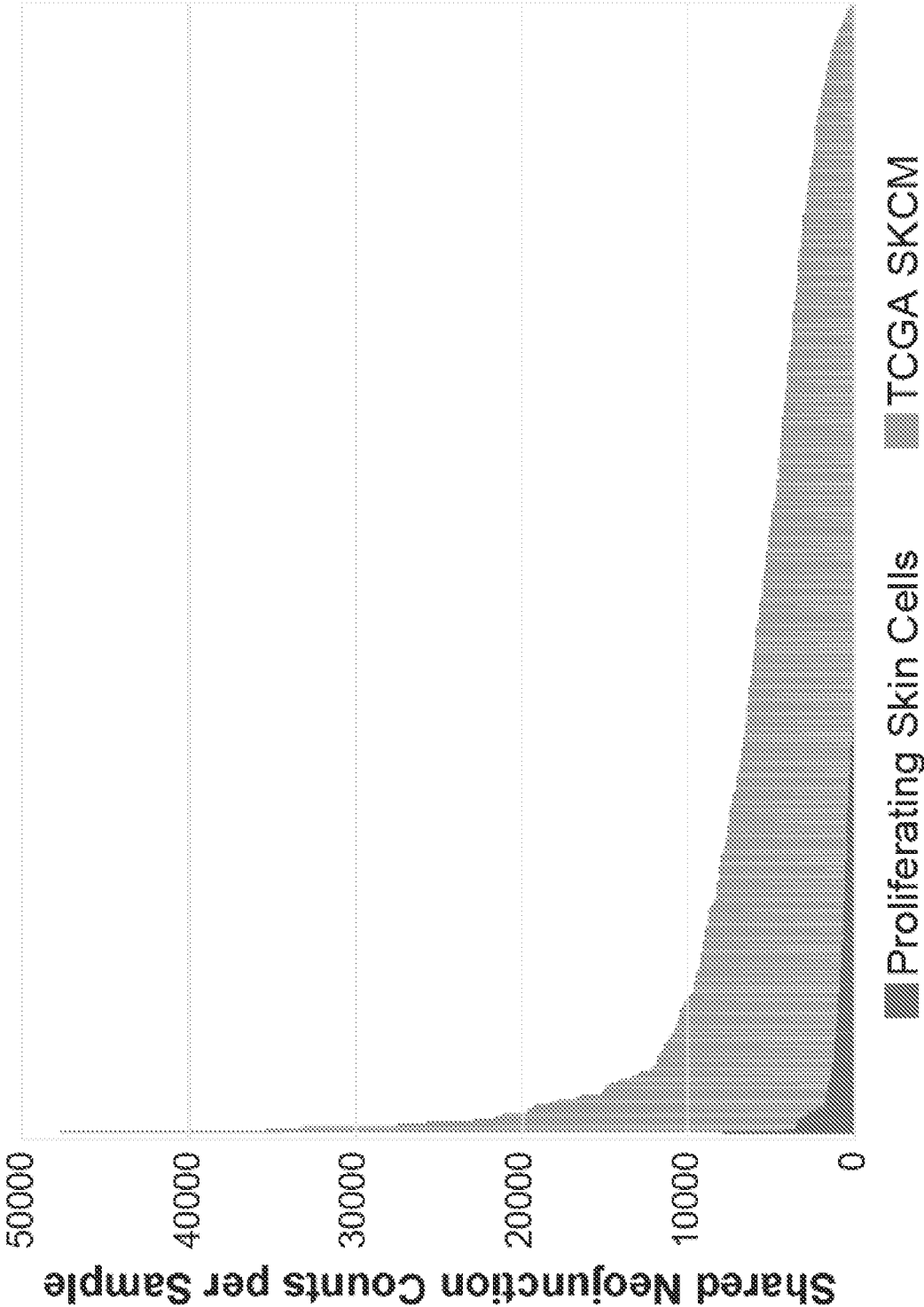


FIG. 17C

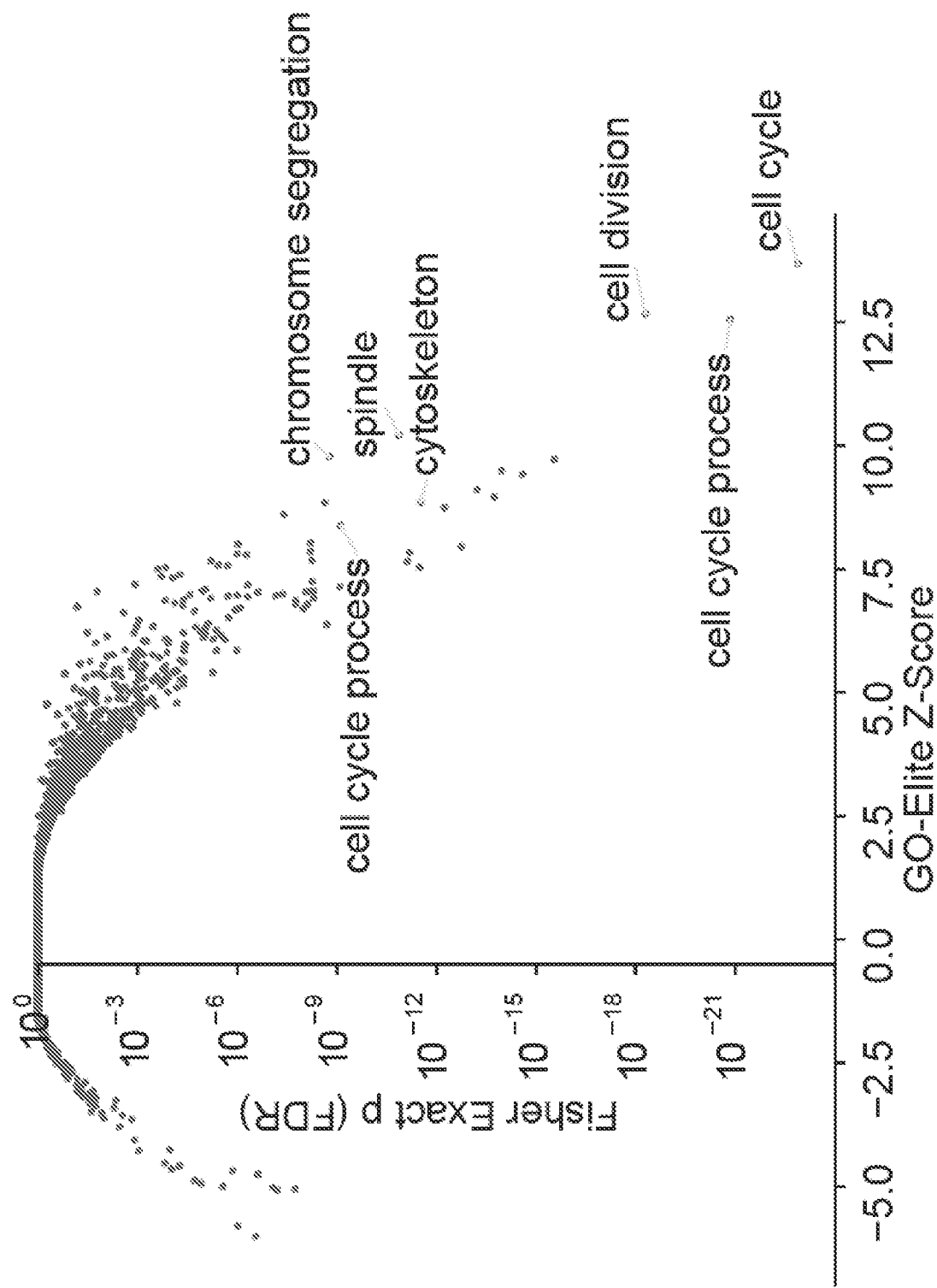
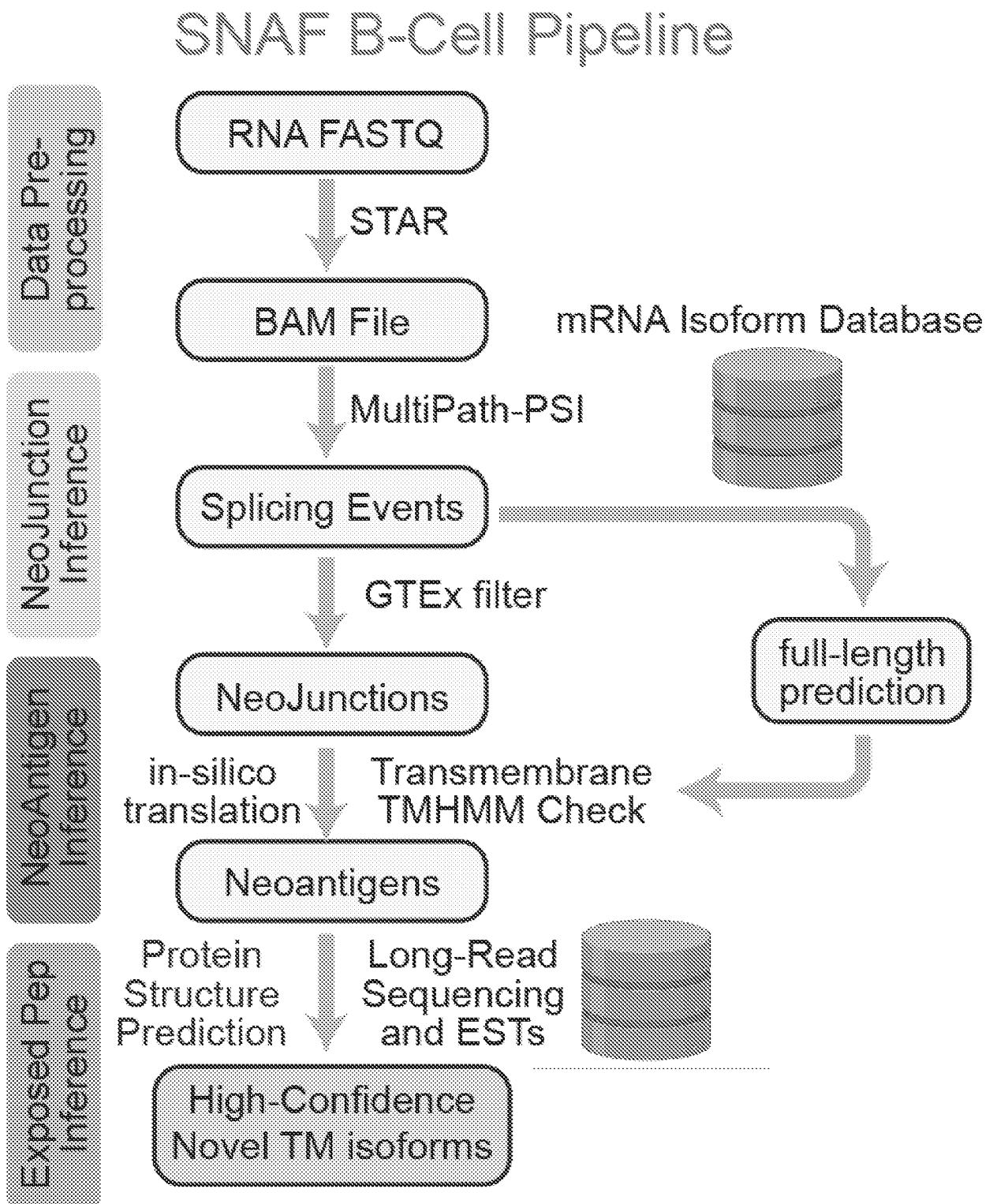


FIG. 17D

47/78

**FIG. 18A**

48/78

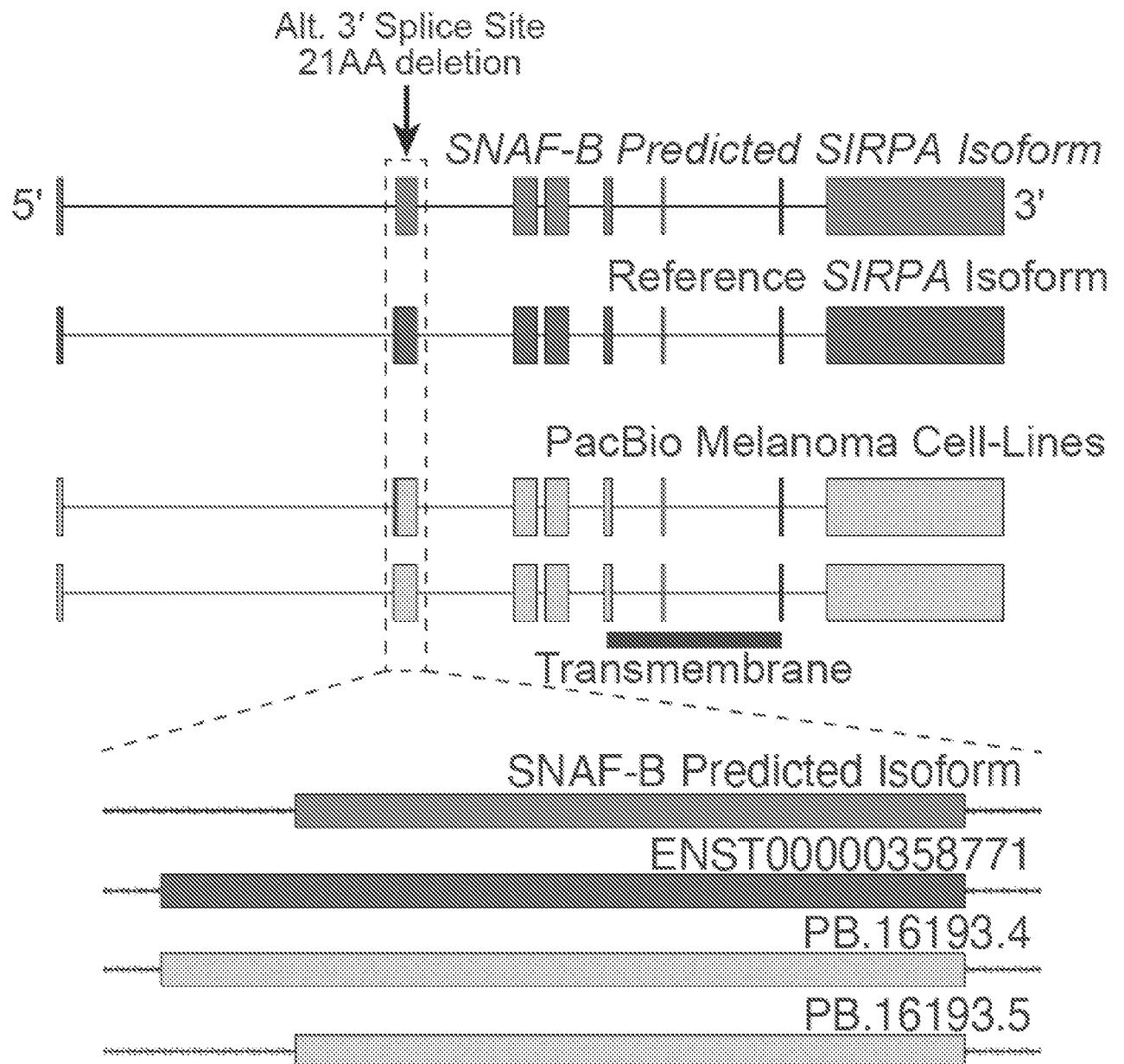


FIG. 18B

49/78

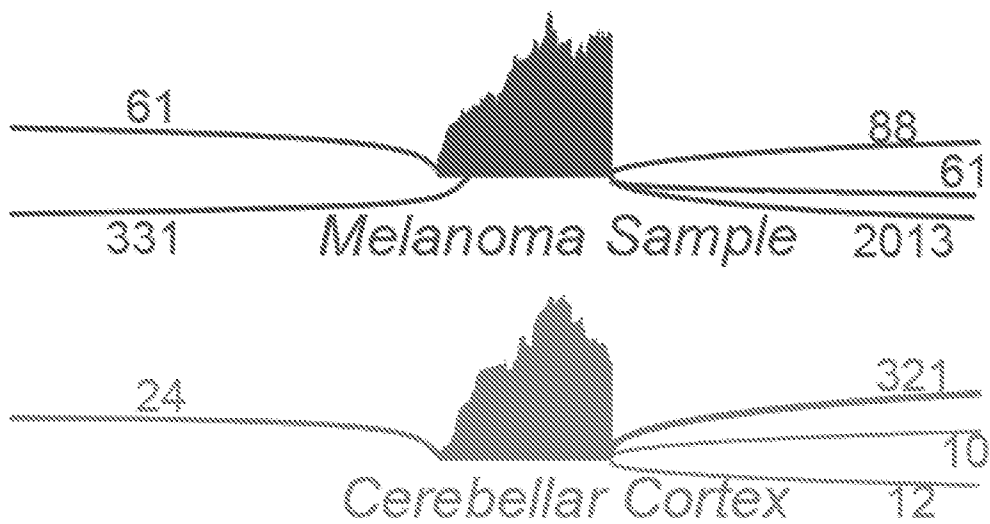


FIG. 18C

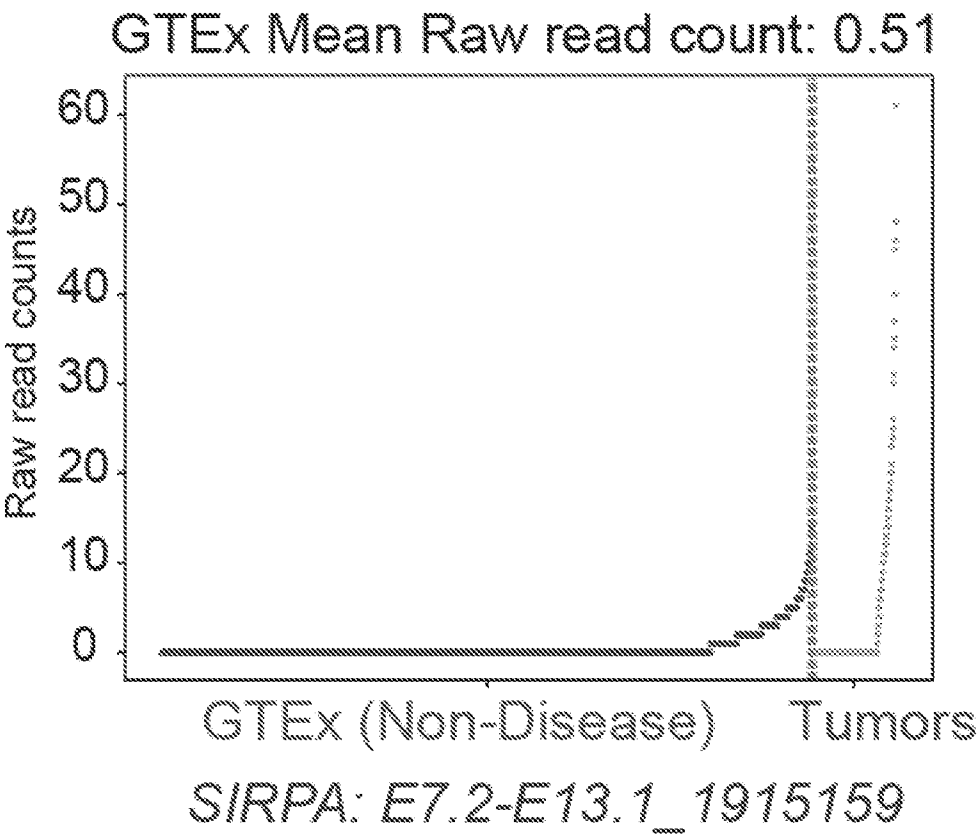


FIG. 18D

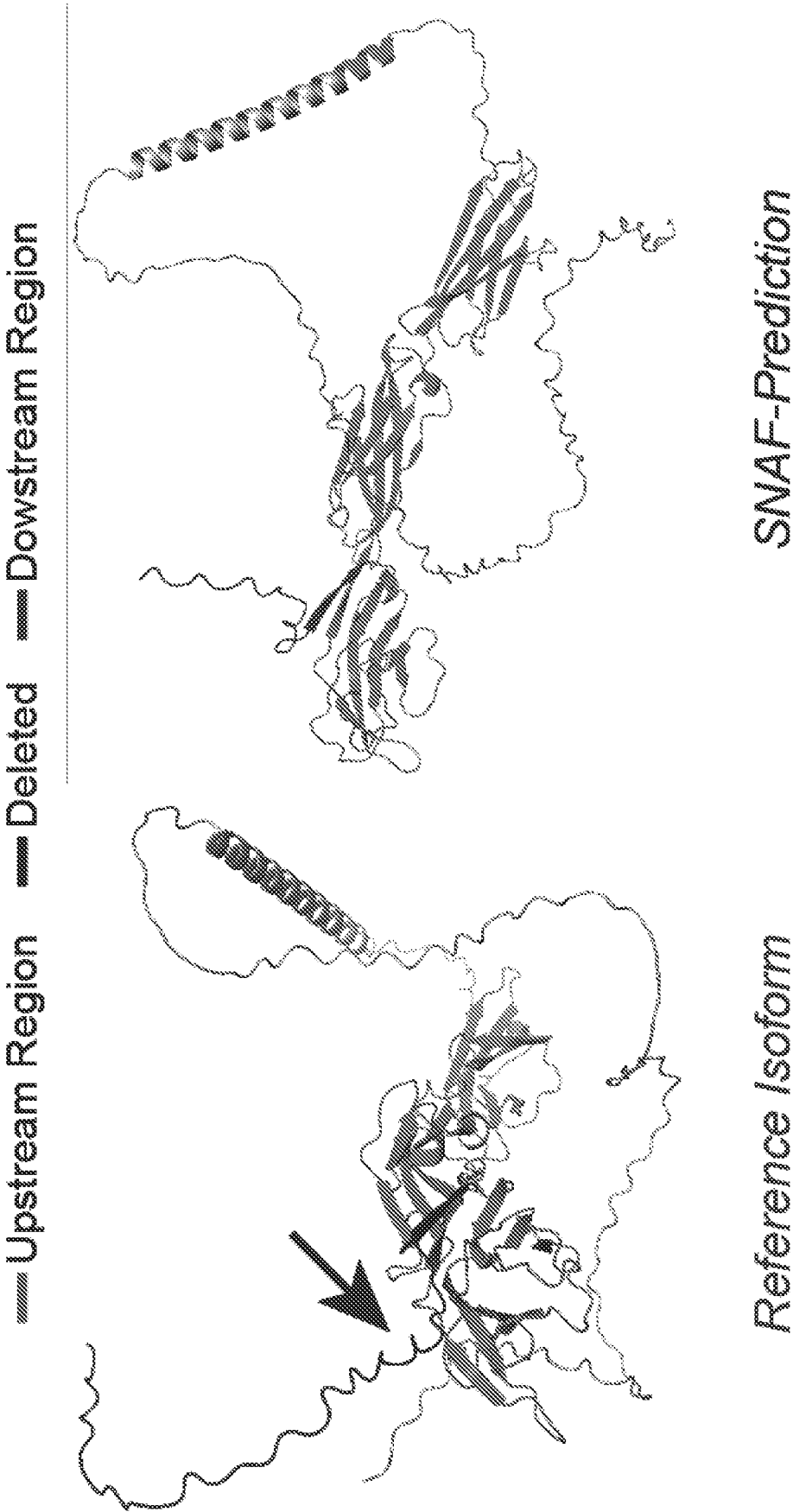


FIG. 18E

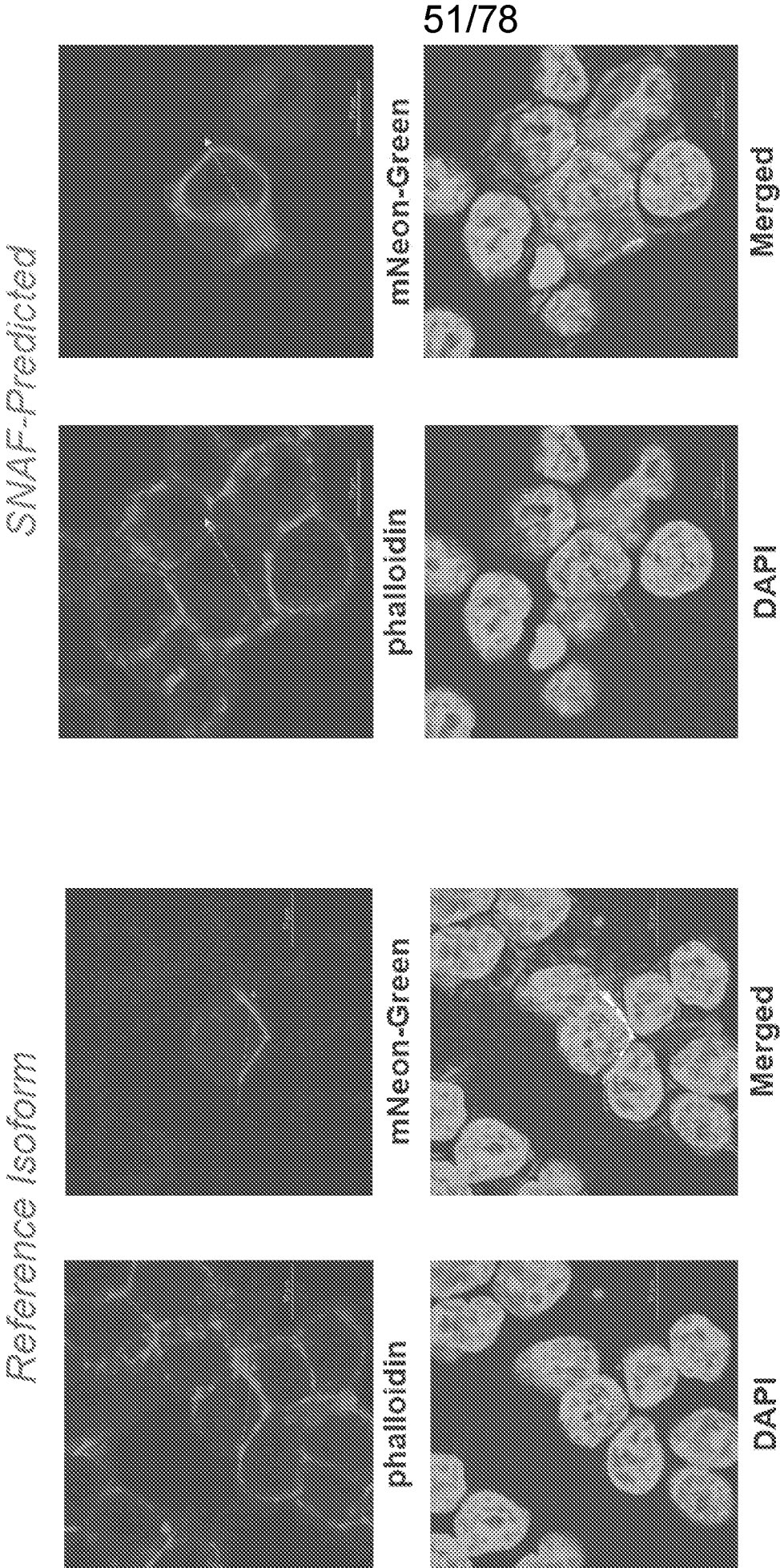


FIG. 18G

FIG. 18F

52/78

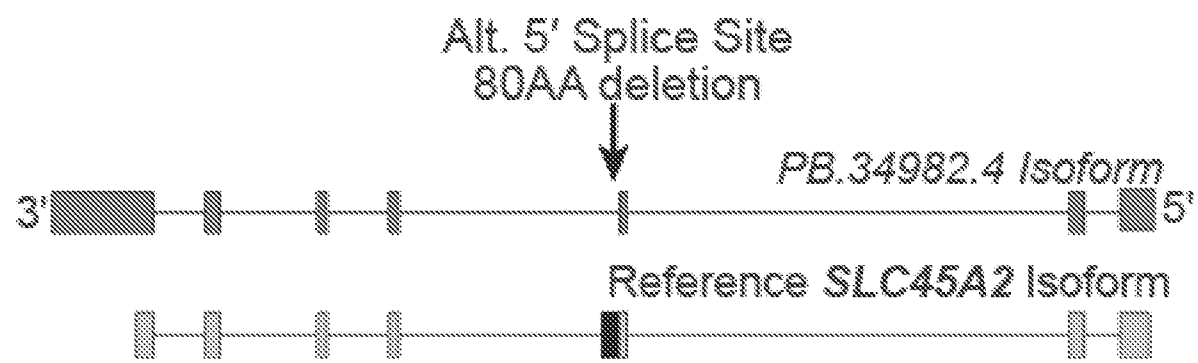


FIG. 19A

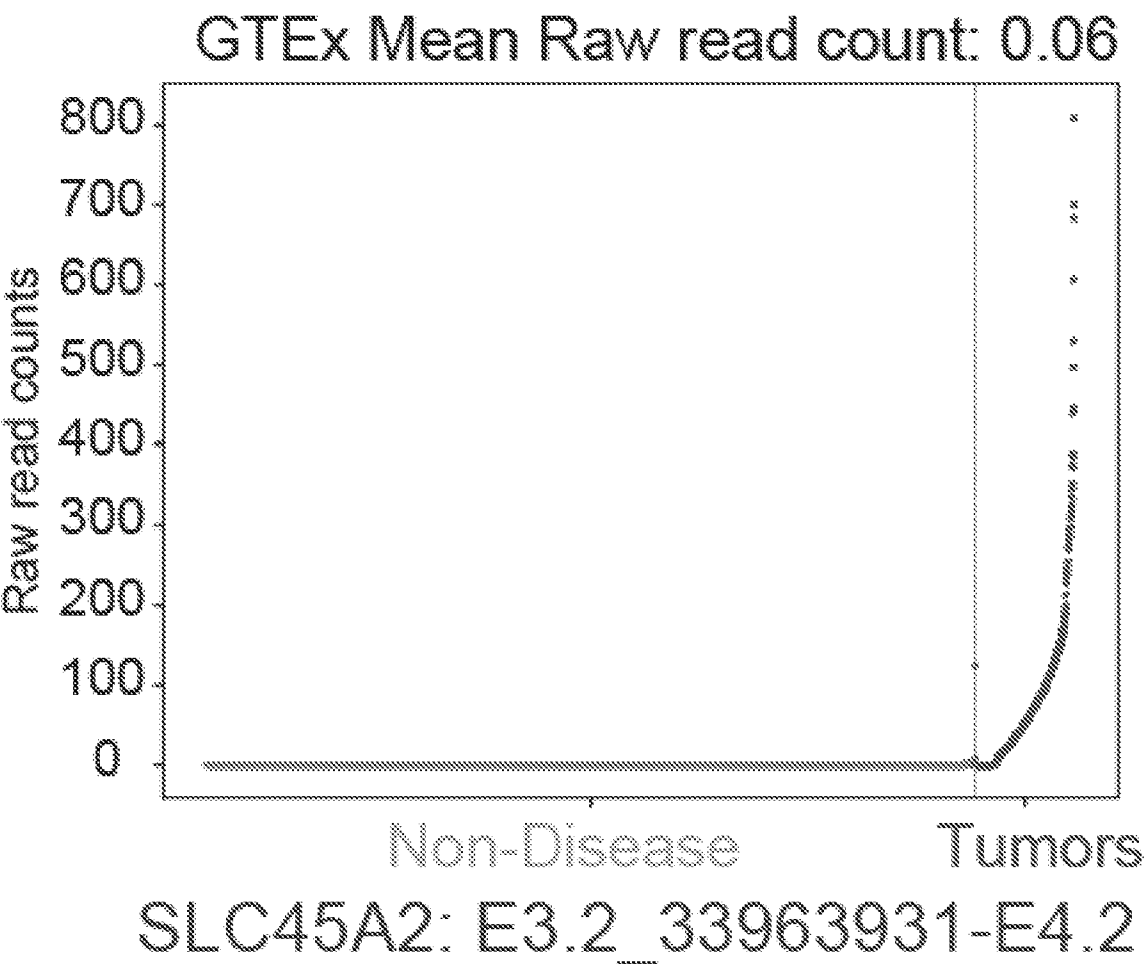
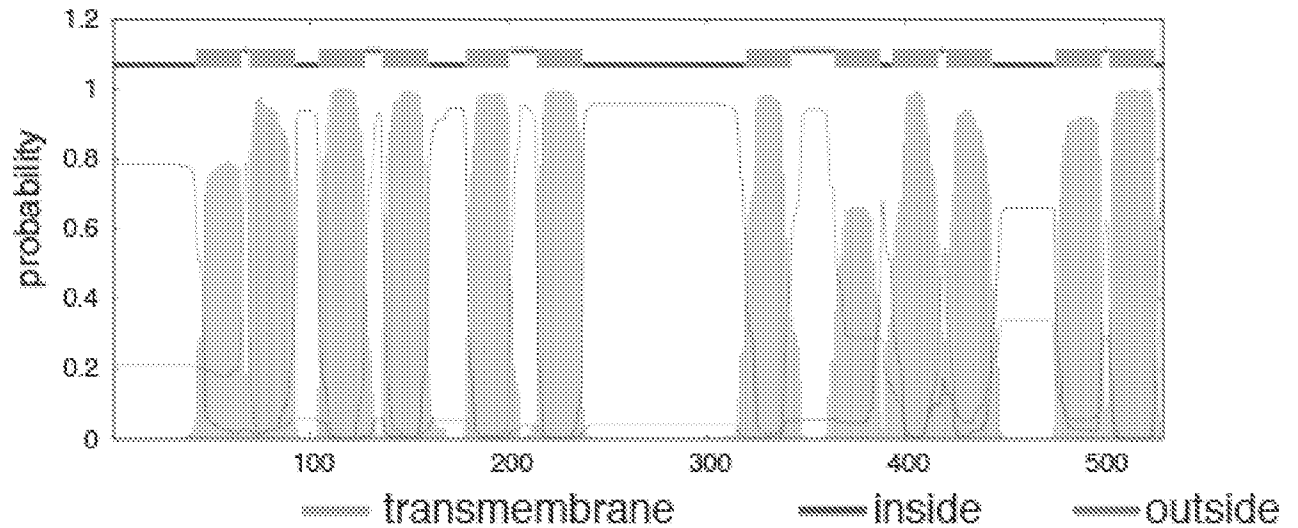
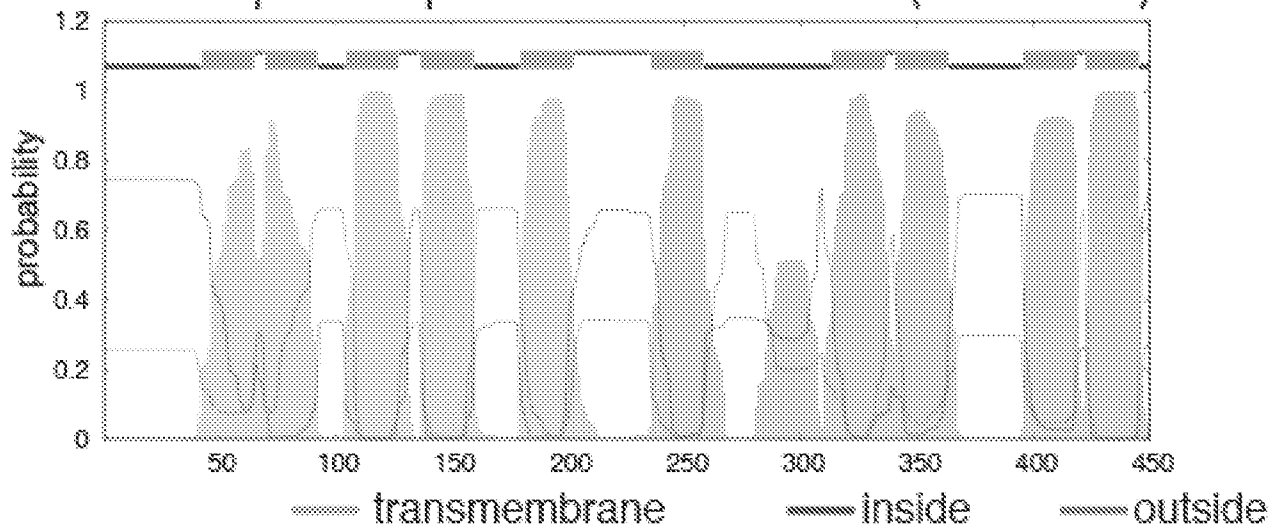


FIG. 19B

53/78

TMHMM posterior probabilities - **SLC45A2** (ENST00000296589)TMHMM posterior probabilities for **SLC45A2** (PB.34982.4)**FIG. 19C**

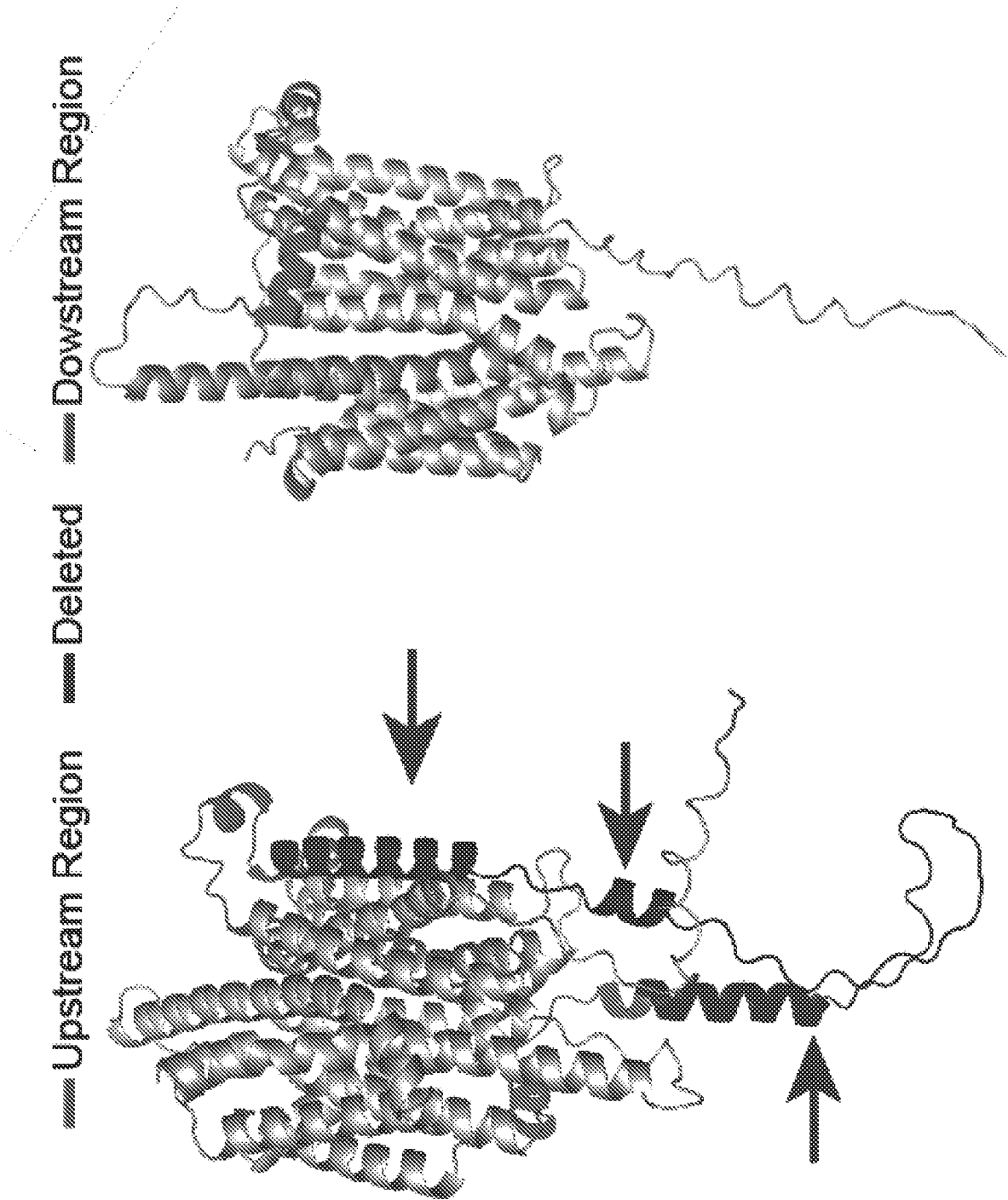
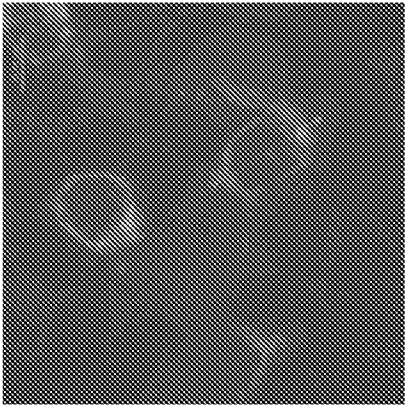


FIG. 19D

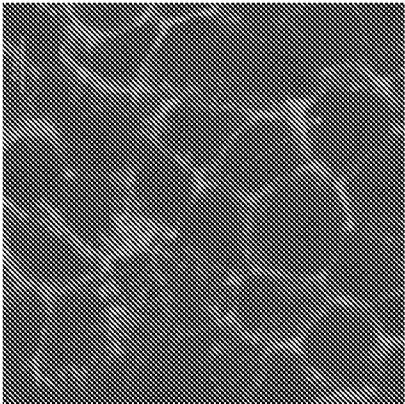
SNAF-Predicted

Reference Isoform

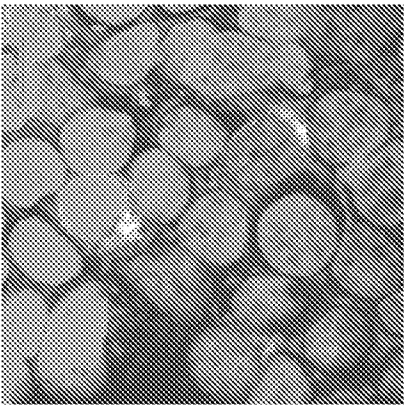
SNAF-Predicted



mNeon-Green

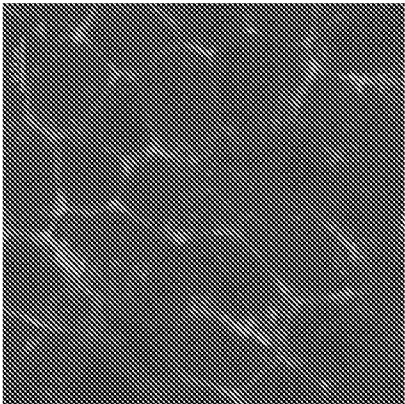


phalloidin

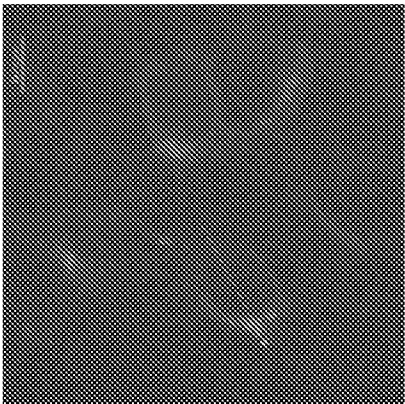


Merged

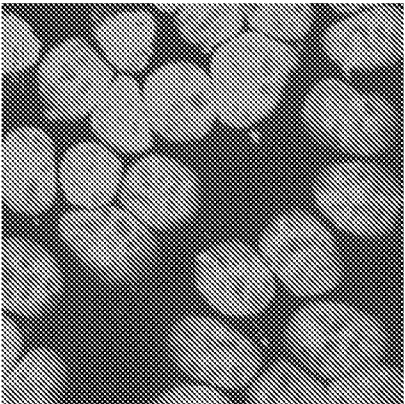
Reference Isoform



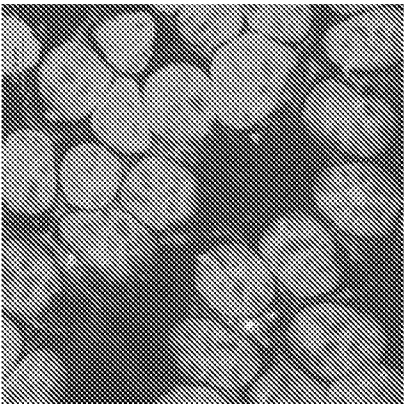
phalloidin



mNeon-Green



DAPI

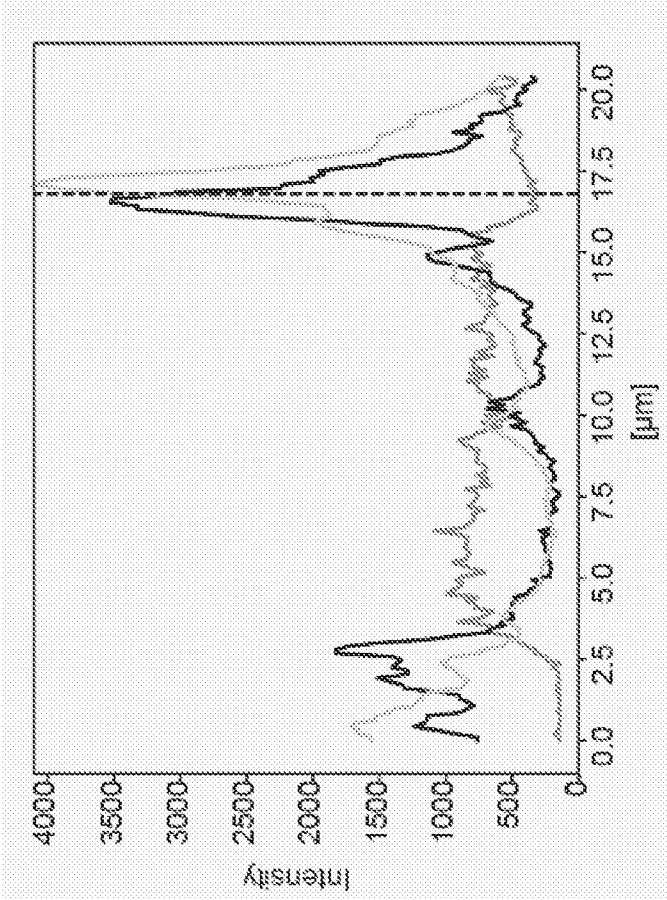


Merged

DAPI

FIG. 19E

SNAP-PREDICTED



REFERENCE ISOFORM

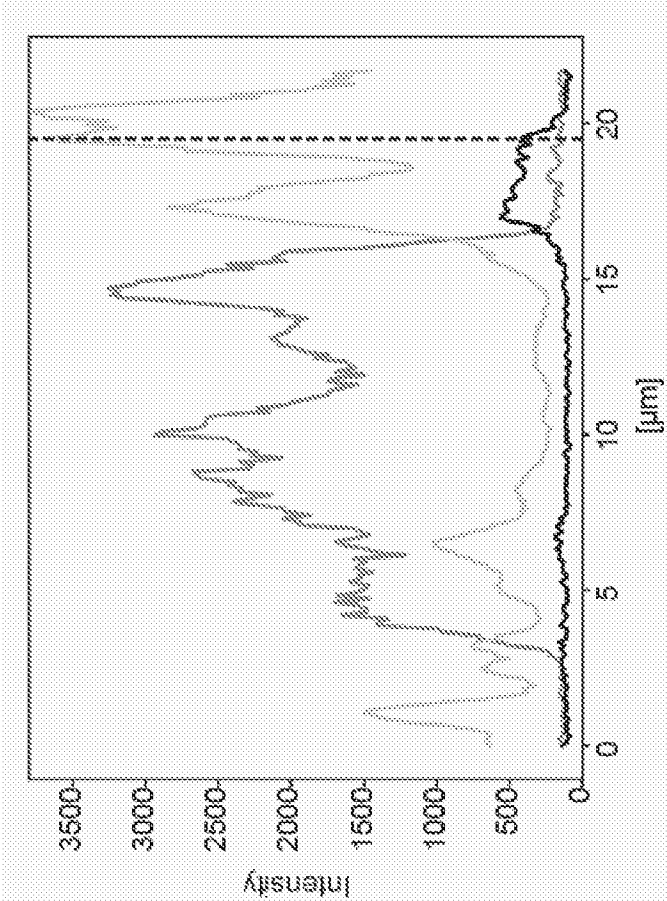


FIG. 19E (CONT.)

57/78

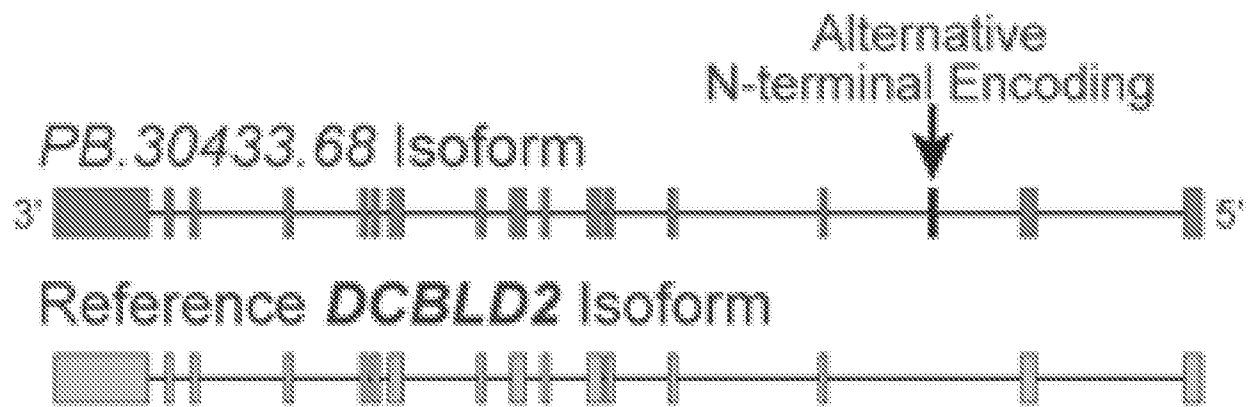


FIG. 20A

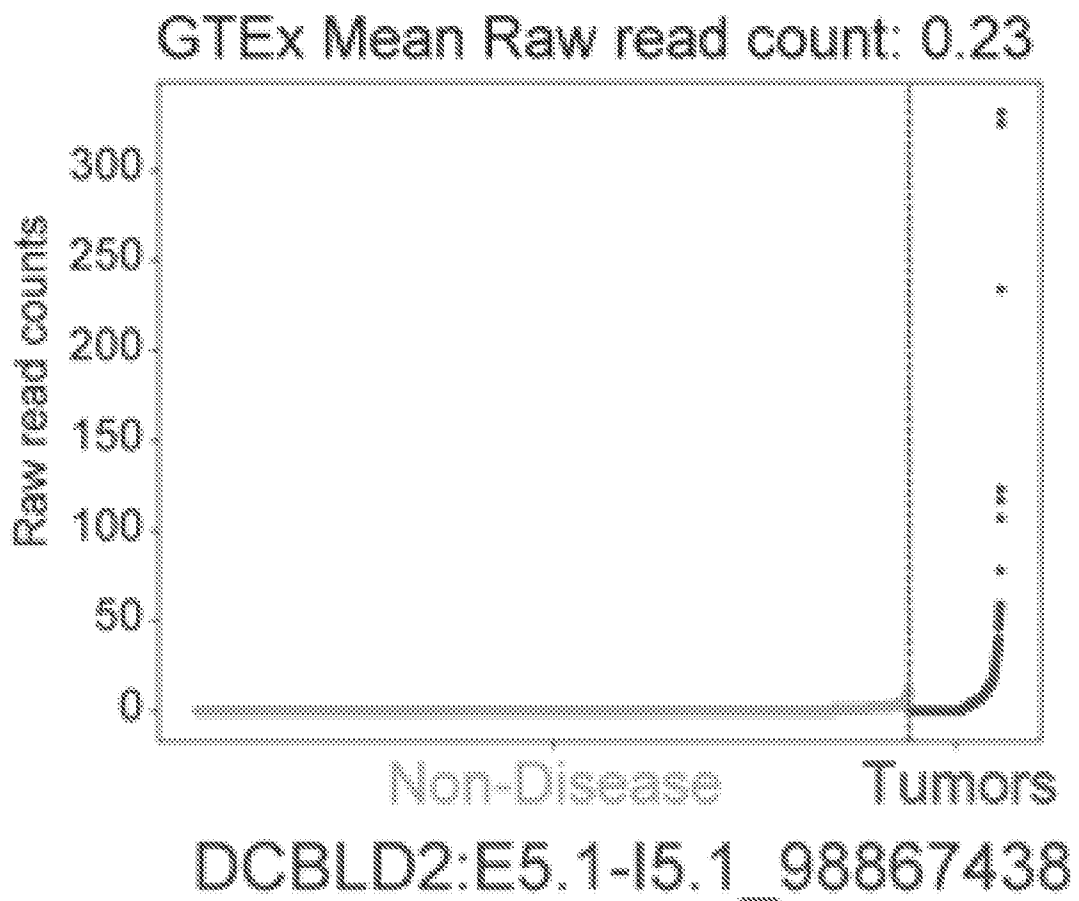


FIG. 20B

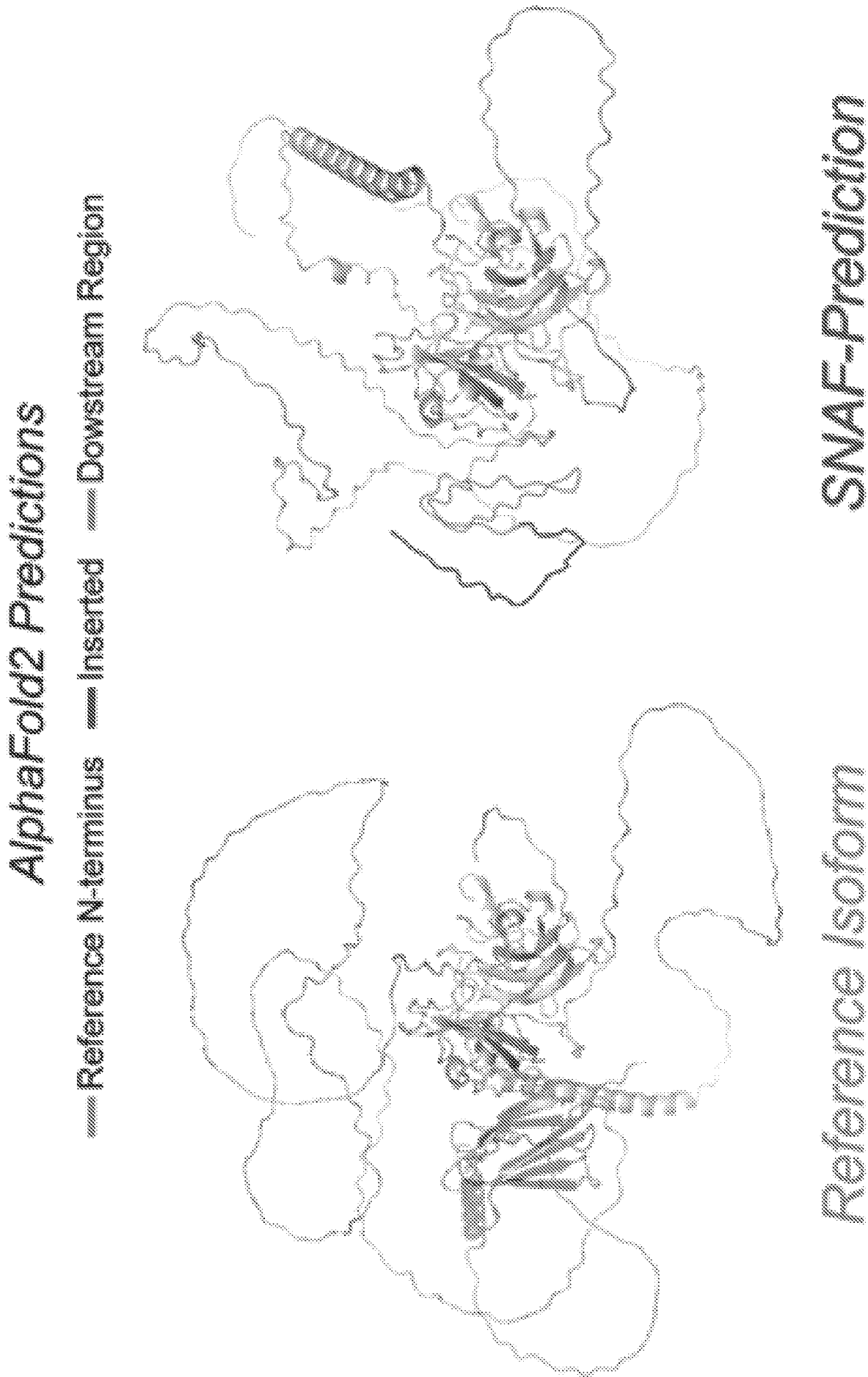


FIG. 20C

59/78

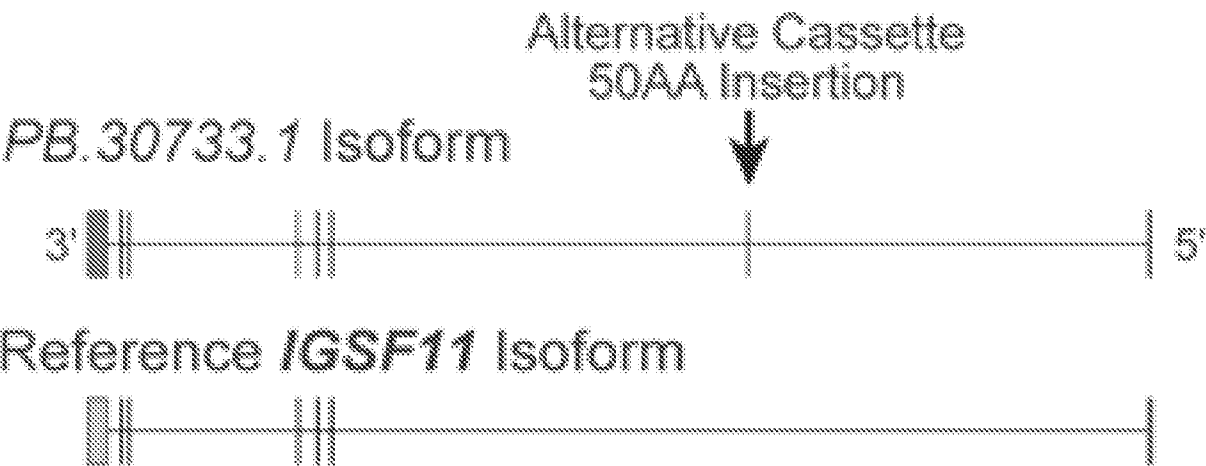
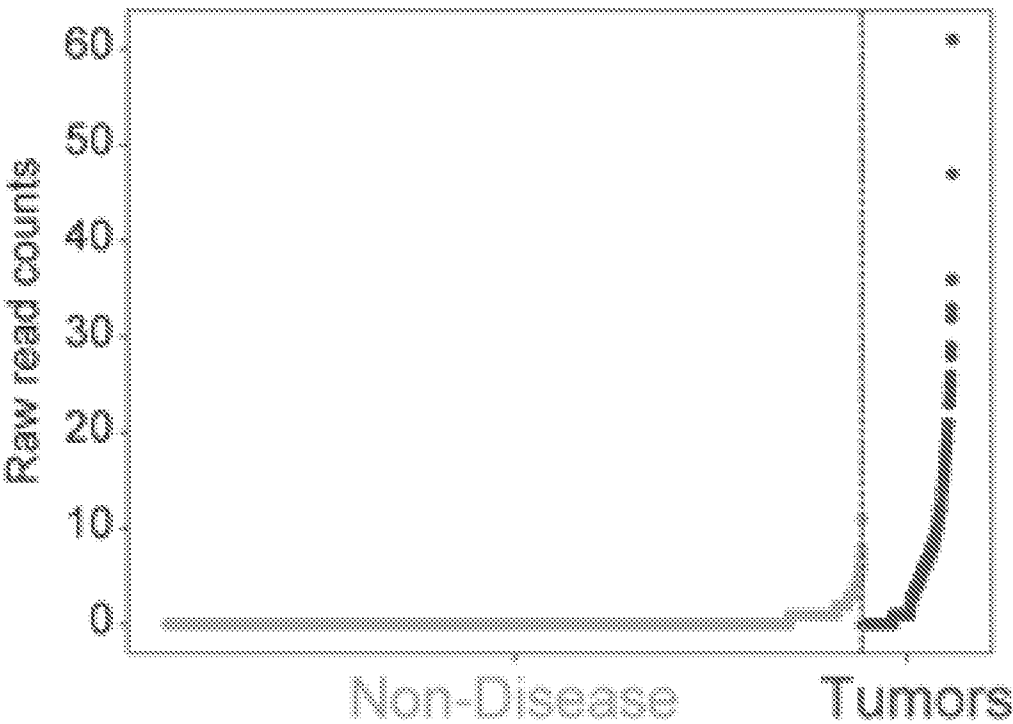


FIG. 20D

GTEX Mean Raw read count: 0.18



IGSF11:E4.4-E8.1

FIG. 20E



FIG. 20F

61/78

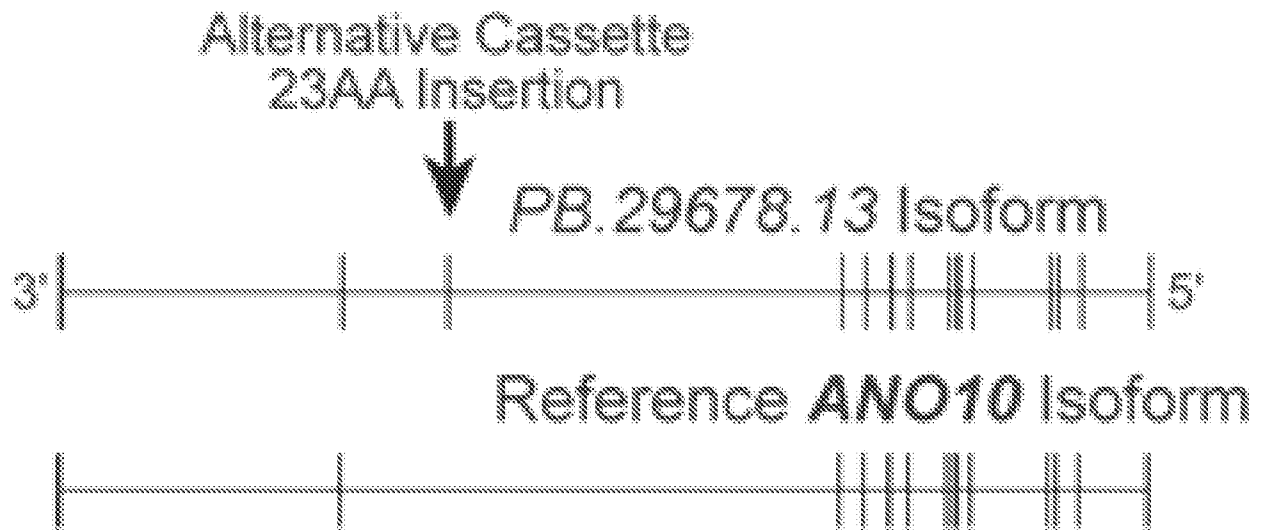
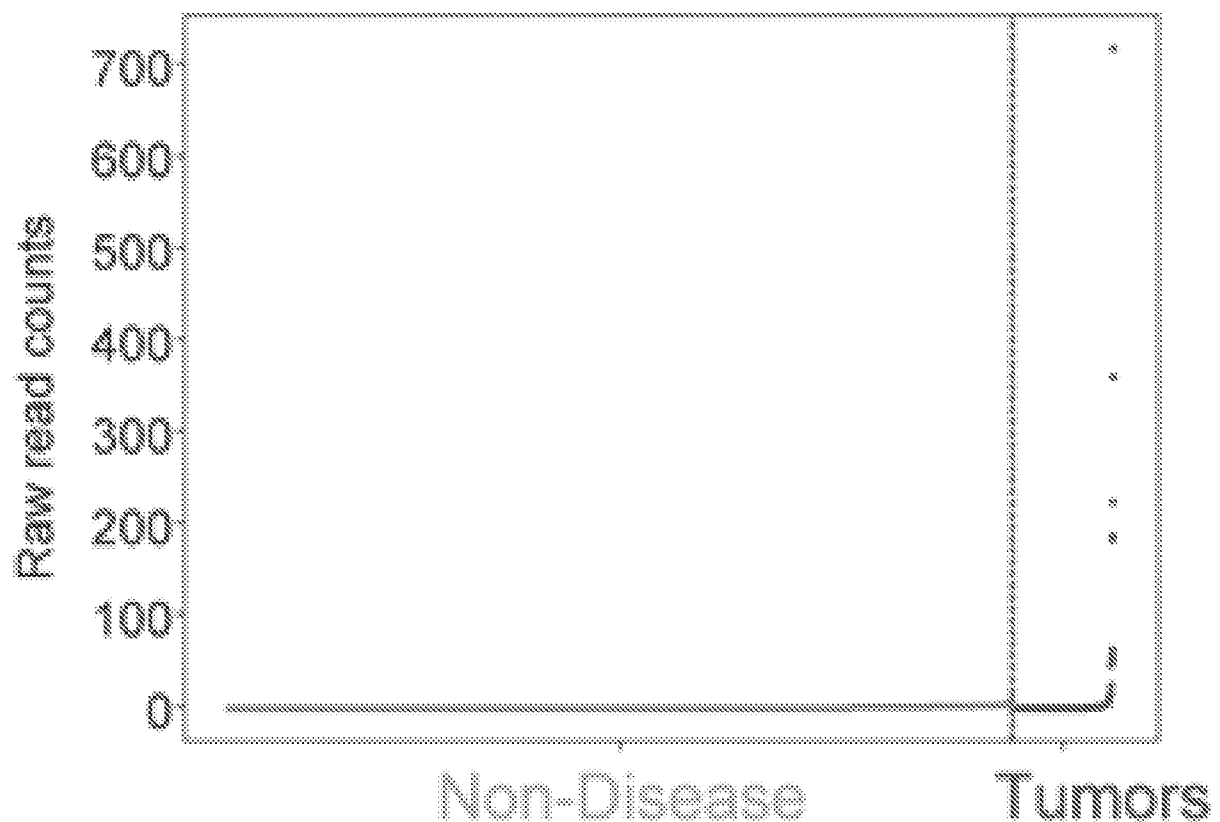


FIG. 20G

GTEx Mean Raw read count: 0.39

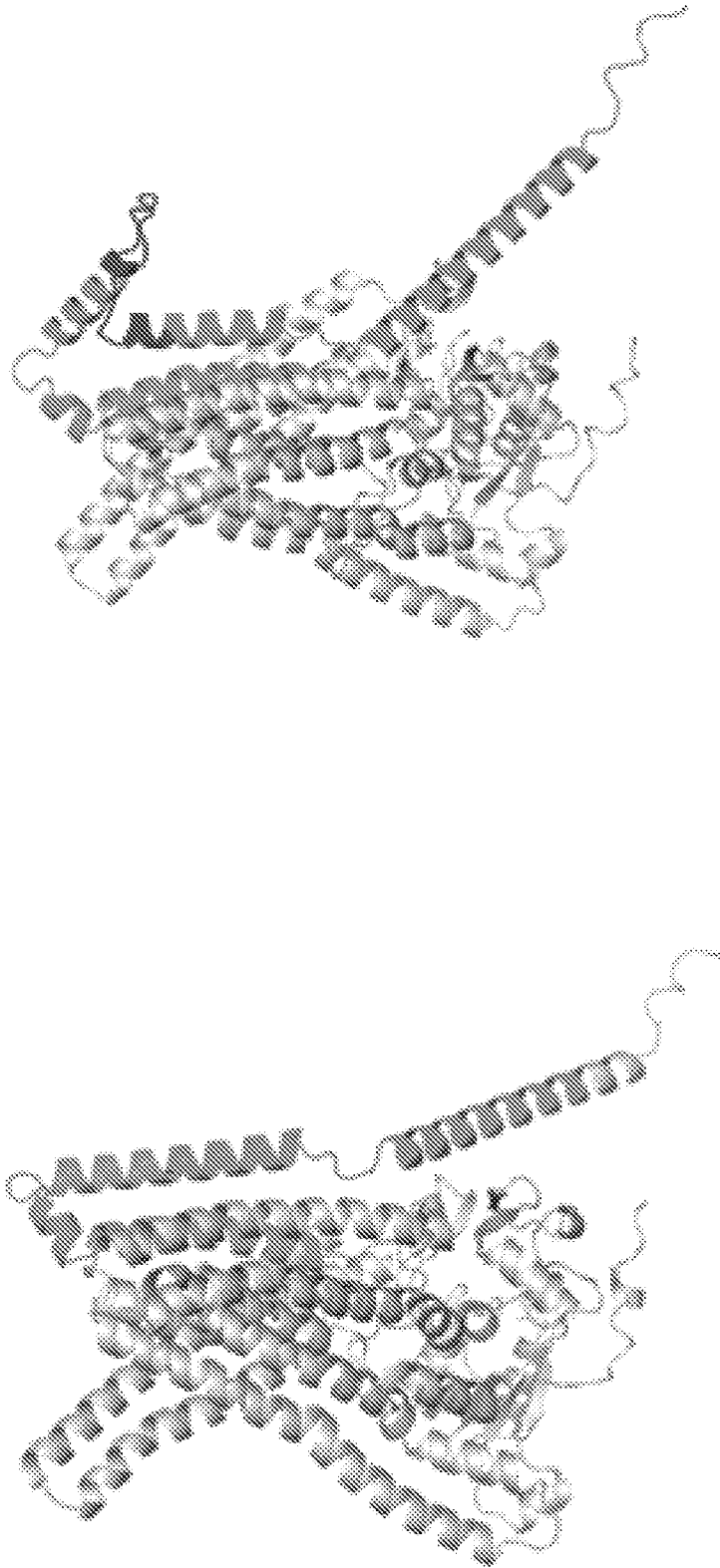


ANO10:I21.1 43457363-E22.1

FIG. 20H

AlphaFold2 Predictions

Upstream Region Insertion Downstream Region



Reference Isoform

SNAF-Prediction

FIG. 20I

63/78

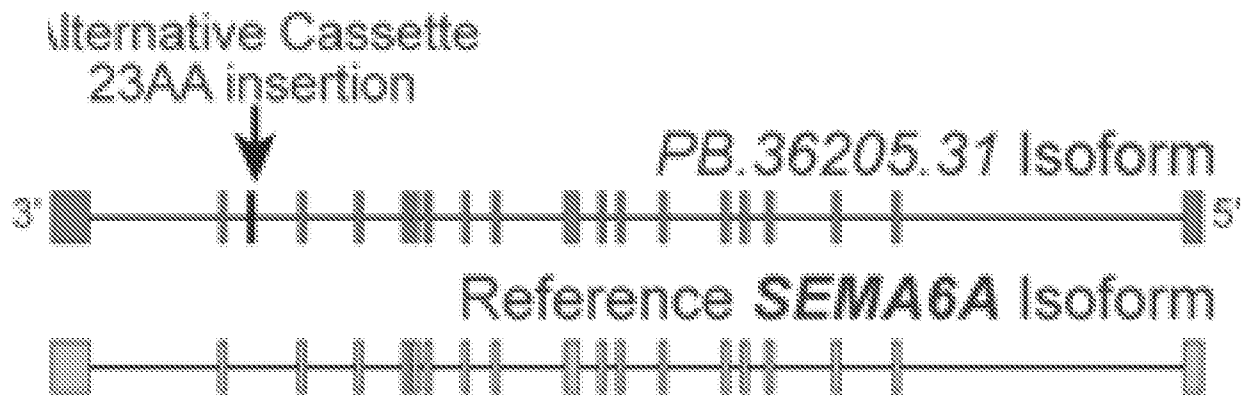


FIG. 20J

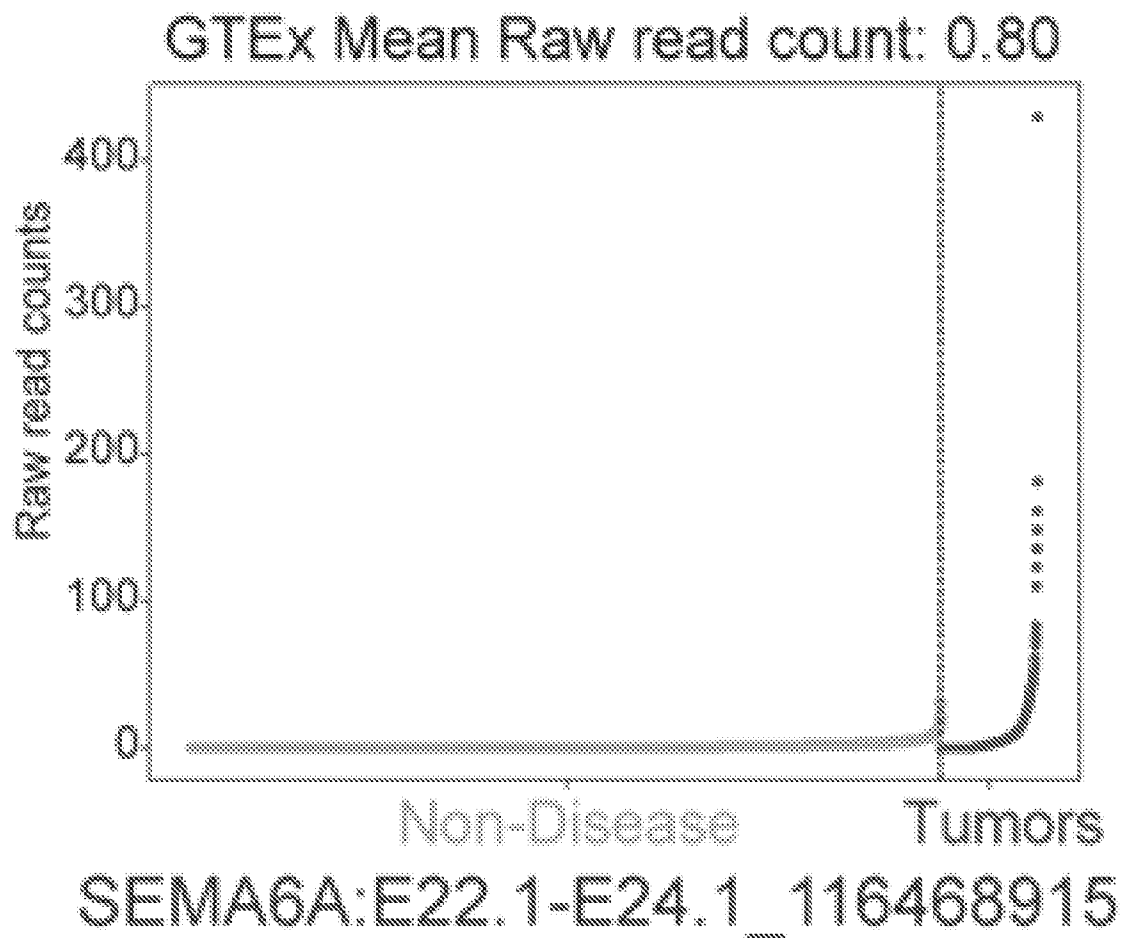


FIG. 20K

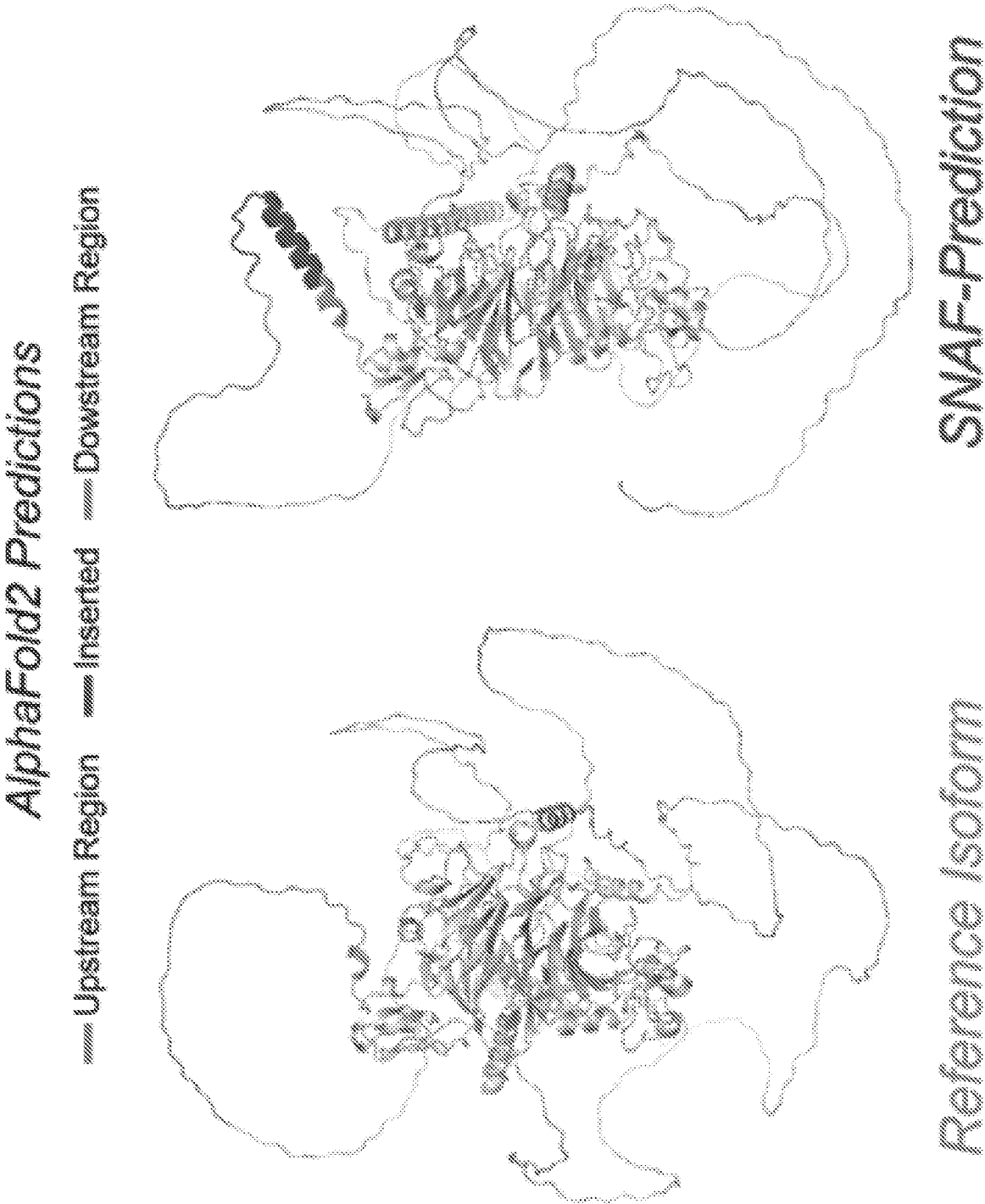
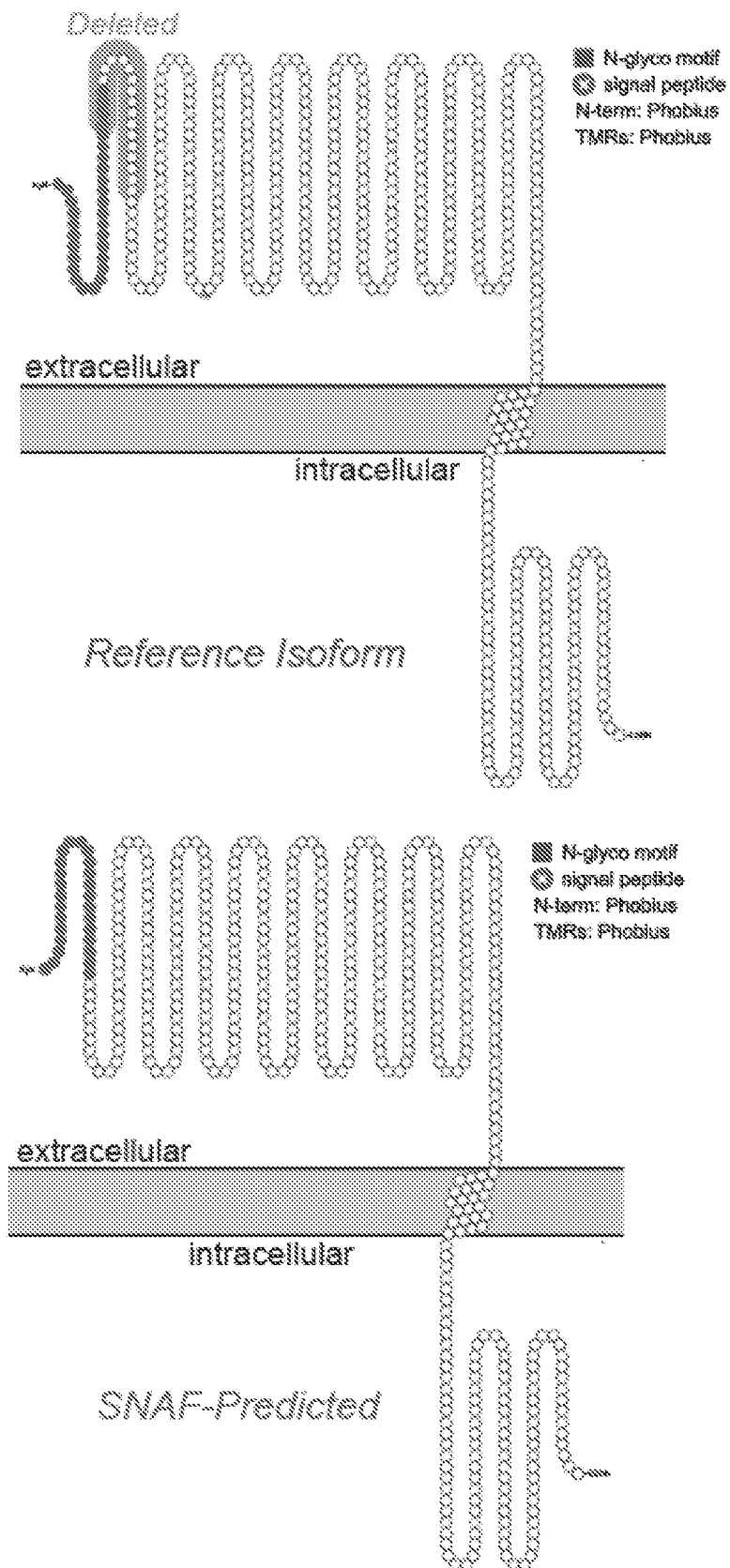


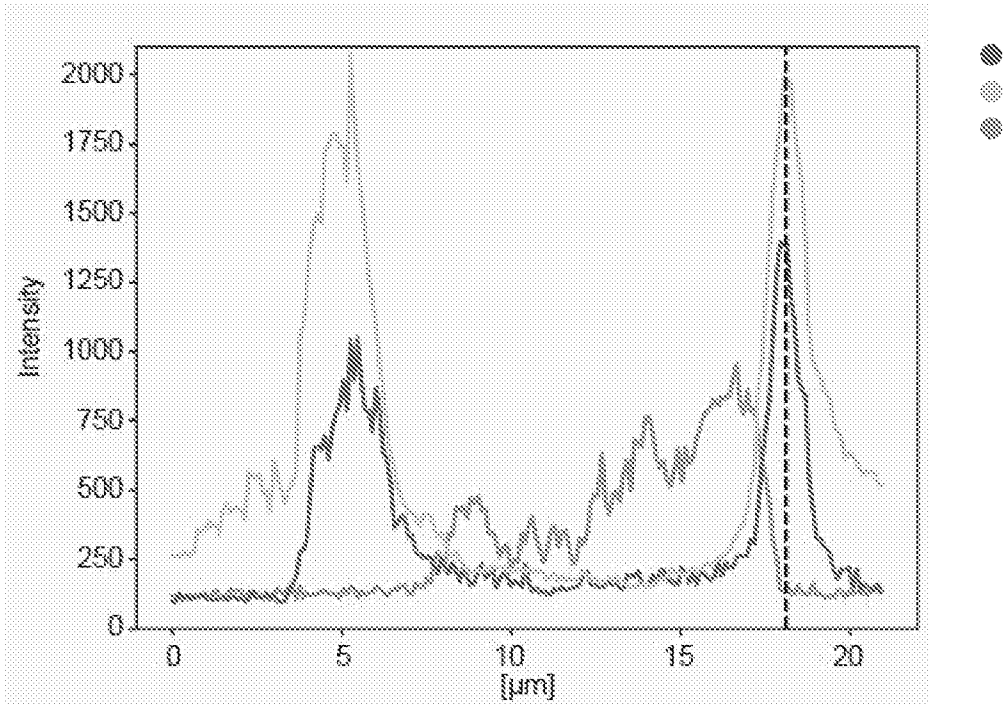
FIG. 20L

65/78

**FIG. 21A**

66/78

REFERENCE ISOFORM



SNAP-PREDICTED

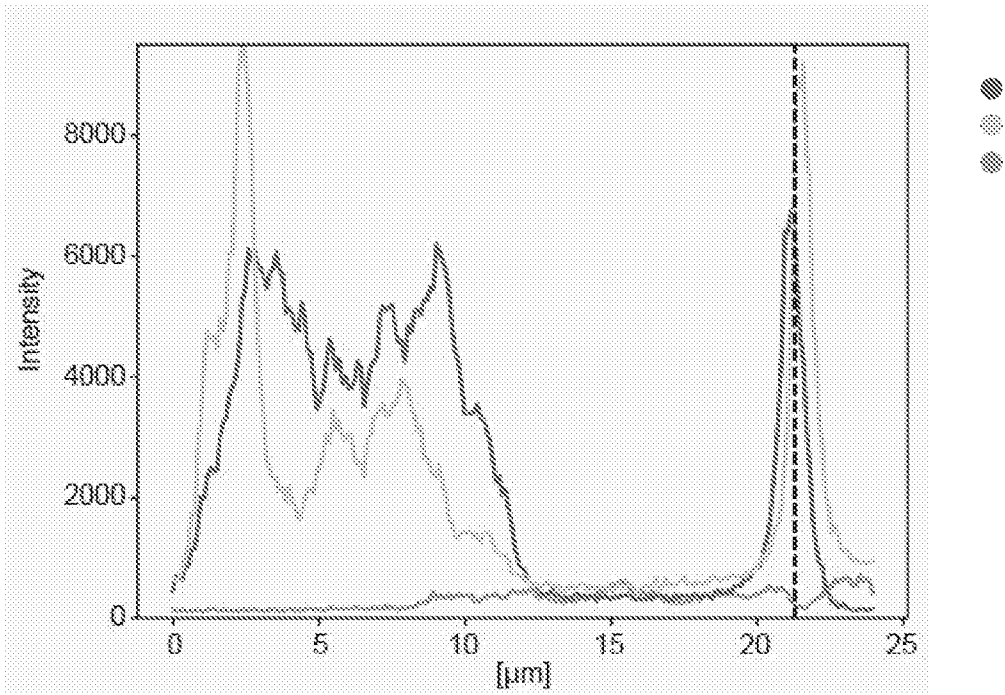
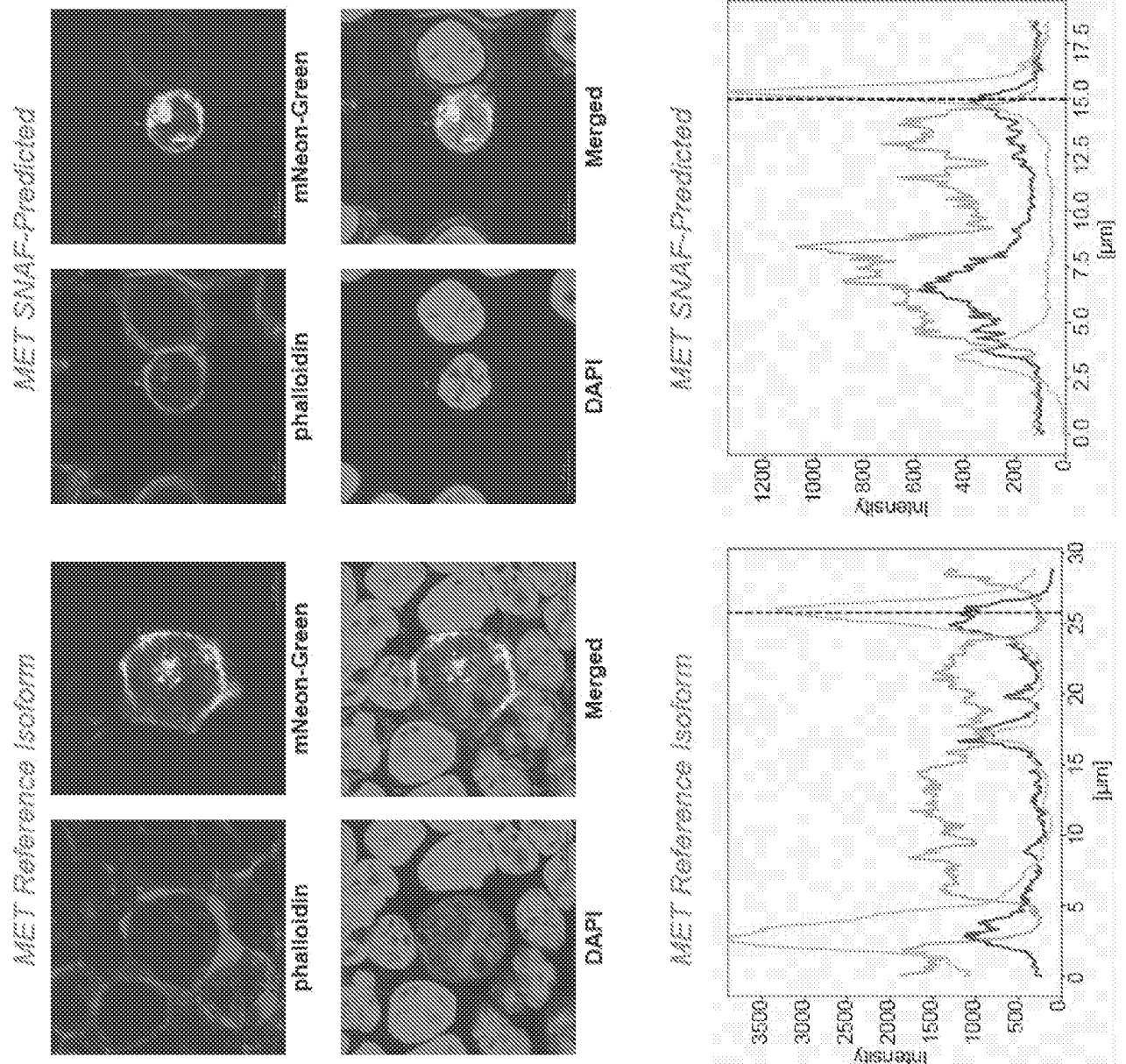


FIG. 21B

FIG. 21C



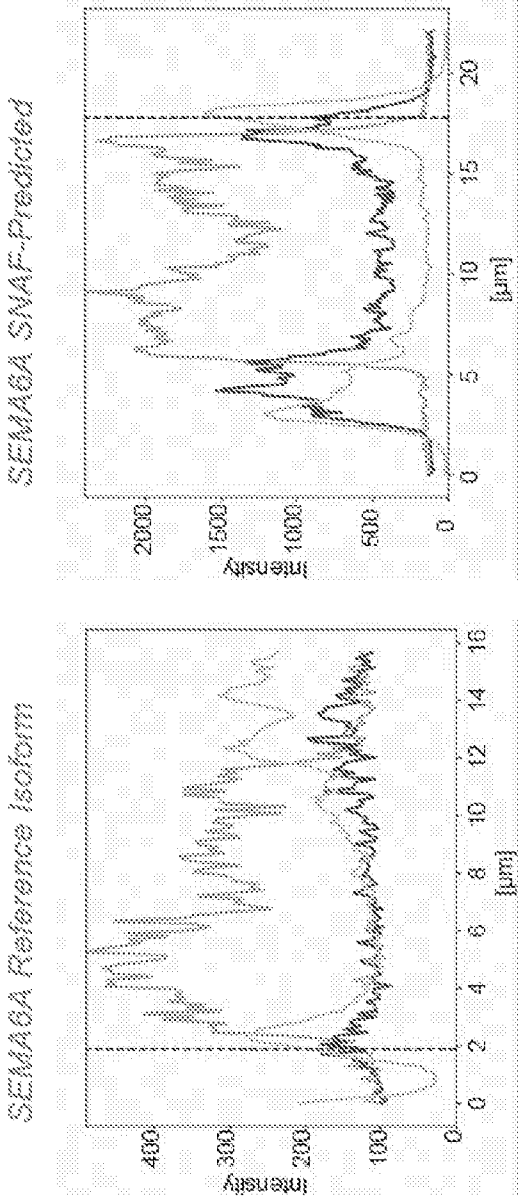
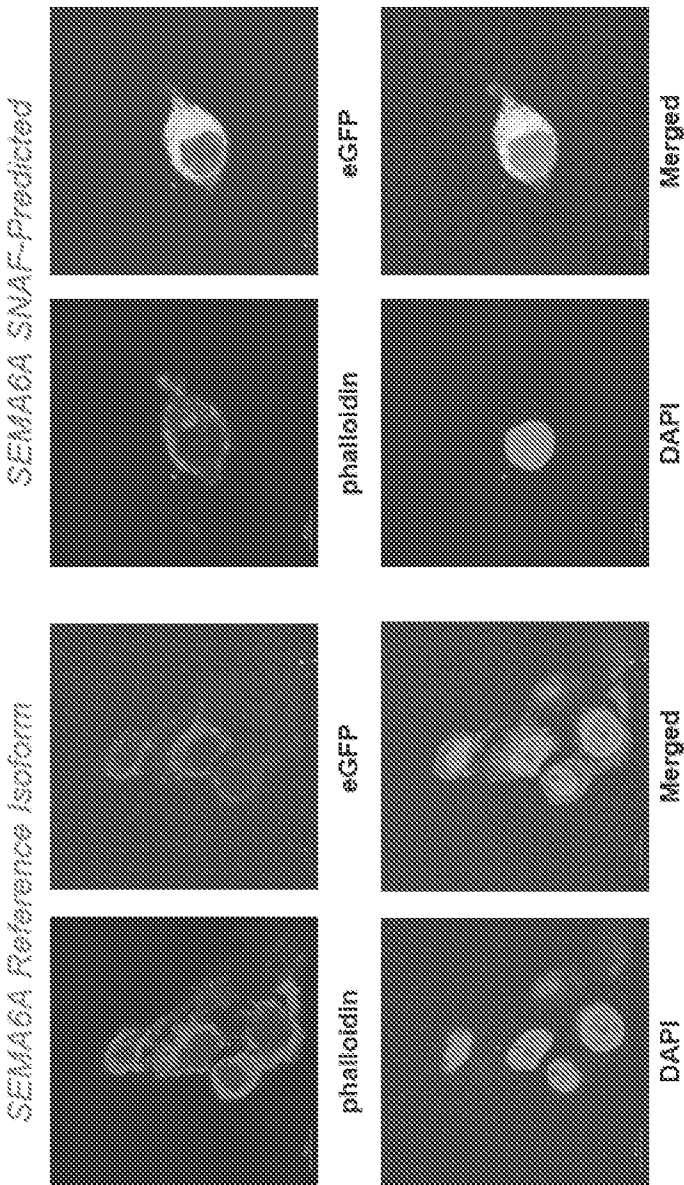


FIG. 21D

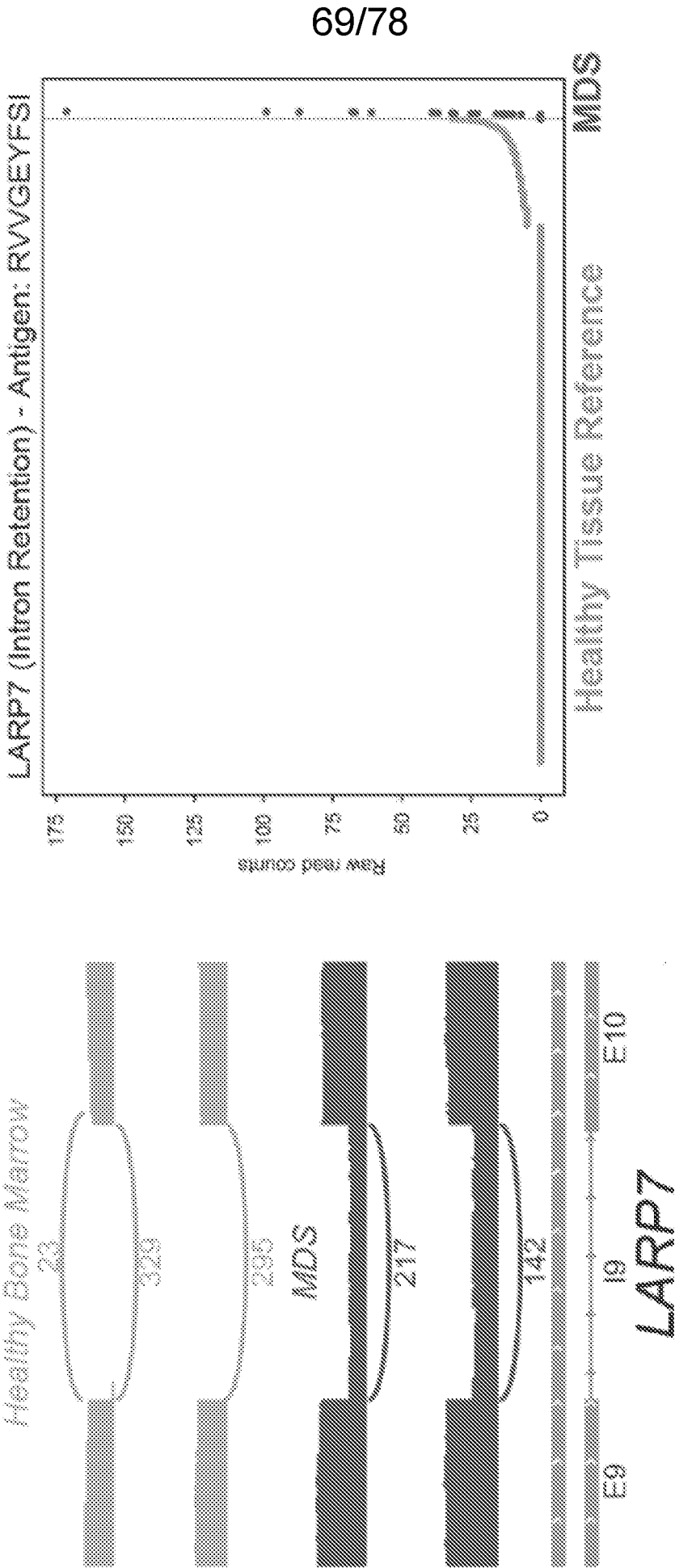


FIG. 22A

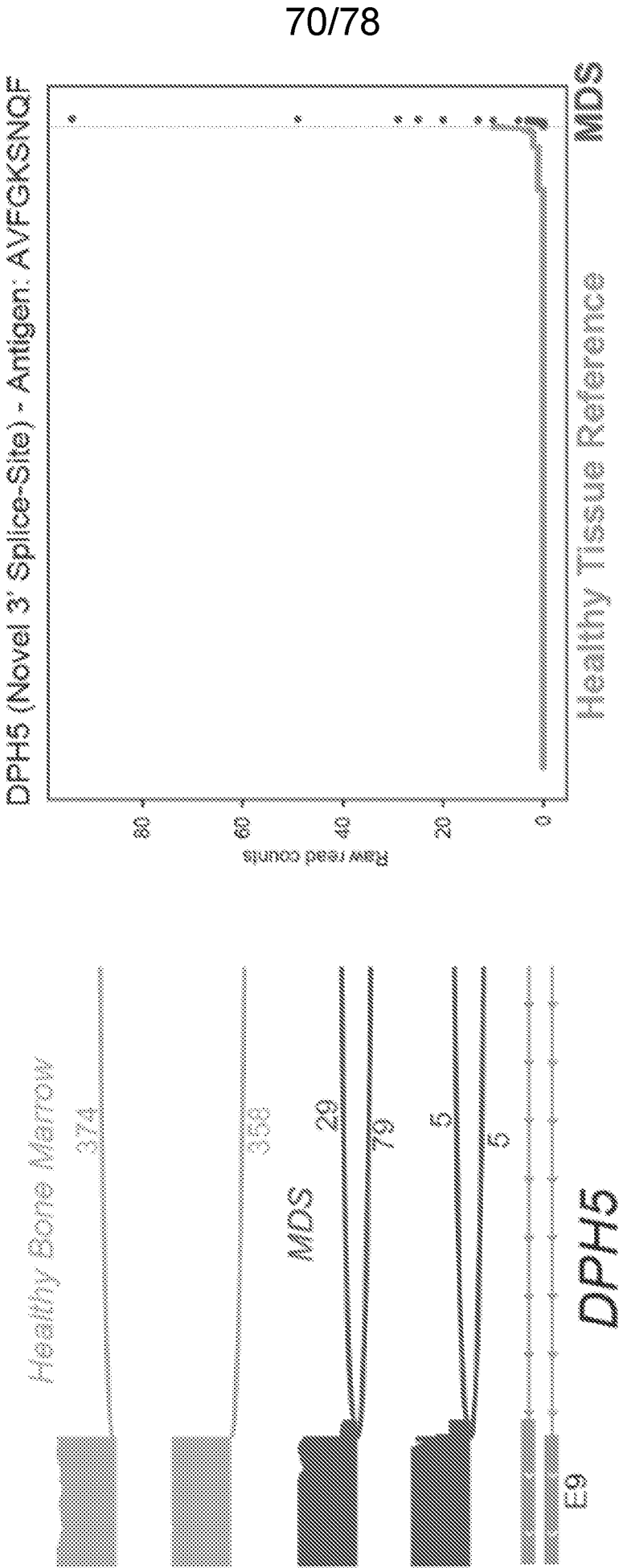


FIG. 22B

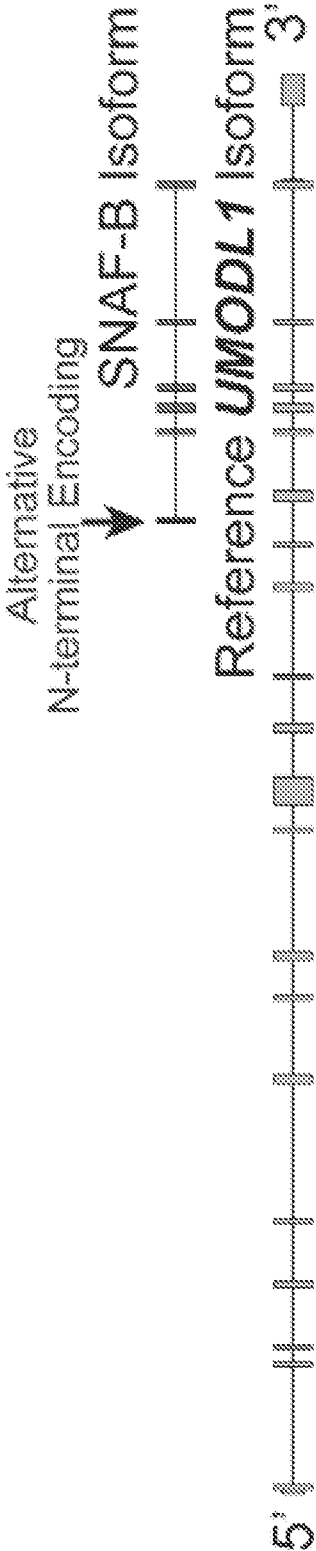


FIG. 23A

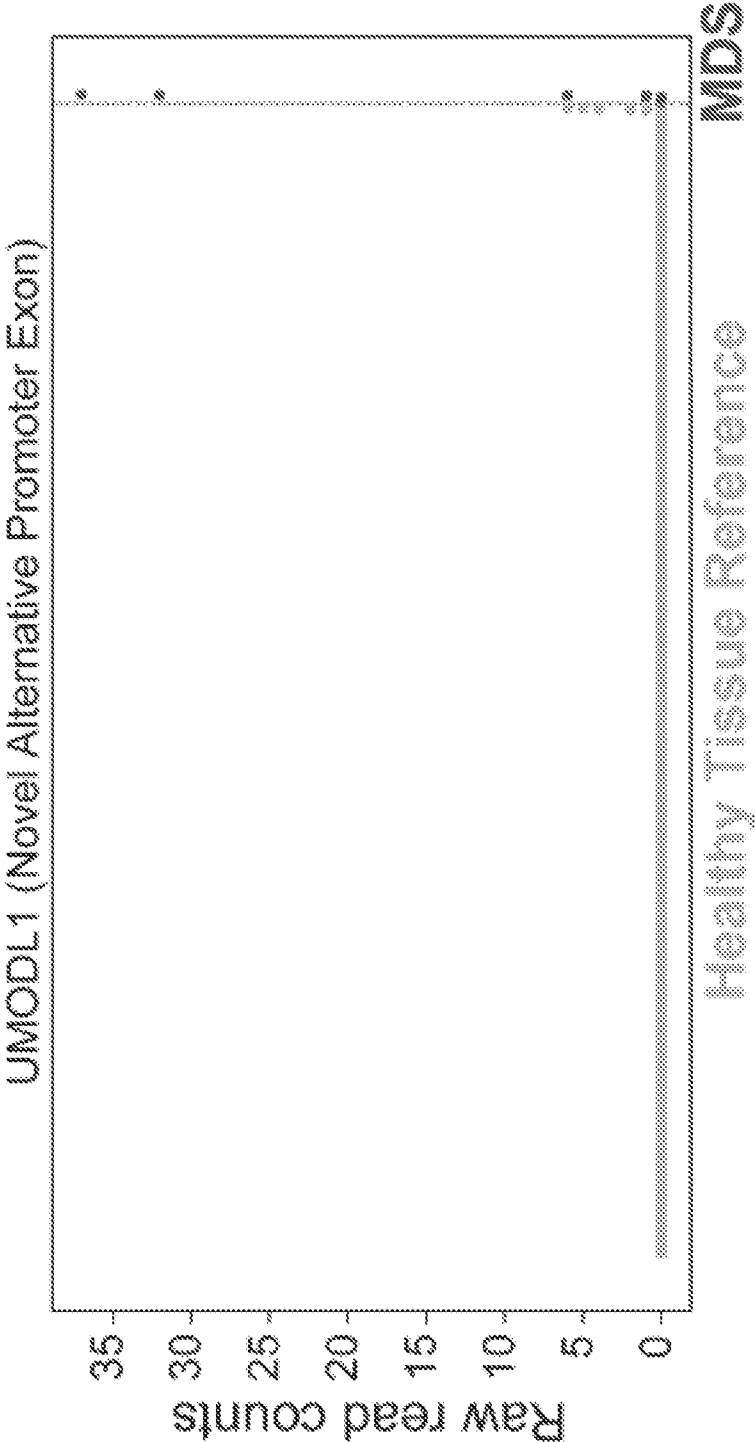
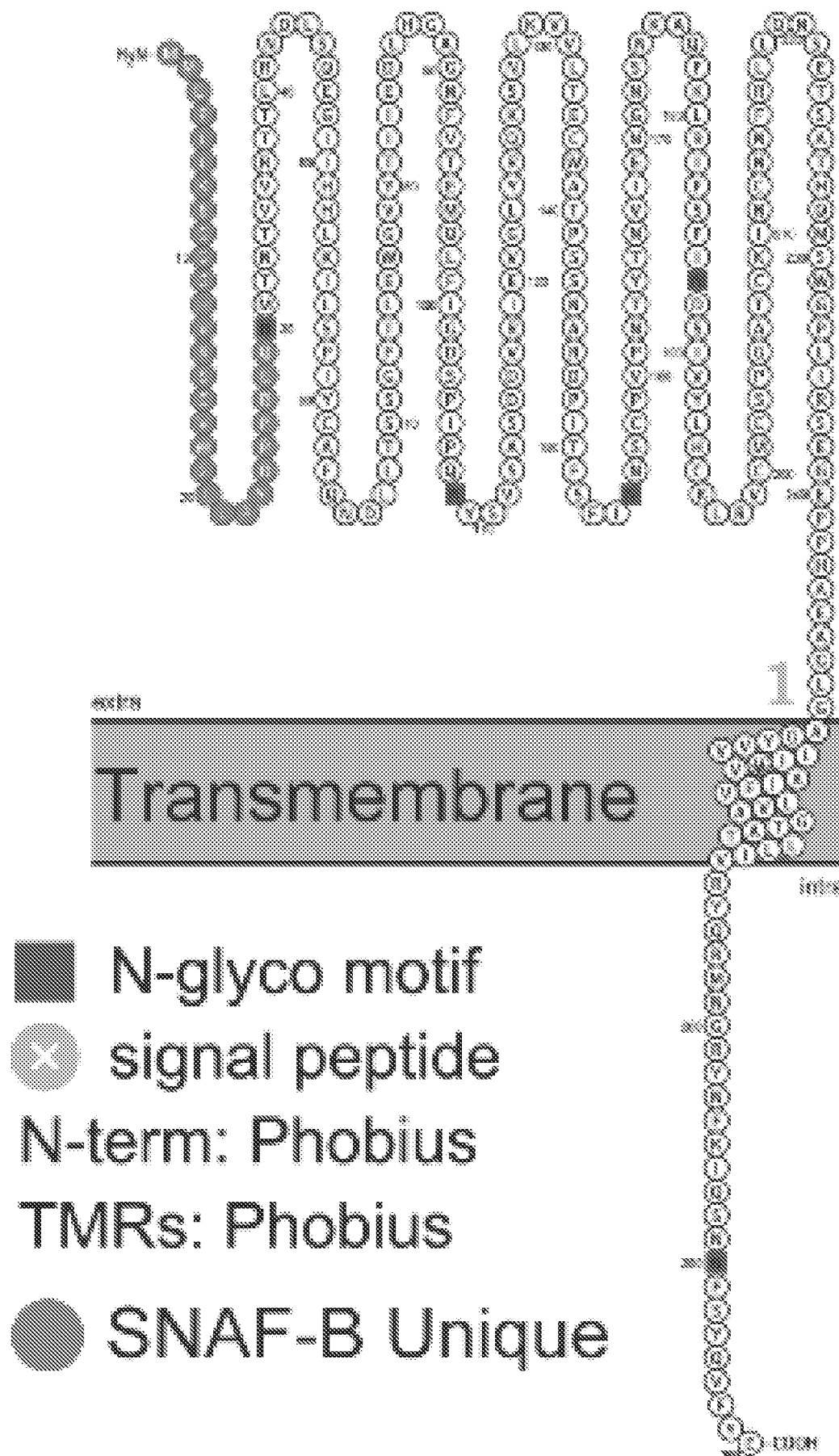
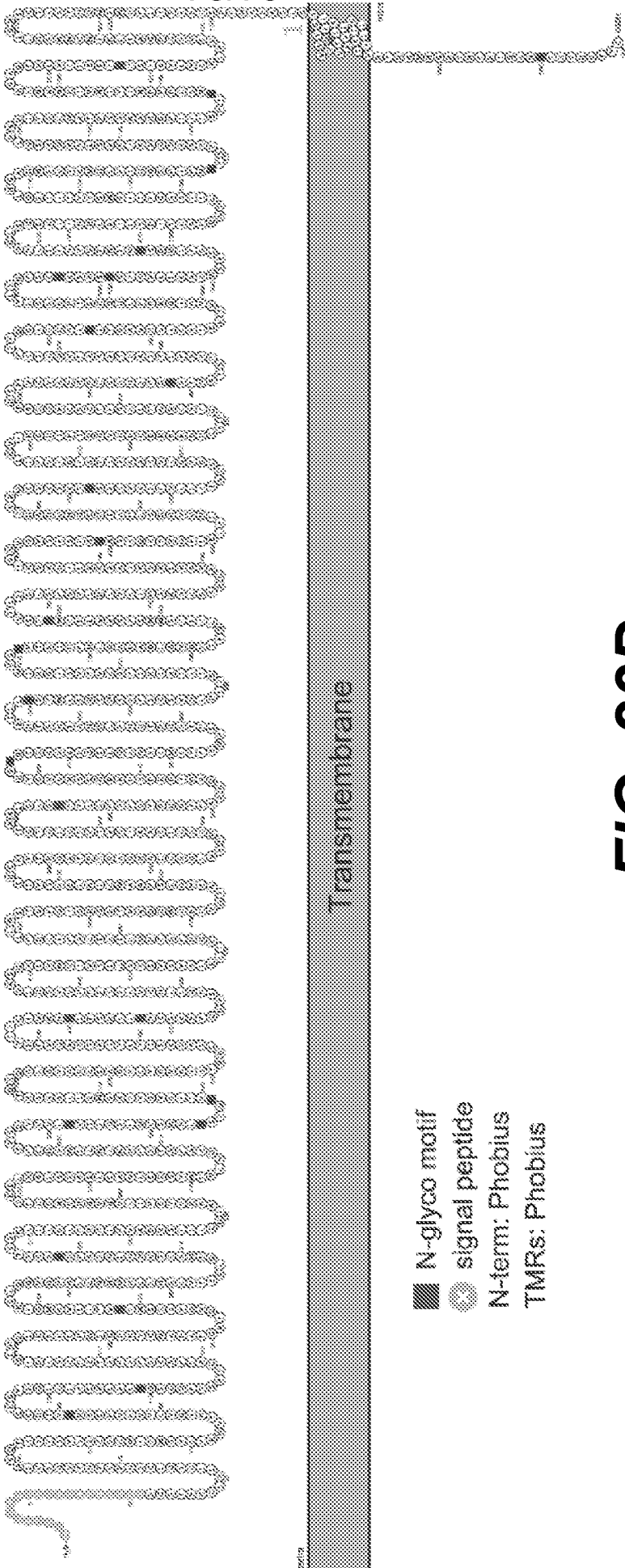


FIG. 23B

72/78

**FIG. 23C**



■ N-glyco motif
⊗ signal peptide
N-term: Phobius
TMRs: Phobius

FIG. 23D

FROM FIG. 24A

↓

Selected Neointigen

neointigen	uid	mean_percent_samples_junction_present	actual_percent_samples_junction_present	identity	length
GREVSKARR	ENS000000102316:E5.1-E5.3	0.3427865101694916	0.0933203389830508	low	9
LYATHRR	ENS000000168918:E3.1-E4.1	0.8961864486779662	0.068864866779661	low	9
LFLAAZER	ENS000000234961:E1.1-E2.1	1	0.0647437627118644	low	9
DSIFQMLR	ENS000000077438:E4.1-E5.1	0.9004237398135594	0.0783898395084745	low	9
YQVRFQMR	ENS000000100318:E2.1-E3.1	0.85593229033898306	0.0763731864486779	low	9
FLREYFMR	ENS000000100246:E4.0-E5.1	0.809322033898305	0.0741525423728613	low	9
LFRGRTSR	ENS000000129877:E7.2-E7.1_3221420	0.8850847457627118	0.0699157542372663	low	9
HRLLECIK	ENS0000000212472:E1.1-ENS000000211899:E1.1_185856217	0.6673728313559322	0.0636779661016949	low	9
YFERRQMR	ENS000000100946:E4.0-E5.1	0.809322033898305	0.0656779661016949	low	9
APLREYFER	ENS000000100946:E4.0-E5.1	0.809322033898305	0.0636779661016949	low	9

FIG. 24A (CONT. 1)

All Exons in AltAnalyze Gene Model

subexon	chr	strand	start	end
E1.1	chr20	+	1875384	1875387
E1.2	chr20	+	1875387	1875389
E1.1	chr20	+	1875384	1875424
E2.1	chr20	+	1875425	1875442
E2.1	chr20	+	1875443	1875585
E2.1	chr20	+	1875587	1875612
E3.1	chr20	+	1875613	1875825
E4.1	chr20	+	1875826	1876013
E4.2	chr20	+	1876013	1876080
E4.3	chr20	+	1876080	1876082

All related existing transcripts

index	transcript	transcript	transcript
5	ENSG00000198853	ENST00000358825	ENSP000002348387
9	ENSG00000198853	ENST00000404852	ENSP00000332941
17	ENSG00000198853	ENST00000358771	ENSP00000331621

FIG. 24A (CONT. 2)

SNAF B-antigen Viewer

Splicing Event

Splicing Event UID: ENSG00000198053:E7.2-El3.1_1915159

SIRPA -- chr20: 1895526-1915159(+)

Expression (Normal -> Tumor)

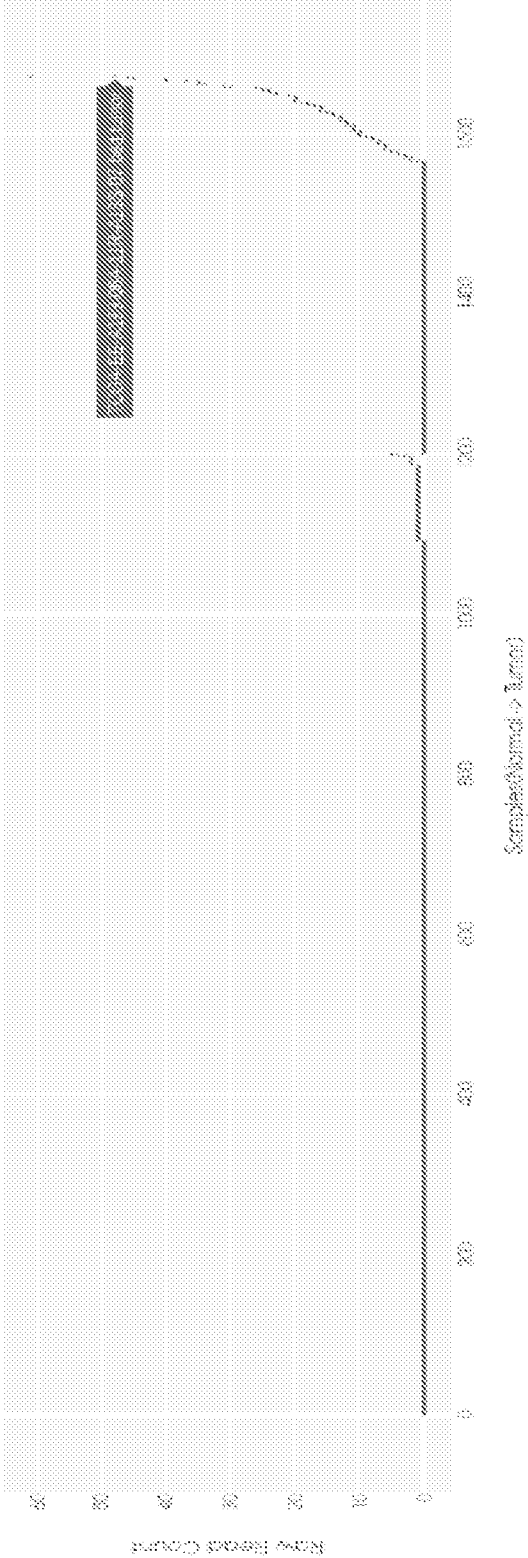


FIG. 24B

Transcript index

5

✕

cDNA or peptide

☐ full length ☐ orft ☒ orfp ☐ junction ☐ score

Submit



Compare with
Existing Transcripts

Sequence

MEPAGPAPGRLGPLLCLLLAASCWVSAGETATLRCTATSLIPVGPQWFRGAGPGRELINQKEGHFRVTIVSDLTNRNNMDFSIRIGNITPADAGTYCYVKFRKGSPPDDVEFKSGAGTE

Ensembl Reference

MEPAGPAPGRLGPLLCLLLAASCWVSAGETATLRCTATSLIPVGPQWFRGAGPGRELINQKEGHFRVTIVSDLTNRNNMDFSIRIGNITPADAGTYCYV

Ensembl Reference

ENR0055_001	1	MEPAGPAPGRLGPLLCLLLAASCWVSAGETATLRCTATSLIPVGPQWFRGAGPGRELINQKEGHFRVTIVSDLTNRNNMDFSIRIGNITPADAGTYCYVKFRKGSPPDDVEFKSGAGTE	50
ENR0055_001	1	MEPAGPAPGRLGPLLCLLLAASCWVSAGETATLRCTATSLIPVGPQWFRGAGPGRELINQKEGHFRVTIVSDLTNRNNMDFSIRIGNITPADAGTYCYV	50
ENR0055_001	101	MMHFSLRIGITPPADAGTYCYVKFRKGSPPDDVEFKSGAGTELSIRAKPQSA	150
ENR0055_001	81	MMDFSLRIGITPPADAGTYCYVKFRKGSPPDDVEFKSGAGTELSIRAKPQSA	130
ENR0055_001	151	PVYSGPAPAPATPGHIVSFTCESHGFSPKQITLNMFRNMSLSDPQTNRDP	200
ENR0055_001	131	PVYSGPAPAPATPGHIVSFTCESHGFSPKQITLNMFRNMSLSDPQTNRDP	180
ENR0055_001	201	VGESVQSYSLHGLAKVNLTREDNHSQVCELVAPVCLQKQPLNQTARLSETI	250
ENR0055_001	181	VGESVQSYSLHGLAKVNLTREDNHSQVCELVAPVCLQKQPLNQTARLSETI	230

Emboss Needle Alignment (peptide)

FIG. 24B (CONT.)