(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
04 October 2018 (04.10.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/183544 A1

(51) International Patent Classification:
C12Q 1/6809 (2018.01) *C12Q 1/6886* (2018.01)
C12Q 1/6881 (2018.01) *G01N 33/52* (2006.01)

(21) International Application Number:
PCT/US2018/024905

(22) International Filing Date:
28 March 2018 (28.03.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/479,787 31 March 2017 (31.03.2017) US

(71) Applicant: **DANA-FARBER CANCER INSTITUTE, INC.** [US/US]; 450 Brookline Avenue, Boston, Massachusetts 02215 (US).

(72) Inventors: **VAN ALLEN, Eliezer**; 42 Addington Road #1, Brookline, Massachusetts 02445 (US). **SMART, Alicia**; 450 Brookline Avenue, Boston, Massachusetts 02215 (US). **MARGOLIS, Claire**; 1 Nashua Street, #1911, Boston, Massachusetts 02114 (US). **MIAO, Diana**; 63 Pope Road, Acton, Massachusetts 01720 (US).

(74) Agent: **ELRIFI, Ivor** et al.; Cooley LLP, 1299 Pennsylvania Avenue, Suite 700, Washington, District of Columbia 20004 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,

(54) Title: METHOD FOR IDENTIFICATION OF RETAINED INTRON TUMOR NEOANTIGENS FROM PATIENT TRANSCRIPTOME

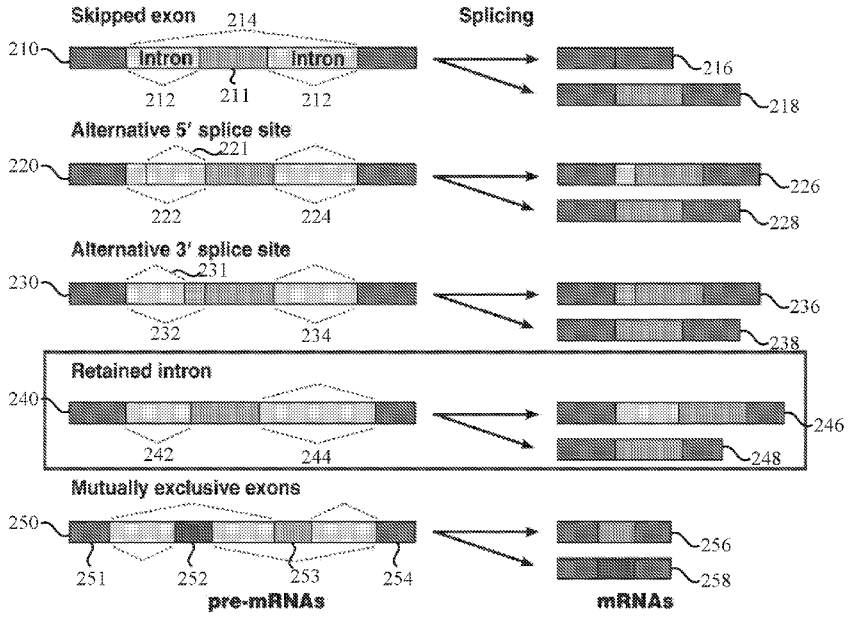

FIG. 2

(57) Abstract: Tumor tissue may be used to generate raw RNA sequencing data. The sequencing data may be aligned and the aligned sequence data may be analyzed to identify retained intron transcripts. The retained introns may be further selected using an expression threshold. The selected retained introns may be converted into corresponding retained intron nucleotide sequences using a reference genomic database, and then translated into retained intron peptides using the open reading frame orientation additionally obtained from the genomic database. DNA from the tumor tissue may be used to generate whole exome sequencing data and analyzed to determine subject-specific HLA alleles. Binding affinities between the retained intron peptides and the HLA alleles may be predicted and retained intron neoantigens may be identified using the retained intron peptides and a binding affinity threshold to identify potential vaccine candidates.

WO 2018/183544 A1

[Continued on next page]

OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, 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))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METHOD FOR IDENTIFICATION OF RETAINED INTRON TUMOR NEOANTIGENS FROM PATIENT TRANSCRIPTOME

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application Serial No. 62/479,787, filed on March 31, 2017, the content of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] This invention relates generally to the identification of tumor specific retained intron neoantigens, and their use in drug discovery and therapeutic compositions, including vaccines.

GOVERNMENT INTEREST

[0003] This invention was made with government support under NIH K08 CA188615 awarded by the National Institutes of Health (NIH). The government has certain rights in the invention.

BACKGROUND

[0004] Cancer immunotherapy is the use of a patient's immune system to treat cancer or to prevent its recurrence. It may be desirable for immunotherapy strategies to stimulate potent, specific and long-lasting anti-tumor immune responses. Vaccines may have the potential to provide long-term protection by inducing endogenous immunological memory. For example, some tumor vaccines comprise one or more tumor antigen epitopes, that may optionally be used in conjunction with adjunctive therapies, including but not limited to immunomodulators (e.g., checkpoint inhibitors) or immunostimulatory molecules (e.g., TLR ligands or cytokines) that may together induce antigen-specific cytotoxic T cells (CTLs) that recognize and lyse tumor cells. These vaccines may contain shared tumor antigens and/or whole tumor cell preparations. The shared tumor antigens may include immunogenic proteins with selective expression in tumors across many patients and are commonly delivered as synthetic peptides or recombinant proteins. Whole tumor cell preparations, on the other hand, may include autologous irradiated cells, cell lysates, cell fusions, heat-shock

protein preparations, and/or total mRNA. Since whole tumor cells are isolated from an autologous patient, the cells may express patient-specific tumor antigens as well as shared tumor antigens.

[0005] Tumor antigens may include proteins with tumor-specific mutations that result in altered amino acid sequences. These mutated proteins may be used to uniquely mark a tumor (relative to non-tumor cells) for recognition and destruction by the immune system, and may also be used to avoid central and/or peripheral T cell tolerance, and thus be recognized by more effective, high avidity T cell receptors. For example, tumor neoantigens arising from somatic mutations in expressed genes may generate novel peptides that are foreign to the immune system and stimulate an immune response. Genes that are not normally expressed in adult somatic tissues, such as cancer germline antigens, may also generate immunogenic peptides. However, conventional approaches to identifying tumor neoantigens rely solely on somatic mutations. As such, additional methods and systems may be desirable to identify subject-specific neoepitopes and neoantigens useful for cancer immunotherapy.

SUMMARY

[0006] Abnormal RNA splicing (e.g., intron retention) is common across cancers, even in the absence of mutations directly related to splicing. Aberrant transcriptional splicing is commonly dysregulated in cancer transcriptomes such that aberrant peptide products generated through the translation and degradation of retained introns may be a source of tumor neoantigens. Transcripts containing retained introns are normally degraded by the nonsense-mediated decay (NMD) pathway and do not lead to expression of full-length proteins. However, the NMD pathway relies on recognition of premature termination codons, which requires that transcripts undergo translation in order to be targeted for degradation. Peptides generated through this pioneer round of translation are a source of antigens presented by the major histocompatibility complex (MHC) I pathway. Presentation of an antigen may occur even when synthesis of its corresponding full-length protein is disrupted. For example, aberrant transcriptional splicing events (e.g., intron retention) may undergo proteolytic cleavage into eight to ten amino acid peptides and then bind to MHC class I or II

molecules for display on a cell surface. T lymphocytes recognize specific antigens through interaction of the T cell receptor (TCR) with short peptides presented by the MHC molecules.

[0007] Described herein are methods and systems for identifying retained intron tumor neoantigens from a subject's transcriptome. Identification of subject-specific retained intron tumor neoantigens may permit study of biologic mechanisms in response to immunotherapy treatment and may be used for cancer immunotherapy. For example, one or more identified retained intron tumor neoantigens may serve as an epitope for stimulation or inducing antibodies in a subject-specific monovalent or polyvalent vaccine composition against a cancer. Additionally, retained intron neoantigens may prove useful as novel clinical biomarkers and/or additions to existing clinical biomarkers, of response or resistance to immune checkpoint blockade therapies. Conventional methods of identifying tumor neoantigens are limited as they rely solely on somatic mutations in DNA and do not identify neoantigens generated by RNA-based aberrant transcriptional splicing events. Thus, the methods and systems described herein may expand the landscape of tumor neoantigens that may be investigated and/or provided therapeutically as a cancer neoantigen vaccine.

[0008] Generally, the methods and systems described herein may identify tumor neoantigens resulting from aberrant splicing events by detecting intron sequences found in RNA sequence data, and identifying resulting neoantigenic epitopes based on expressed intronic sequences and binding affinity to the subject's specific HLA allele types. In general, these methods include the steps of receiving short RNA sequences generated from tumor tissue or other tissue sample from a patient, aligning or pseudoaligning the RNA sequences to reconstruct the transcriptome augmented with intronic sequences, receiving aligned or pseudoaligned transcriptome sequence data and the corresponding retained intron (RI) transcript expression level values, exome data sequenced from the subject's tumor, and HLA class I alleles identified from the exome data. RIs may be identified using the RI transcript expression level values derived from the aligned or pseudoaligned transcriptome sequence data. RI nucleotide sequences may be identified using an RI chromosomal loci and a reference genomic database. The RI nucleotide sequences may be translated into RI

peptide sequences. Binding affinity data may be received by using the RI peptide sequences and the HLA class I alleles. RI neoantigens may be identified using the binding affinity data.

[0009] In some variations, a subject-specific immunogenic composition for administration to the subject may be formulated using one or more of the RI neoantigens. In some variations, genomic coordinates and window sizes corresponding to the set of RIs, RI nucleotide sequences, and RI peptides may be identified. In some of these variations, open reading frame orientations may be identified using the genomic coordinates and the reference genomic database. In some of these variations, identifying the RI nucleotide sequences comprises using the genomic coordinates, window sizes, and open reading frame orientations. In some of these variations, the genomic coordinates may correspond to mutually exclusive intron chromosomal coordinates and include an intron start location. The window size may correspond to a number of amino acids around the intron start location.

[0010] In some variations, the RIs may be expressed at a level of at least about one transcript per million. In some variations, the RI nucleotide sequences may comprise at least one amino acid in an intron. In some variations, the RI neoantigens may comprise the binding affinity of rank less than about 2 percentile or less than about 500 nanomolar. In some of these variations, the RI neoantigens may comprise the binding affinity of rank less than about 0.5 percentile or less than about 50 nanomolar.

[0011] In some variations, identifying the RIs neoantigens may comprise excluding false-positive retained intron events or false-positive RI neoantigens. In some of these variations, excluding false-positive retained introns may comprise applying a zero-coverage filter. In some variations, excluding false-positive retained introns may comprise applying a percent-spliced-in (PSI) filter. In some variations, excluding false-positive retained introns may comprise applying an expression filter. In some variations, excluding false-positive retained intron neoantigens may comprise applying a filter removing RI neoantigens with peptide sequences that are present in a normal proteome to eliminate peptides that likely will not provoke an immune reaction due to host tolerance. In some variations, excluding false-positive RI neoantigens may comprise removing RI

neoantigens originating from introns that are likely retained in normal tissue. In some variations, excluding false-positive retained introns may comprise applying a filter to eliminate likely artefactual or erroneously-annotated retained introns.

[0012] In some variations, raw RNA sequencing data may be received. The aligned or pseudoaligned transcripts may comprise aligned raw RNA sequences. In some variations, transcript information or summary information of the RI neoantigens and distribution information of HLA allele types may be generated.

[0013] In another variation, a method is provided, comprising, receiving a set of RNA sequences from a sample of a subject's tumor, aligning or pseudoaligning the RNA sequences to a transcriptome comprising both intronic and exonic sequences, quantifying an expression level for each transcript in the set of RNA sequences as transcript expression data, identifying retained introns from the set of transcript expression data by excluding exonic sequences or wherein the transcript expression level is below a predetermined level, generating a set of retained intron nucleotide sequences by referencing retained intron loci in a reference genomic database, and translating the set of retained intron sequences into a set of retained intron peptides. The method may further comprise determining a set of binding affinities for the set of retained intron peptides using a set of HLA class I alleles and selecting retained intron neoantigens from the set of retained intron peptides using a pre-determined binding affinity value. The selected retained intron neoantigens may be incorporated into a monovalent or multivalent vaccine.

[0014] In some variations, a system for characterizing a subject's genome is provided, comprising a transceiver configured to receive RNA sequences sequenced from a subject's tumor and aligned or pseudoaligned to a reference transcriptome comprising exonic and intronic sequences, corresponding retained intron (RI) transcript expression values, whole exome sequence data sequenced from the subject's tumor, HLA class I alleles corresponding to the whole exome sequence data, and binding affinities between RI peptides and the HLA class I alleles. A controller may comprise a processor and a memory. The controller may be configured to perform steps including identifying retained introns (RIs) using the RI transcript expression values and an

expression level threshold. RI nucleotide sequences may be identified using RI chromosomal loci and a reference genomic database. The RI nucleotide sequences may be translated into the RI peptides. RI neoantigens may be identified using the binding affinities between the HLA class I alleles and the RI peptides.

[0015] In some variations, the system may further comprise one or more of an RNA sequencing system configured to generate raw RNA sequence data, an alignment or pseudoalignment system configured to align or pseudoalign the raw RNA sequence data to the reference transcriptome or genome comprising the intronic and exonic sequences, a quantification system configured to generate the RI transcript expression values, a DNA sequencing system configured to generate the whole exome sequence data, an HLA typing system configured to generate the HLA class I alleles, and a peptide binding system configured to generate the binding affinities, each coupled to the transceiver.

[0016] In some variations, the controller may be further configured to identify genomic coordinates and window sizes corresponding to the RIs, RI nucleotide sequences, and RI peptides. In some of these variations, the controller may be further configured to identify open reading frame orientations using the intron genomic coordinates and the reference genomic database. In some variations, identifying the RIs comprises excluding false-positive retained introns.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] FIG. 1 is an illustrative diagram of DNA-based and RNA-based neoantigen creation and presentation pathways.

[0018] FIG. 2 is an illustrative diagram of aberrant RNA splicing.

[0019] FIG. 3 is an illustrative graph of subject immunotherapy response and putative contribution of DNA-based and RNA-based neoantigens to the overall neoantigen burden.

[0020] FIGS. 4A-4D are illustrative flowcharts of variations of a method of identifying retained intron neoantigens.

[0021] FIGS. 5A-5B are illustrative block diagrams of variations of a retained intron neoantigen identification system.

[0022] FIG. 6 is an illustrative graph of retained intron counts and corresponding retained intron neoantigen counts of the Snyder and Hugo patient study cohorts.

[0023] FIG. 7 is an illustrative graph of various neoantigen counts for patients of the Snyder and Hugo patient study cohort.

[0024] FIG. 8 is an illustrative graph of various counts and response statuses for subjects of the Snyder and Hugo patient study cohorts.

[0025] FIG. 9 is an illustrative flowchart of a retained intron neoantigen identification process for a tumor cell line.

[0026] FIG. 10 is an illustrative flowchart of a retained intron neoantigen identification process for a set of tumor cell lines.

DETAILED DESCRIPTION

[0027] Generally described here are methods and systems for characterizing a sample genome. More particularly described are methods and systems for identifying sample-specific retained intron tumor neoantigens. Accordingly, it may be helpful to briefly describe neoantigen formation, aberrant RNA splicing, and neoantigen burden.

[0028] The schematic diagram of FIG. 1 generally illustrates various mechanisms by which neoantigens may arise including via somatic mutations and retained introns. Neoantigens may arise from DNA-based events (e.g., somatic mutations) and RNA-based events (e.g., retained introns). A cell nucleus is illustrated with two molecules (101, 111) of DNA. A first DNA molecule (101) may

be subject to a spontaneous somatic mutation while a second DNA molecule (111) (e.g., wild type (WT)) does not undergo mutation. An RNA transcript (102) may be transcribed from the mutated DNA (101). This mutation-bearing transcript may then be spliced into mRNA (103) and translated to produce a mutated protein (104). The mutated protein (104) may undergo proteosomal degradation (105) to 9- or 10-mer peptides (106). A peptide (106) bearing the mutation may then presented on a cell surface by a MHC class I molecule as a neoantigen.

[0029] The second DNA molecule (111) may be transcribed into a WT RNA transcript (112). A splicing error may result in an mRNA (113) including a retained intron. This mRNA (113) including a retained intron may be translated into an abnormal protein (114) with a sequence corresponding to the mRNA (113) with the retained intron. The abnormal protein (114) may then be subject to degradation (115) through the NMD pathway and processed into 9- or 10-mer peptides (116). Any abnormal peptide (116) arising from the retained intron may function as a neoantigen when presented on a cell surface by a MHC class I molecule.

[0030] FIG. 2 is a schematic diagram of aberrant RNA splicing where abnormally-spliced mRNAs may arise from a pre-mRNA transcript through various splicing mechanisms. For example, first pre-mRNA (210) may undergo splicing that skips an exon where an RNA segment (214) includes introns (212) and an exon (211) which are spliced to form first mRNA (216) without an exon. Introns (212) may also be removed to yield second mRNA (218). A second pre-mRNA (220) may undergo alternative 5′ splicing where intron (224) may be spliced with a shorter intron portion (221) of the intron (222) at an alternative 5′ splice site to yield a third mRNA (226). Introns (222, 224) may be spliced from the second pre-mRNA (220) to yield a fourth mRNA (228) shorter than the third mRNA (226). A third pre-mRNA (230) may undergo alternative 3′ splicing where intron (234) may be spliced with a shorter intron portion (231) at an alternative 3′ splice site to yield a fifth mRNA (236) longer than a sixth mRNA (238). Introns (232, 234) may be spliced from the third pre-mRNA (230) to yield the sixth mRNA (238). A fourth pre-mRNA (240) may undergo splicing to yield retained intron mRNA (246, 248). Introns (242, 244) may be spliced out the fourth pre-mRNA (240) to yield eighth mRNA (248). Intron (242) may be spliced out of the fourth pre-mRNA (240)

while intron (242) may be retained to yield an abnormal seventh mRNA (246). A fifth pre-mRNA (250) may undergo mutually exclusive exon splicing. For example, exons (251, 253, 254) may be spliced to yield ninth mRNA (256). Exons (251, 252, 254) may also be spliced to yield tenth mRNA (258).

[0031] FIG. 3 is a graph of cancer immunotherapy response where somatic mutation and retained intron neoantigen burden may be predictive of subject response to immunotherapy. As shown in FIG. 3, responders to tumor immunotherapy may have a higher total neoantigen burden than non-responders lacking an anti-tumor immune response. The total neoantigen load may include somatic neoantigens (310) derived from DNA and retained intron neoantigens (320) derived from RNA.

I. Methods

[0032] Described herein are methods for identifying a retained intron neoantigen using the systems and devices described herein. This may have one or more benefits, such as improved tumor immune surveillance, as well as identification of additional types of subject-specific neoantigens (e.g., RNA based) for cancer vaccine development or immunotherapy biomarker application. For example, the methods described herein may permit tumor vaccines to be formulated based on neoantigens arising from aberrant transcriptional splicing events. Vaccine compositions may be formulated to treat one or more of skin cancer, breast cancer, cervical cancer, colon cancer, liver cancer, lung cancer, non-Hodgkin lymphoma, ovarian cancer, prostate cancer, renal cancer, autoimmune disorders, genetic disorders, and the like. As described in detail herein, a controller of a computing device (e.g., server, server cluster, distributed computing system, desktop PC, laptop, and the like) may be configured to perform one or more steps of the neoantigen identification process. One of more of the steps in the methods described may be performed by separate computing devices and/or computing devices over one or more networks.

[0033] Generally, the methods described here include identifying retained intron neoantigen epitopes based on a tumor sample. The process may begin by using a tissue sample of a tumor to generate raw (e.g., unaligned, short read) RNA sequence data. The raw RNA sequence data may be

aligned with respect to a reference genome or transcriptome (e.g., hg19, hg38) which may be augmented to contain both intronic and exonic sequences, and output as aligned transcriptome sequence data. The retained intron expression values (e.g., intron retention event abundance estimates) of the aligned sequence data may be determined and used to identify a set of retained introns, which may be identified by their locations in terms of genomic coordinates. For example, retained intron locations may be identified using the retained intron expression values compared to a retained intron expression threshold. The retained intron locations may be converted into retained intron nucleotide sequences using a reference genomic database (e.g., UCSC Table Browser; Karolchik et al. 2004) and then translated into retained intron amino acid, or peptide, sequences. A DNA sample of the tumor may be used to generate exome data and analyzed to determine subject-specific HLA class I and/or class II alleles. Binding affinities between the retained intron peptide sequences and the HLA class I and/or class II alleles may be predicted computationally. Retained intron neoantigens may then be identified based upon the binding affinities of the retained intron peptides to the subject's HLA class I and/or class II alleles. For example, retained intron neoantigens may be identified using the predicted binding affinity of the retained intron peptides compared to a binding affinity threshold.

[0034] In some variations, a subject-specific immunogenic composition (e.g., vaccine) comprising one or more of the identified retained intron neoantigens may be formulated for administration to the subject. Additionally or alternatively, identified retained intron neoantigens may be validated in vitro in subject-derived cell lines to assess immunogenicity. The methods described herein may be applied to normal tissue controls to investigate the extent of intron retention and self-antigen expression. In some variations, binding assays may be used to empirically assess neoantigen peptide binding to MHC molecules. In other variations, peripheral blood mononuclear cells (PBMC) that match the determined HLA alleles may be harvested from healthy subjects, loaded with peptides *ex vivo*, and administered to the subject.

[0035] FIGS. 4A-4D are flowcharts that generally describe a method (400) of characterizing or assessing a sample (e.g., subject, tumor) genome. Generally, the method (400) identifies intron

retention events from RNA sequencing data, determines open reading frames that extend from exonic into intronic sequences, and identifies retained intron neoantigens predicted to bind to the subject's HLA alleles. The method (400) may begin with generating a set of raw RNA sequence data sequenced from a tumor sample (402) (e.g., from a subject) using an RNA sequencing system. In some variations, the set of raw RNA sequence data may correspond to patient, tissue, and/or cancer specific cohorts. For example, the method (400) may generate RNA sequence data sequenced from any of a skin cancer cohort, breast cancer cohort, cervical cancer cohort, colon cancer cohort, liver cancer cohort, lung cancer cohort, non-Hodgkin lymphoma cohort, ovarian cancer cohort, prostate cancer cohort, renal cancer cohort, autoimmune disorder cohort, genetic disorder cohort, and the like. In some variations, the set of raw transcriptome sequence data may be transmitted by the RNA sequencing system and be received by a neoantigen identification system (403) such as those described herein. The neoantigen identification system may store the set of raw RNA sequence data in memory. In some variations, the raw RNA sequence data may be stored in a FASTQ file for subsequent alignment or pseudoalignment steps.

[0036] The set of raw RNA sequence data may be aligned or pseudoaligned to a transcriptome (404) using an alignment or pseudoalignment system. In some variations, raw RNA sequence data may be aligned or pseudoaligned to an augmented version of the hg19 human transcriptome build containing both exonic and intronic sequences. In some variations, the alignment or pseudoalignment system may execute a sequence alignment or pseudoalignment algorithm and output one or more Binary Alignment Map (BAM) files. BAM is a compressed binary version of a text-based format for storing biological sequences aligned to a reference sequence. The alignment or pseudoalignment system may perform alignment or pseudoalignment steps using a spliced RNA sequence aligner algorithm. The algorithm may include kallisto (Bray et al. 2016), Bowtie, Bowtie2, TopHat, TopHat2, Spliced Transcripts Alignment to a Reference (STAR), and the like. For example, kallisto is a program for efficiently quantifying abundances of transcripts from RNA sequencing data based on the idea of pseudoalignment for rapidly determining the compatibility of reads with targets, without the need for alignment. In some variations, a subject's FASTQ files may be input to the kallisto algorithm for pseudoaligment and quantification. In some variations, the set

of aligned or pseudoaligned transcripts may be transmitted by the alignment or pseudoalignment system and received by the neoantigen identification system (405).

[0037] A set of retained intron transcript expression values corresponding to the set of aligned or pseudoaligned transcriptome data may be determined using an expression quantification system (406). In some variations, the expression quantification system may generate retained intron transcript abundance estimates using the kallisto algorithm for pseudoalignment and quantification (Bray et al., 2016). In some variations, the set of retained intron transcript expression values may be transmitted by the expression quantification system and received by the neoantigen identification system (407).

[0038] A first threshold may be set (408) corresponding to a retained intron transcript expression value. In some variations, the first threshold may be set to a retained intron transcript expression value of at least about one transcript per million (TPM), and any retained intron transcripts with expression values below this threshold may be eliminated. In other variations, the first threshold may be between about 0.25 TPM and about 0.5 TPM. In some further examples, the threshold may be variable depending on the tumor type, patient genotype, intronic chromosome, expression level of the surrounding exons, and/or other characteristics.

[0039] In some variations, false-positive retained introns may be excluded from the set of retained intron transcripts (409). A filter may be applied to the set of retained intron transcripts to exclude false-positive retained introns. False-positive retained introns may be more frequent in low coverage exonic regions and regions having high variations in coverage due to biases or repetitive sequences. In some variations, a zero-coverage filter may be applied to the retained intron transcript data to exclude false-positive retained introns. For example, a zero coverage filter may be configured to determine the longest continuous region in an intron which has no reads starting in that region. The probability of observing a region of a predetermined length with no reads starting in it is determined (given the intron's expression). The intron may be classified as a false-positive and removed if the determined probability is below a predetermined probability threshold. In some variations, a percent-spliced-in (PSI) filter may be applied to exclude false-positive retained introns. For

example, if all samples in a cohort contain exactly 0% or 100% retention of an intron (or other predetermined values), the intron may be classified as false-positive and removed. In some variations, further expression-based filters may be applied to exclude false-positive retained introns. For example, if fewer than a predetermined percentage of samples in a cohort contain greater than a predetermined number of unique intron transcript counts and/or greater than a predetermined intron expression level, then the intron may be classified as a false-positive and removed. For example, for a particular intron, if fewer than 25% of samples in a cohort have expression of the given intron transcript at a level of greater than one TPM and/or fewer than 25% of samples in a cohort have at least five unique counts of the given intron transcript, then the given intron may be classified as a false-positive and removed. In some variations, a normal tissue filter may be applied to remove false-positive retained introns. Retained introns that are present in normal tissue are not likely to yield retained intron neoantigens due to likely subject immune tolerance to peptides produced by these retained introns. Step 409 may serve as a pre-processing step to retained intron detection in step 410.

[0040] A set of retained introns and related characteristics may be identified from the set of aligned transcriptome data using an intron retention detection system given the retained intron expression values, the first threshold, and the false-positive filtering criteria (410). In some variations, the identified characteristics may include a set of retained intron chromosomal loci. In some variations, the intron retention detection system may identify the set of retained introns and related characteristics using the Keep Me Around (KMA) algorithm (Pimentel et al., 2015). KMA incorporates biological replicates to reduce the number of false-positives when detecting intron retention events. In some variations, the intron retention detection system may use the KMA algorithm to output a KMA file comprising a set of retained intron chromosomal loci. The first threshold and false-positive filtering criteria may then be applied to the KMA file to identify the set of retained introns.

[0041] In some variations, a set of genomic coordinates and window sizes corresponding to the set of retained introns may be identified by, for example, the neoantigen identification system (411).

For example, the genomic coordinates and window sizes may be identified by the neoantigen identification system from the KMA file. The set of genomic coordinates may correspond to mutually exclusive retained intron chromosomal coordinates and include an intron start location. The window sizes may correspond to a number of nucleotides around the intron start location.

[0042] A set of retained intron nucleotide sequences may be identified using the set of retained intron genomic coordinates (e.g., unique intron chromosomal locations) and a reference genomic database. In some variations, the reference genomic database may include the UCSC Table Browser online database that may be configured to output the set of nucleotide sequences based upon input genomic coordinates and window size. Furthermore, the reference genomic database may be used to determine a set of open reading frame orientations of the retained intron nucleotide sequences at the intron start loci using the genomic coordinates (412). For example, the reference genomic database may determine whether a retained intron nucleotide sequence is on a plus or minus strand and whether the start of the intron is in the 0, +1, or +2 open reading frame orientation.

[0043] Once the open reading frame orientations have been determined for each of the retained intron start coordinates, the reference genomic database may identify the set of retained intron nucleotide sequences using the sets of genomic coordinates, window sizes, and open reading frame orientations (413). In some variations, the set of retained intron nucleotide sequences may be output in FASTA format.

[0044] A second threshold of at least one amino acid in an intron may be set and applied to each of the retained intron nucleotide sequences (414). The set of retained intron nucleotide sequences may be translated into a set of retained intron amino acid sequences, or peptides, by the neoantigen identification system (415). For example, a codon table mapping nucleotide triplets to corresponding amino acids may be used to translate an input FASTA containing retained intron nucleotide sequences into an output FASTA containing retained intron peptides.

[0045] As depicted in FIG. 4C, a set of raw whole exome sequencing data may be generated by sequencing the exonic DNA from a sample of a subject's tumor (420) using a DNA sequencing

system. In some variations, the whole exome sequencing data may be transmitted by the DNA sequencing system and received by the neoantigen identification system (421). A set of HLA class I alleles corresponding to the whole exome sequencing data may be determined using an HLA typing system (422). In some variations, the whole exome sequencing data may be processed by the HLA typing system using Polysolver (polymorphic loci resolver) (Shukla et al., 2015). Polysolver is an algorithm for inferring alleles of the three major MHC class I (HLA-A, HLA-B, HLA-C) genes. For example, subject HLA class I alleles may be inferred from whole exome sequencing data by the HLA typing system using Polysolver. In other variations, in the absence of whole exome sequencing data, subject HLA class I alleles may be inferred from transcriptome sequencing data by an alternative HLA typing algorithm. In some variations, the set of HLA class I alleles may be transmitted by the HLA typing system and received by the neoantigen identification system (423).

[0046] A set of binding affinities between the set of retained intron peptides and the set of HLA class I alleles may be determined using a peptide binding assessment system (424). In some variations, the set of peptide-HLA binding affinities may be determined by the peptide binding assessment system using a peptide binding algorithm for predicting the strength of binding between a peptide and an MHC molecule. For example, a FASTA file containing retained intron peptides and the set of HLA class I alleles may be input to the peptide binding assessment system where one or more versions of a peptide binding algorithm may be run. In some variations, sequences of 9-10 amino acids generated from the retained intron peptide sequences may be analyzed using one or more of NetMHCpan v3.0 (Nielsen et al., 2016), NetMHC, NetMHCcons, and the like, to identify peptides that are predicted to bind to the HLA alleles. In some variations, a set of predicted binding affinities may be transmitted by the peptide binding assessment system and received by the neoantigen identification system (425).

[0047] A third threshold may be set (426) corresponding to a binding affinity. In some variations, the third threshold may comprise a binding affinity of rank less than about 2 percentile or less than about 500 nanomolar. In some variations, the third threshold may comprise a binding affinity of rank less than about 0.5 percentile or less than about 50 nanomolar. In some variations, false-

positive retained intron peptides may be excluded where these peptides may have the same amino acid sequence as peptides in the normal proteome or arise from introns retained in normal tissue (427). These retained intron peptides are not likely to be neoantigens because of pre-existing immune tolerance, and they may be excluded. In some variations, a set of retained intron neoantigens may be identified using the sets of peptide-HLA binding affinities, the third threshold, and the set of retained intron peptides (428). For example, retained intron neoantigens may be identified from the retained intron peptides using the set of binding affinities compared to the third threshold.

[0048] In some variations, summary and/or aggregate information from the set of retained intron neoantigens and HLA alleles from each sample may be generated by the neoantigen identification system. For example, for every sample, data including one or more of therapy response status, number of retained introns, number of total, strong, and weak retained intron neoantigens, number of total neoantigens (somatic neoantigens plus retained intron neoantigens), and HLA type may be generated as a set of analysis characteristics. A distribution of HLA types may be generated from a cohort of subjects. In some variations, retained intron neoantigen data may be generated and include one or more of the retained intron peptide, intron chromosome location, sample data such as response status to immunotherapy, and HLA type.

[0049] In some variations, a subject-specific immunogenic composition may be formulated for administration to the subject using one or more of the retained intron neoantigens (429). For example, a vaccine composition may be configured to raise a specific cytotoxic T cell response and/or a specific helper T cell response. A vaccine composition may comprise between one peptide and about 20 peptides. In some variations, different peptides and/or polypeptides may be selected such that the vaccine composition comprises peptides and/or polypeptides capable of associating with different MHC molecules such as different MHC class I molecules. In some variations, the selection, number, and/or amount of neoantigen peptides in the composition may be tissue, cancer, and/or patient-specific. For example, vaccine compositions may be specific to any cancer such as skin cancer, breast cancer, cervical cancer, colon cancer, liver cancer, lung cancer, non-Hodgkin

lymphoma, ovarian cancer, prostate cancer, renal cancer, an autoimmune disorders, and a genetic disorder. The vaccine composition comprising at least one antigen-presenting cell may be pulsed or loaded with one or more of the identified neoantigen peptides. Alternatively, peripheral blood mononuclear cells (PBMC) isolated from a subject may be loaded with peptides ex vivo and injected into the subject.

[0050] In some variations, the subject-specific immunogenic composition may comprise an adjuvant and/or a carrier. Adjuvants may comprise a substance whose admixture into a vaccine composition may increase or otherwise modify the subject's immune response to the neoantigen peptide. Additionally or alternatively, adjuvants may be conjugated covalently or non-covalently to the identified neoantigen peptides. A vaccine composition may comprise one or more types of adjuvants. The one or more neoantigen peptides and the adjuvant may be administered separately in any appropriate sequence.

[0051] In some variations, one or more neoantigen peptides in the composition may be associated with a carrier such as a protein or an antigen-presenting cell (e.g., dendritic cell) configured to present the peptide to a T cell. Carriers may comprise scaffold structures such as a polypeptide and/or polysaccharide, to which the neoantigen peptides are capable of being associated. A carrier may be present independently of an adjuvant. The carrier may be configured to increase the molecular weight of one or more neoantigen peptides in order to increase their activity or immunogenicity, to confer stability, to increase the biological activity, or to increase serum half-life. Furthermore, a carrier may aid in presenting the neoantigen peptides to T cells. In some variations, the carrier may be a protein or an antigen presenting cell. For example, a carrier protein may be a keyhole limpet hemocyanin, serum protein such as transferrin, bovine serum albumin, human serum albumin, thyroglobulin or ovalbumin, immunoglobulins, or hormones such as insulin or palmitic acid. In some variations, the carrier may comprise one or more of tetanus toxoid, diptheria toxoid, and dextrans.

[0052] The subject-specific immunogenic composition may be administered alone or in combination with other therapeutic agents. The therapeutic agent may be, for example, a

chemotherapeutic, radiation, or immunotherapy agent. In some variations, the subject may be further administered an anti-immunosuppressive and/or immunostimulatory agent. The immunogenic composition may be configured for intravenous, sub-cutaneous, intradermal, intraperitoneal, and intramuscular injection. In some variations, the immunogenic composition may be administered at the site of surgical excision to induce a local immune response to the tumor. The immunogenic composition may also be administered using liposomes configured to target particular cell tissue such as lymphoid tissue. In therapeutic applications, an immunogenic composition may be administered to a patient in an amount sufficient to elicit an effective response to the tumor antigen. The concentration of neoantigen peptides may range from less than about 0.1% to about 50% or more by weight.

II. Systems

[0053] FIGS. 5A-5B are block diagrams of a variation of a processing system (500). The system (500) may comprise a retained intron neoantigen identification system (520). The identification system (520) may be coupled to one or more networks (570), databases (540), and/or servers (550) through one or more wired or wireless communication channels. The network (570) may comprise one or more databases (540), servers (550), and computing devices (560). The one or more databases (540), servers (550), and computing devices (560) may include, for example, an RNA sequencing system, alignment or pseudoalignment system, expression quantification system, intron retention detection system, reference genomic database, DNA sequencing system, HLA typing system, and peptide binding assessment system as described herein.

[0054] In some variations, a remote operator (not shown) may be coupled to one or more networks (570), databases (540), and servers (550) through a computing device (560). One or more of the steps described herein for identifying retained intron neoantigens may be performed at any one of the computing devices of the system (500) or distributed throughout a plurality of computing devices. For example, generation of raw RNA sequence data may be performed by an RNA sequencing system and generation of whole exome sequence data may be performed by a DNA sequencing system. Other steps such as applying thresholds to identify different data sets (e.g.,

identify retained introns or retained intron neoantigens) may be performed by the identification system (520). In some variations, alignment or pseudoalignment to a reference transcriptome and quantification of RNA sequence data may be performed by the identification system (520).

Identification System

[0055] FIG. 5B is a block diagram of the identification system (520). The identification system (520) may comprise a controller (522) comprising a processor (524) and a memory (526). In some variations, the identification system (520) may further comprise one or more of a communication interface (530). The controller (522) may be coupled to a communication interface (530) to permit an operator to remotely control the identification system (520) and any other component of the system (500). The communication interface (530) may comprise a network interface (532) configured to connect the identification system (520) to another system (e.g., Internet, remote server, database, computing device) over a wired and/or wireless network. The communication interface (530) may further comprise a user interface (534) configured to permit an operator to directly control the identification system (520).

A. Controller

[0056] The identification system (520), as depicted in FIG. 5B, may comprise a controller (522) in communication with the processing system (500). The controller (520) may comprise one or more processors (524) and one or more machine-readable memories (526) in communication with the one or more processors (524). The processor (524) may incorporate data received from memory (526) and operator input to control the system (500). The memory (526) may further store instructions to cause one or more processors (524) to execute modules, processes, and/or functions associated with the system (500). The controller (522) may be configured to control one or more components of the system (500), such as database (540), server (550), network (570), communication interface (530), and the like. For example, the identification system (520) may provide input (e.g., genomic coordinates) to a reference genomic database (e.g., UCSC Table Browser). The reference genomic database may return a set of nucleotide sequences to the memory

(526) of the identification system (520). The identification system (520) may further process received data (e.g., format data) such that the output data of one system is compatible with the input data requirements of another system.

[0057] The controller (522) may be implemented consistent with numerous general purpose or special purpose computing systems or configurations. Various exemplary computing systems, environments, and/or configurations that may be suitable for use with the systems and devices disclosed herein may include, but are not limited to software or other components within or embodied on servers or server computing devices such as routing/connectivity components, multiprocessor systems, microprocessor-based systems, distributed computing networks, personal computing devices, network appliances, portable (e.g., hand-held) or laptop devices. Examples of portable computing devices include smartphones, personal digital assistants (PDAs), cell phones, tablet PCs, wearable computers taking the form of portable or wearable augmented reality devices that interface with an operator's environment through sensors and may use head-mounted displays for visualization, eye gaze tracking, and user input.

i. Processor

[0058] The processor (524) may be any suitable processing device configured to run and/or execute a set of instructions or code and may include one or more data processors, image processors, graphics processing units, physics processing units, digital signal processors, and/or central processing units. The processor (524) may be, for example, a general purpose processor, Field Programmable Gate Array (FPGA), an Application Specific Integrated Circuit (ASIC), and the like. The processor (524) may be configured to run and/or execute application processes and/or other modules, processes and/or functions associated with the system and/or a network associated therewith. The underlying device technologies may be provided in a variety of component types including metal-oxide semiconductor field-effect transistor (MOSFET) technologies like complementary metal-oxide semiconductor (CMOS), bipolar technologies like emitter-coupled logic (ECL), polymer technologies (e.g., silicon-conjugated polymer and metal-conjugated polymer-metal structures), mixed analog and digital, and the like.

ii. Memory

[0059] In some variations, the memory (526) may include a database and may be, for example, a random access memory (RAM), a memory buffer, a hard drive, an erasable programmable read-only memory (EPROM), an electrically erasable read-only memory (EEPROM), a read-only memory (ROM), Flash memory, and the like. As used herein, memory refers to a data storage resource. The memory (526) may store instructions to cause the processor (524) to execute modules, processes and/or functions associated with the identification system (520), such as data processing, communication, and/or user settings. In some variations, storage may be network-based and accessible for one or more authorized users. Network-based storage may be referred to as remote data storage or cloud data storage. Sequence data and neoantigen data stored in cloud data storage (e.g., database) may be accessible to respective users via a network (570), such as the Internet. In some variations, database (540) may be a cloud-based FPGA.

[0060] Some variations described herein relate to a computer storage product with a non-transitory computer-readable medium (also may be referred to as a non-transitory processor-readable medium) having instructions or computer code thereon for performing various computer-implemented operations. The computer-readable medium (or processor-readable medium) is non-transitory in the sense that it does not include transitory propagating signals per se (e.g., a propagating electromagnetic wave carrying information on a transmission medium such as space or a cable). The media and computer code (also may be referred to as code or algorithm) may be those designed and constructed for the specific purpose or purposes. Examples of non-transitory computer-readable media include, but are not limited to, magnetic storage media such as hard disks, floppy disks, and magnetic tape; optical storage media such as Compact Disc/Digital Video Discs (CD/DVDs); Compact Disc-Read Only Memories (CD-ROMs); holographic devices; magneto-optical storage media such as optical disks; solid state storage devices such as a solid state drive (SSD) and a solid state hybrid drive (SSHD); carrier wave signal processing modules; and hardware devices that are specially configured to store and execute program code, such as Application-Specific Integrated Circuits (ASICs), Programmable Logic Devices (PLDs), Read-Only Memory

(ROM), and Random-Access Memory (RAM) devices. Other variations described herein relate to a computer program product, which may include, for example, the instructions and/or computer code disclosed herein.

[0061] The systems, devices, and/or methods described herein may be performed by software (executed on hardware), hardware, or a combination thereof. Hardware modules may include, for example, a general-purpose processor (or microprocessor or microcontroller), a field programmable gate array (FPGA), and/or an application specific integrated circuit (ASIC). Software modules (executed on hardware) may be expressed in a variety of software languages (e.g., computer code), including C, C++, Java®, Python, Ruby, Visual Basic®, and/or other object-oriented, procedural, or other programming language and development tools. Examples of computer code include, but are not limited to, micro-code or micro-instructions, machine instructions, such as produced by a compiler, code used to produce a web service, and files containing higher-level instructions that are executed by a computer using an interpreter. Additional examples of computer code include, but are not limited to, control signals, encrypted code, and compressed code.

B. Communication interface

[0062] The communication interface (530) may permit an operator to interact with and/or control the system (500) directly and/or remotely. For example, a user interface (534) of the system (500) may include an input device for an operator to input commands and an output device for an operator and/or other observers to receive output (e.g., view subject data on a display device) related to operation of the system (500). In some variations, a network interface (532) may permit the control system (520) to communicate with one or more of a network (570) (e.g., Internet), remote server (550), database (540), and computing devices (560) as described in more detail herein.

i. User interface

[0063] User interface (534) may serve as a communication interface between a user (e.g., operator) and the control system (520). In some variations, the user interface (534) may comprise an input device and output device (e.g., touch screen and display) and be configured to receive input

data and output data from one or more of the input device, output device, network (570), database (540), and server (550). For example, data generated by server (550) may be processed by processor (524) and memory (526), and output visually by one or more output devices. Neoantigen data may be received by user interface (534) from memory (526) and output visually, audibly, and/or through haptic feedback through one or more output devices. As another example, operator control of an input device (e.g., joystick, keyboard, touch screen) may be received by user interface (534) and then processed by processor (524) and memory (526) for user interface (534) to output a control signal to one or more of the network (570), database (540), and server (550).

1. Output device

[0064] An output device of a user interface (534) may output data corresponding to a subject and/or system (500), and may comprise one or more of a display device, audio device, and haptic device. The display device may be configured to display a graphical user interface (GUI). A display device may permit an operator to view one or more of sample data, subject data, system data, sequence data, binding and allele data, neoantigen data, peptide data, and/or other data received and/or processed by the controller (522). In some variations, an output device may comprise a display device including at least one of a light emitting diode (LED), liquid crystal display (LCD), electroluminescent display (ELD), plasma display panel (PDP), thin film transistor (TFT), organic light emitting diodes (OLED), electronic paper/e-ink display, laser display, and/or holographic display.

[0065] An audio device may audibly output sequence data, binding and allele data, neoantigen data, peptide data, sample data, subject data, system data, alarms and/or warnings. In some variations, an audio device may comprise at least one of a speaker, piezoelectric audio device, magnetostrictive speaker, and/or digital speaker. In some variations, an operator may communicate with other users using the audio device and a communication channel.

[0066] A haptic device may be incorporated into one or more of the input and output devices to provide additional sensory output (e.g., force feedback) to the operator. For example, a haptic

device may generate a tactile response (e.g., vibration) to confirm operator input to an input device (e.g., joystick, keyboard, touch surface).

2. Input device

[0067] Some variations of an input device may comprise at least one switch configured to generate a control signal. In some variations, the input device may comprise a wired and/or wireless transmitter configured to transmit a control signal to a wired and/or wireless receiver of a controller (522). For example, an input device may comprise a touch surface for an operator to provide input (e.g., finger contact to the touch surface) corresponding to a control signal. An input device comprising a touch surface may be configured to detect contact and movement on the touch surface using any of a plurality of touch sensitivity technologies including capacitive, resistive, infrared, optical imaging, dispersive signal, acoustic pulse recognition, and surface acoustic wave technologies. In variations of an input device comprising at least one switch, a switch may comprise, for example, at least one of a button (e.g., hard key, soft key), touch surface, keyboard, analog stick (e.g., joystick), directional pad, pointing device (e.g., mouse), trackball, jog dial, step switch, rocker switch, pointer device (e.g., stylus), motion sensor, image sensor, and microphone. A motion sensor may receive operator movement data from an optical sensor and classify an operator gesture as a control signal. A microphone may receive audio and recognize an operator voice as a control signal.

ii. Network interface

[0068] As depicted in FIG. 5A, an identification system (520) described herein may communicate with one or more networks (570) and servers (550) through a network interface (532). In some variations, the identification system (520) may be in communication with other devices via one or more wired and/or wireless networks. In some variations, the network interface (532) may facilitate communication with other devices over one or more external ports (e.g., Universal Serial Bus (USB), multi-pin connector) configured to couple directly to other devices or indirectly over a network (e.g., the Internet, wireless LAN).

[0069] In some variations, the network interface (530) may comprise radiofrequency (RF) circuitry (e.g., RF transceiver) including one or more of a receiver, transmitter, and/or optical (e.g., infrared) receiver and transmitter configured to communicate with one or more devices and/or networks. RF circuitry may receive and transmit RF signals (e.g., electromagnetic signals). The RF circuitry converts electrical signals to/from electromagnetic signals and communicates with communications networks and other communications devices via the electromagnetic signals. The RF circuitry may include one or more of an antenna system, an RF transceiver, one or more amplifiers, a tuner, one or more oscillators, a digital signal processor, a CODEC chipset, a subscriber identity module (SIM) card, memory, and the like. A wireless network may refer to any type of digital network that is not connected by cables of any kind. Examples of wireless communication in a wireless network include, but are not limited to cellular, radio, satellite, and microwave communication. The wireless communication may use any of a plurality of communications standards, protocols and technologies, including but not limited to Global System for Mobile Communications (GSM), Enhanced Data GSM Environment (EDGE), high-speed downlink packet access (HSDPA), wideband code division multiple access (W-CDMA), code division multiple access (CDMA), time division multiple access (TDMA), Bluetooth, near-field communication (NFC), radio-frequency identification (RFID), Wireless Fidelity (Wi-Fi) (e.g., IEEE 802.11a, IEEE 802.11b, IEEE 802.11g, IEEE 802.11n), Voice over Internet Protocol (VoIP), Wi-MAX, a protocol for email (e.g., Internet Message Access Protocol (IMAP), Post Office Protocol (POP)), instant messaging (e.g., eXtensible Messaging and Presence Protocol (XMPP), Session Initiation Protocol for Instant Messaging, Presence Leveraging Extensions (SIMPLE), Instant Messaging and Presence Service (IMPS)), Short Message Service (SMS), or any other suitable communication protocol. Some wireless network deployments combine networks from multiple cellular networks or use a mix of cellular, Wi-Fi, and satellite communication. In some variations, a wireless network may connect to a wired network in order to interface with the Internet, other carrier voice and data networks, business networks, and personal networks. A wired network is typically carried over copper twisted pair, coaxial cable, and/or fiber optic cables. There are many different types of wired networks including wide area networks (WAN), metropolitan area networks

(MAN), local area networks (LAN), Internet area networks (IAN), campus area networks (CAN), global area networks (GAN), like the Internet, and virtual private networks (VPN). As used herein, network refers to any combination of wireless, wired, public, and private data networks that are typically interconnected through the Internet, to provide a unified networking and information access system.

III. Examples

[0070] Treatment of metastatic melanoma has benefited from the development of immune checkpoint inhibitors, which for example have improved outcomes and increased response rates by about 60%. However, only a minority of subjects treated experience long-term clinical benefit, and predictors of response are limited. For example, predictors of response to checkpoint inhibitor therapy may include somatic mutations in oncogenes, tumor expression of immune markers such as PD-L1, somatic mutation load, and neoantigen burden. The methods described herein were applied to two cohorts of melanoma patients (Hugo, Snyder) treated with immune checkpoint inhibitor therapy to identify retained intron neoantigens.

[0071] Tumor-specific RI neoantigens may be identified as described herein in both patient-derived and cell line-derived samples and a subset may be validated experimentally via mass spectrometry in complex with MHC I. As described herein, patient-specific neoantigens arising from intron retention events and identified RI neoantigens in tumor samples may be identified from the Hugo and Snyder cohorts of melanoma patients. As described herein, putative RI-neoantigen peptides predicted *in silico* from multiple human tumor cell lines may be found experimentally to be bound to the MHC Class I molecule *in vitro* through mass spectrometry. This suggests that aberrant splicing results in intron retention, which generates abnormal transcripts that are translated into immunogenic peptides and presented to the immune system, underscoring their clinical relevance in patients receiving immunotherapy. Additionally, although RI neoantigen load is not necessarily predictive of response to immune checkpoint blockade therapy as a global measure, a subset of RI neoantigens may be associated with treatment response and may have further clinical relevance for both cancer vaccine formulation and immunotherapy response prediction.

[0072] Identification of tumor neoantigens including those derived from somatic mutation, aberrant gene expression, and splicing dysregulation may contribute to a more complete understanding of the tumor immune landscape, and may improve prediction of individual response to therapy. Prediction of patient-specific RI-neoantigens may contribute to the development and further improvement of personalized cancer vaccines.

A. Cohort composition

[0073] Analysis was conducted on published cohorts of melanoma patients treated with checkpoint inhibitors, as summarized below in Table 1. The Hugo cohort included samples from 27 melanoma patients (26 pre-treatment, 1 on treatment) treated with the PD-1 inhibitor pembrolizumab (Hugo et al., 2016). 14 patients received clinical benefit (R) and 13 patients received no clinical benefit (NR). These samples were sequenced from fresh frozen tissue using a standard, poly(A) selected protocol. The Snyder cohort included samples from 21 melanoma patients treated with a PD-1 inhibitor where 8 patients received long-term clinical benefit (LB) and 13 patients received no clinical benefit (NB). RNA sequencing was performed on fresh frozen tissue using a standard, poly(A) selected protocol.

Table 1

Cohort	Tumor type	Immuno target	Sample size	Response types
Hugo et al. *Cell* 2016	Melanoma	PD-1	27	R: 14 NR: 13
Snyder et al. *NEJM* 2014	Melanoma	PD-1	21	R: 8 NR: 13

B. Cohort Analysis

[0074] FIG. 6 is an illustrative graph of retained intron counts (e.g., load) and corresponding neoantigen counts of the Snyder and Hugo patient study cohorts. The methods described herein may be applied to the two cohorts to generate the retained intron neoantigen counts. In particular,

retained intron loads and retained intron neoantigen loads are shown for each of the Snyder (610) and Hugo (620) cohorts. A correlation may be calculated between the total number of retained introns and the RI neoantigen load. For example, the Hugo cohort has a R^2 of 0.93 and the Snyder cohort has a R^2 of 0.86. Thus, the total number of retained introns is tightly correlated with RI neoantigen load in both cohorts.

[0075] FIG. 7 is an illustrative graph of retained intron and somatic neoantigen counts for patients of the Snyder and Hugo patient study cohorts. The neoantigen identification methods described herein were applied to the Snyder and Hugo cohorts to identify neoantigens. FIG. 7 illustrates retained intron counts (710) and neoantigen counts (720) for each subject in the Snyder and Hugo cohorts. Across the Snyder and Hugo cohorts, the mean somatic neoantigen load is 2,218 and the mean RI neoantigen load is 1,515. This corresponds to the RI neoantigens contributing to an approximately 0.7 times increase to the total neoantigen load. A correlation may be calculated between the RI neoantigen load and the somatic neoantigen load. The correlation between somatic neoantigen load and RI neoantigen load of $p = 0.63$ is not significant, indicating that neoantigens from both sources merit consideration as independent features.

[0076] Putative somatic neoantigens may be identified *in silico* for each sample as described in Van Allen et al. 2015. Briefly, BAM files from each cohort may undergo sequencing quality control to ensure concordance between tumor and matched normal sequences and adequate depth of sequencing coverage. Single nucleotide variants may be called using MuTect (Cibulskis et al., 2013) and insertions and deletions may be called using Strelka (Saunders et al., 2012). In some variations, sequences of 9-10 amino acid peptides having at least one mutant amino acid may be identified. HLA-peptide binding interactions may be identified using NetMHCpan v3.0 (Nielsen et al., 2016), the identified peptides, and a set of HLA Class I alleles called with POLYSOLVER (Shukla et al., 2015). Somatic neoantigens may be identified by applying a threshold value based on the identified peptides and HLA Class I alleles. For example, all peptides with predicted binding rank $\leq 2.0\%$ for at least one patient HLA Class I allele may be identified as somatic neoantigens for each patient.

[0077] FIG. 8 is an illustrative graph of load counts and response statuses for subjects of the Snyder and Hugo patient study cohorts. The neoantigen identification methods described herein may be applied to each of the cohorts to identify neoantigen load. FIG. 8 depicts response status (e.g., subjects receiving clinical benefit (810) and receiving no clinical benefit (820)) for total retained intron load, neoantigen-yielding retained intron load, and retained intron neoantigen load.

[0078] In some variations of cohort analysis, RNA sequence data such as a raw RNA-Seq FASTQ file may be pseudoaligned to a reference transcriptome such as an augmented hg19 (GENCODE Release 19, GRCh37.p13) (Harrow et al., 2012) transcriptome index comprising both exonic and intronic transcript sequences. Transcript expression values may be quantified via kallisto (Bray et al., 2016). The KMA algorithm (Pimentel et al., 2016), implemented as a suite of Python scripts within an R package, may be used to identify whole exome sequence data such as the genomic loci of expressed intron retention events with limited false-positives. RI nucleotide sequences may be identified using these RI loci and the UCSC Table Browser database (Karolchik et al., 2004). In some of these variations, the identified RI nucleotide sequences may correspond to the intronic regions and fragments of the previous exonic sequences, as well as the open reading frame orientation at the start of the intron. In some variations, RI peptide sequences of 9-10 amino acids having at least one intronic amino acid may be identified by translating open reading frame orientations into intronic sequences until hitting an in-frame stop codon. HLA Class I alleles may be received using the POLYSOLVER algorithm (Shukla et al., 2015). Peptide-MHC I binding affinity may be received using NetMHCPan v3.1 (Nielsen et al., 2016). In some variations, RI neoantigens may be identified using the received binding affinity and HLA Class I alleles. In some of these variations, RI neoantigens may be identified by applying a threshold value based on the identified peptides and HLA Class I alleles. For example, a threshold of rank $< 0.5\%$ may be applied to identify putative RI neoantigens.

[0079] In some variations, false-positive RIs and RI neoantigens may be excluded by applying one or more filters (e.g., zero-coverage, PSI, expression). For example, after expression quantification, RIs expressed at a level ≤ 1 transcript per million may be excluded as they are likely

artefactual. In some variations, expression-based filters may be applied within the KMA algorithm. For example, RIs that do not reach a level of at least five unique counts and whose neighboring exons do not reach a level of at least one transcript per million in at least 25% of samples in a cohort may be excluded as false-positives. As another example, a panel-of-normals filter may be applied to RIs and RI neoantigens to exclude false-positives.

C. Experimental confirmation of predicted RI neoantigens in cancer cell lines

[0080] RI neoantigens may be identified in tumor cell lines that are complexed to MHC I to experimentally demonstrate that RIs are endogenously processed and presented through the MHC class I pathway. RNA-Seq data from multiple human tumor cell lines and their corresponding MHC I mass spectrometry data (including immunopeptidomes) may be used to query for *in vitro* presentation of the RI neoantigens (Barretina et al., 2012; Ritz et al., 2016). As shown in FIG. 9, in the melanoma cell line MeWo (902) may be used to generate CCLE RNA-Seq data (904). The RI neoantigens *EVYAAGKYV* (920) and *YAAGKYVSF* (922) may be identified using the methods (910) described herein with the CCLE data (904), HLA class I alleles (906), and immunopeptidomes (908). For example, both of these RI neoantigens (920, 922) are predicted to arise from a retained intron (940) in the gene *KCNAB2* at genomic locus chr1:6142308-6145287 (930). The identified RI neoantigens may be experimentally confirmed in complex with MHC I via mass spectrometry.

[0081] Similar analysis (1010) may be performed for a plurality of melanoma cell lines (1002, 1004, 1006, 1008, 1009), as shown in FIG. 10. CCLE RNA-Seq data (1012) may be derived from each of the cell lines (1002, 1004, 1006, 1008, 1009). RI neoantigens (1020, 1022, 1024, 1026, 1028) may be identified using the methods (1014) described herein with the CCLE data (1012), HLA class I alleles (1016), and immunopeptidomes (1018). These identified RI neoantigens may be experimentally confirmed in complex with MHC I via mass spectrometry. SK-MEL-5 (1002) includes RI neoantigens *AMSDVSHPK* and *LAMSDVSHPK* (1020) from an intron in gene *SMARCD1*. B cell lymphoma cell lines CA46 (1004) includes RI neoantigen *FRYVAQAGL* (1022) from an intron in gene *LRSAM1*. DOHH-2 (1006) includes RI neoantigens *TLFLLSLPL* and *FLLSLPLPV* (1024) from an intron in gene *CYB561A3*. Leukemia cell line HL-60 (1008) includes

RI neoantigen *SVLDDVRGW* (1026) from an intron in gene *TAF1*. THP-1 (1009) includes RI neoantigen *LTSQGKSAF* (1028) from an intron in gene *ZCCHC6*). In some variations, a threshold of predicted binding rank $\leq 2.0\%$ for at least one HLA Class I allele may be used to distinguish cell line RI neoantigens. All pipeline filters applied to patient data described above were implemented on the cell line data except RI neoantigens expected to be retained in normal tissue since these experiments were focused on presentation of RI neoantigens rather than immune system stimulation once presented.

[0082] The foregoing description, for purposes of explanation, used specific nomenclature to provide a thorough understanding of the invention. However, it will be apparent to one skilled in the art that specific details are not required in order to practice the invention. Thus, the foregoing descriptions of specific variations of the invention are presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise forms disclosed; obviously, many modifications and variations are possible in view of the above teachings. The variations were chosen and described in order to best explain the principles of the invention and its practical applications, and they thereby enable others skilled in the art to best utilize the invention and various implementations with various modifications as are suited to the particular use contemplated. It is intended that the following claims and their equivalents define the scope of the invention.

CLAIMS

We claim:

1. A method comprising:

receiving RNA sequence data sequenced from a subject's tumor and aligned or pseudoaligned to a reference transcriptome or genome comprising exonic and intronic sequences, corresponding retained intron (RI) transcript expression values, whole exome sequence data sequenced from the subject's tumor, and corresponding HLA class I alleles;

identifying RIs using the RI expression values and the RNA sequence data,

identifying RI nucleotide sequences using the RIs and a reference genomic database;

translating the RI nucleotide sequences into RI peptides;

receiving binding affinity data between the RI peptides and the HLA class I alleles; and

identifying RI neoantigens using the binding affinity data and the RI peptides.

2. The method of claim 1, further comprising formulating a subject-specific immunogenic composition for administration to the subject using one or more of the RI neoantigens.

3. The method of claim 1, further comprising identifying genomic coordinates and window sizes corresponding to the RIs, RI nucleotide sequences, and RI peptides.

4. The method of claim 3, further comprising identifying open reading frame orientations using the genomic coordinates and the reference genomic database.

5. The method of claim 4, wherein identifying the RI nucleotide sequences comprises using the genomic coordinates, the window sizes, and the open reading frame orientations.

6. The method of any of claims 3, 4, or 5, wherein the genomic coordinates correspond to mutually exclusive intron chromosomal coordinates and comprise an intron start location, and the window size corresponds to a number of amino acids around the intron start location.

7. The method of claim 1, wherein the RIs are expressed at a level of at least about one transcript per million.

8. The method of claim 1, wherein the RI neoantigens comprise the binding affinity of rank less than about 2 percentile or less than about 500 nanomolar.

9. The method of claim 1, wherein the RI neoantigens comprise the binding affinity of rank less than about 0.5 percentile or less than about 50 nanomolar.

10. The method of claim 1, wherein the RI nucleotide sequences comprise at least one intronic amino acid.

11. The method of claim 1, wherein identifying the RI neoantigens comprises excluding false-positive retained introns or false-positive RI neoantigens.

12. The method of claim 11, wherein excluding false-positive retained introns comprises applying a zero-coverage filter.

13. The method of claim 11, wherein excluding false-positive retained introns comprises applying a percent-spliced-in (PSI) filter.

14. The method of claim 11, wherein excluding false-positive retained introns comprises applying an expression filter.

15. The method of claim 11, wherein excluding false-positive retained introns comprises applying a filter removing introns that are typically retained in normal tissue.

16. The method of claim 11, wherein excluding false-positive RI neoantigens comprises applying a filter removing RI neoantigens with peptide sequences that are present in a normal proteome.

17. The method of claim 1, further comprising receiving raw RNA sequencing data, wherein aligned or pseudoaligned transcripts comprise aligned raw RNA sequences.

18. The method of claim 1, further comprising generating transcript or summary information of the RI neoantigens and distribution information of HLA allele types.

19. A method comprising:

receiving a set of RNA sequences from a sample of a subject's tumor;

quantifying an expression level for each transcript of the set of RNA sequences as transcript expression data;

identifying retained introns from the transcript expression data by excluding exonic sequences or wherein the expression level of the transcript expression data is below a predetermined level;

generating a set of retained intron nucleotide sequences by referencing retained intron loci in a reference genomic database; and

translating the set of retained intron sequences into a set of retained intron peptides.

20. The method of claim 19, further comprising:

determining a set of binding affinities for the set of retained intron peptides using a set of HLA class I alleles; and

selecting neoantigens from the set of retained peptides using a pre-determined binding affinity threshold value.

21. A system for characterizing a subject's genome comprising:

a transceiver configured to receive RNA sequences sequenced from a subject's tumor and aligned or pseudoaligned to a reference transcriptome comprising exonic and intronic sequences, corresponding retained intron (RI) transcript expression values, whole exome sequence data sequenced from the subject's tumor, corresponding HLA class I alleles, and binding affinities between RI peptides and the HLA class I alleles; and

a controller comprising a processor and a memory, the controller configured to:

identify RIs using the RI transcript expression values and an expression value threshold;

identify RI nucleotide sequences using RI chromosomal loci and a reference genomic database;

translate the RI nucleotide sequences into the RI peptides; and

identify RI neoantigens using the binding affinities and the RI peptides.

22. The system of claim 21, further comprising one or more of an RNA sequencing system configured to generate raw RNA sequence data, an alignment or pseudoalignment system configured to align the raw RNA sequence data to the reference transcriptome or genome comprising the exonic and intronic sequences, a quantification system configured to generate the RI expression values, a DNA sequencing system configured to generate the whole exome sequence data, an HLA typing system configured to generate the HLA class I alleles, and peptide binding system configured to generate the binding affinities, each coupled to the transceiver.

23. The system of claim 21, wherein the controller is further configured to identify genomic coordinates and window sizes corresponding to the retained introns, RI nucleotide sequences, and RI peptides.

24. The system of claim 21, wherein the controller is further configured to identify open reading frame orientations using intron genomic coordinates and the reference genomic database.

25. The system of claim 21, wherein identifying the retained introns and RI neoantigens comprises excluding false-positive retained introns.

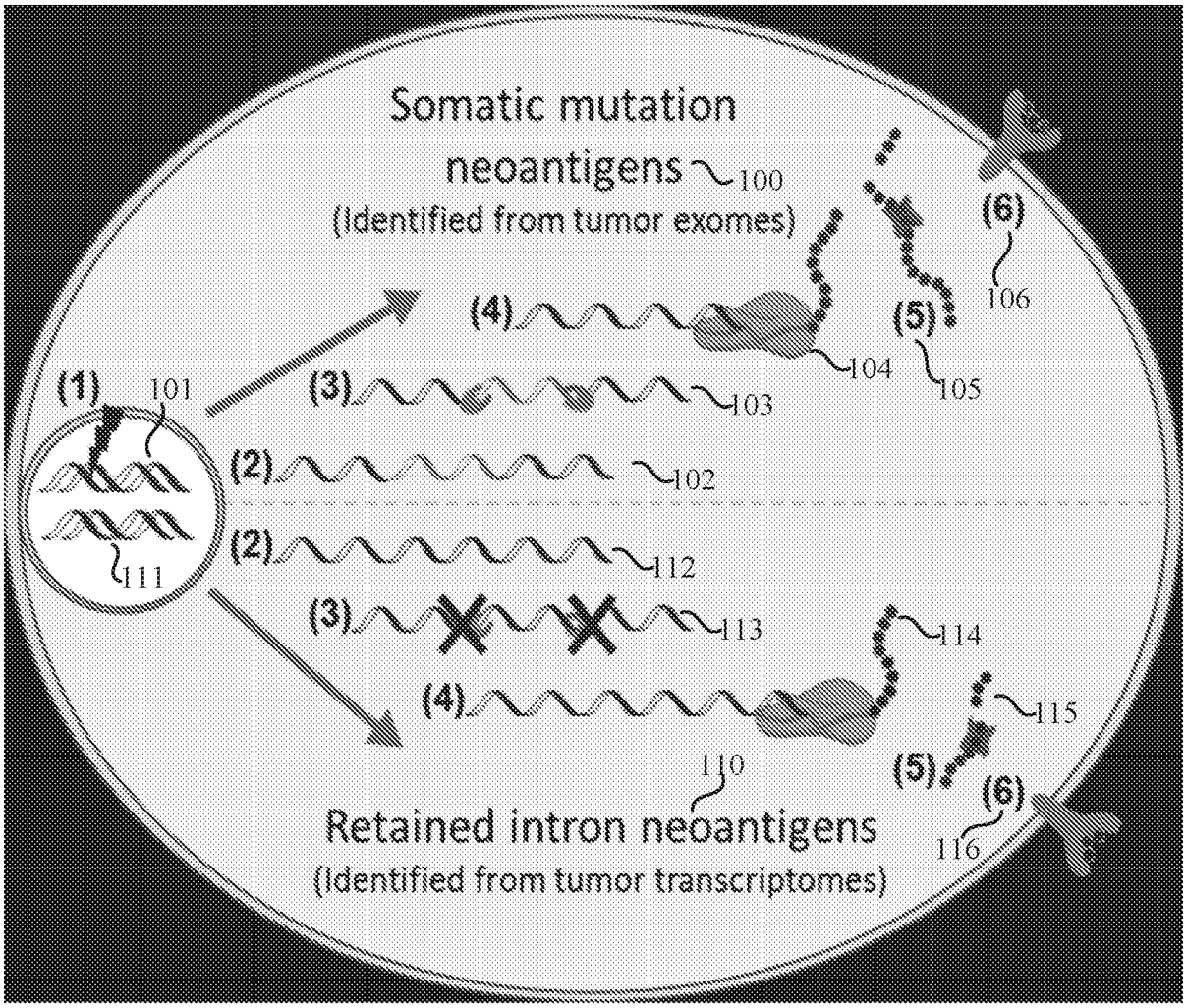
Somatic mutation
neoantigens
100
(Identified from tumor exomes)
(1)
101
111
(2)
102
(3)
103
(4)
104
(5)
105
(6)
106
(2)
112
(3)
113
114
(4)
115
110
(5)
(6)
116
Retained intron neoantigens
(Identified from tumor transcriptomes)

FIG. 1

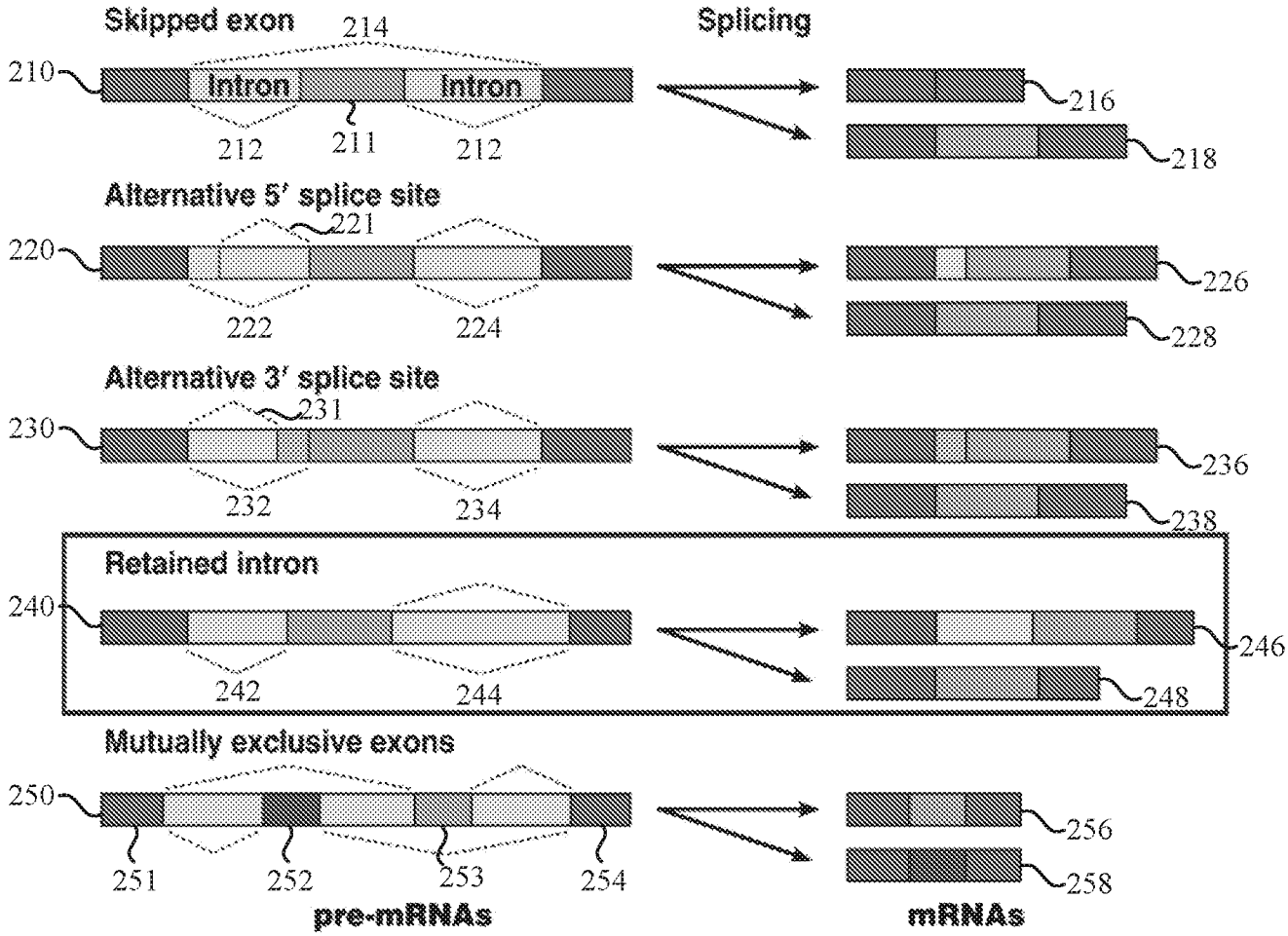
Skipped exon
214
Splicing
210
Intron
Intron
212
211
212
216
218
Alternative 5′ splice site
221
220
222
224
226
228
Alternative 3′ splice site
231
230
232
234
236
238
Retained intron
240
242
244
246
248
Mutually exclusive exons
250
251
252
253
254
256
258
pre-mRNAs
mRNAs

FIG. 2

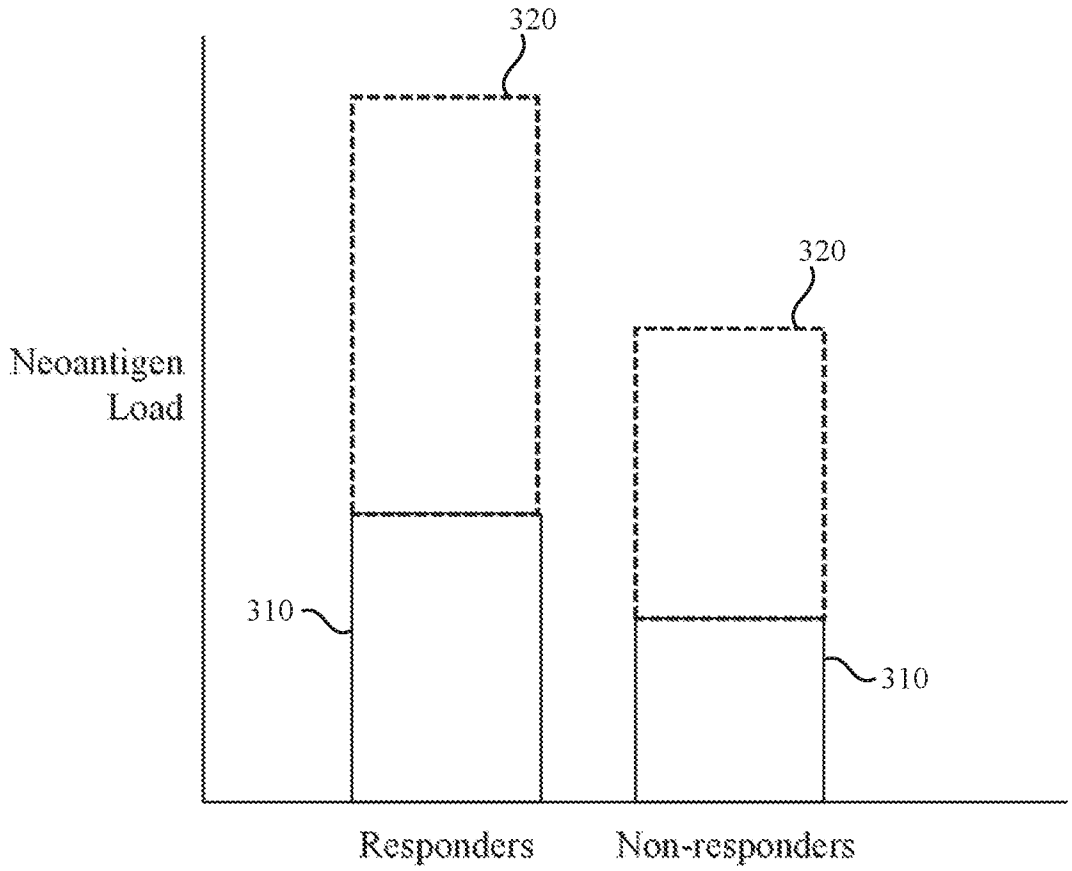

FIG. 3

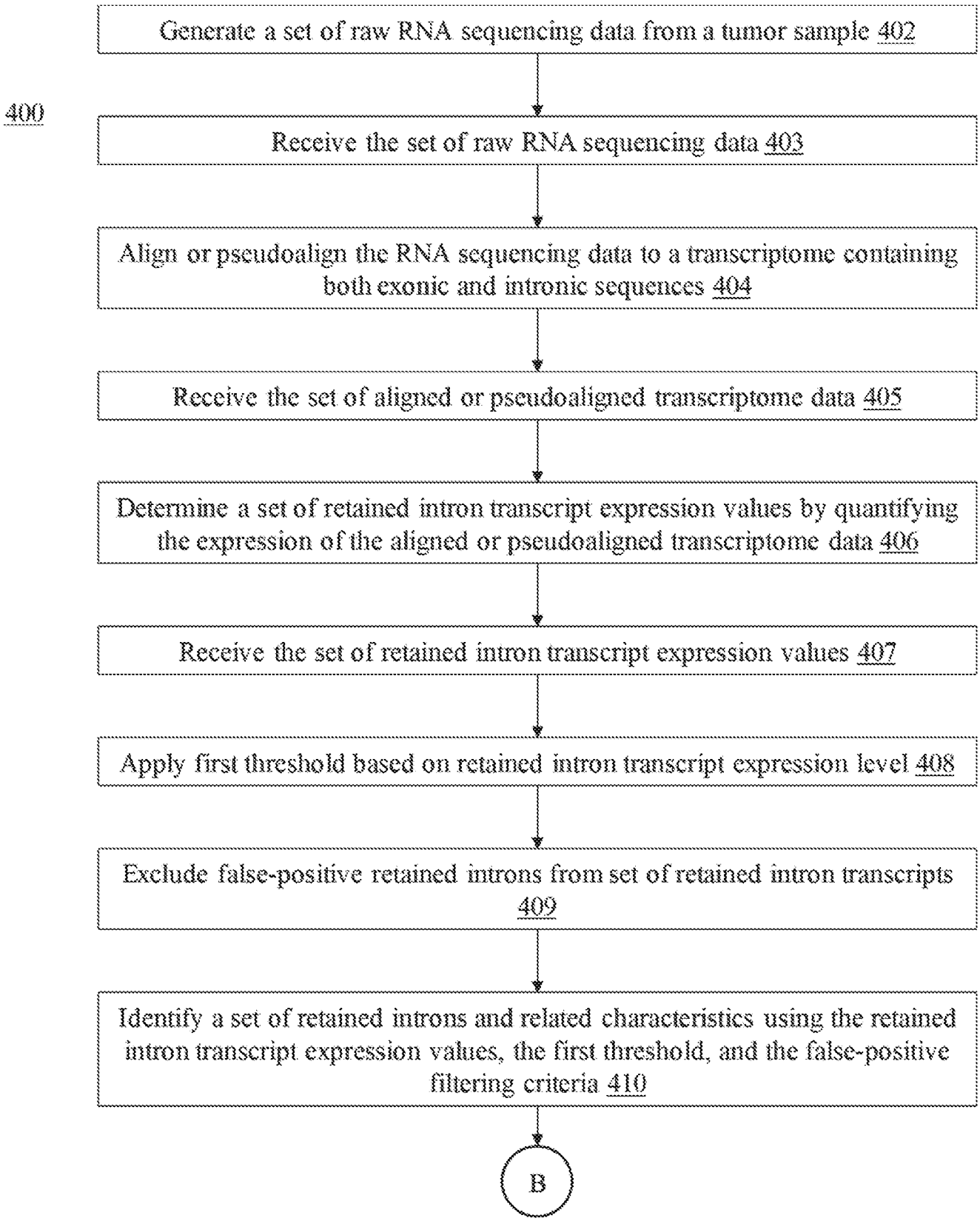
400
Generate a set of raw RNA sequencing data from a tumor sample 402
Receive the set of raw RNA sequencing data 403
Align or pseudoalign the RNA sequencing data to a transcriptome containing both exonic and intronic sequences 404
Receive the set of aligned or pseudoaligned transcriptome data 405
Determine a set of retained intron transcript expression values by quantifying the expression of the aligned or pseudoaligned transcriptome data 406
Receive the set of retained intron transcript expression values 407
Apply first threshold based on retained intron transcript expression level 408
Exclude false-positive retained introns from set of retained intron transcripts 409
Identify a set of retained introns and related characteristics using the retained intron transcript expression values, the first threshold, and the false-positive filtering criteria 410
B

FIG. 4A

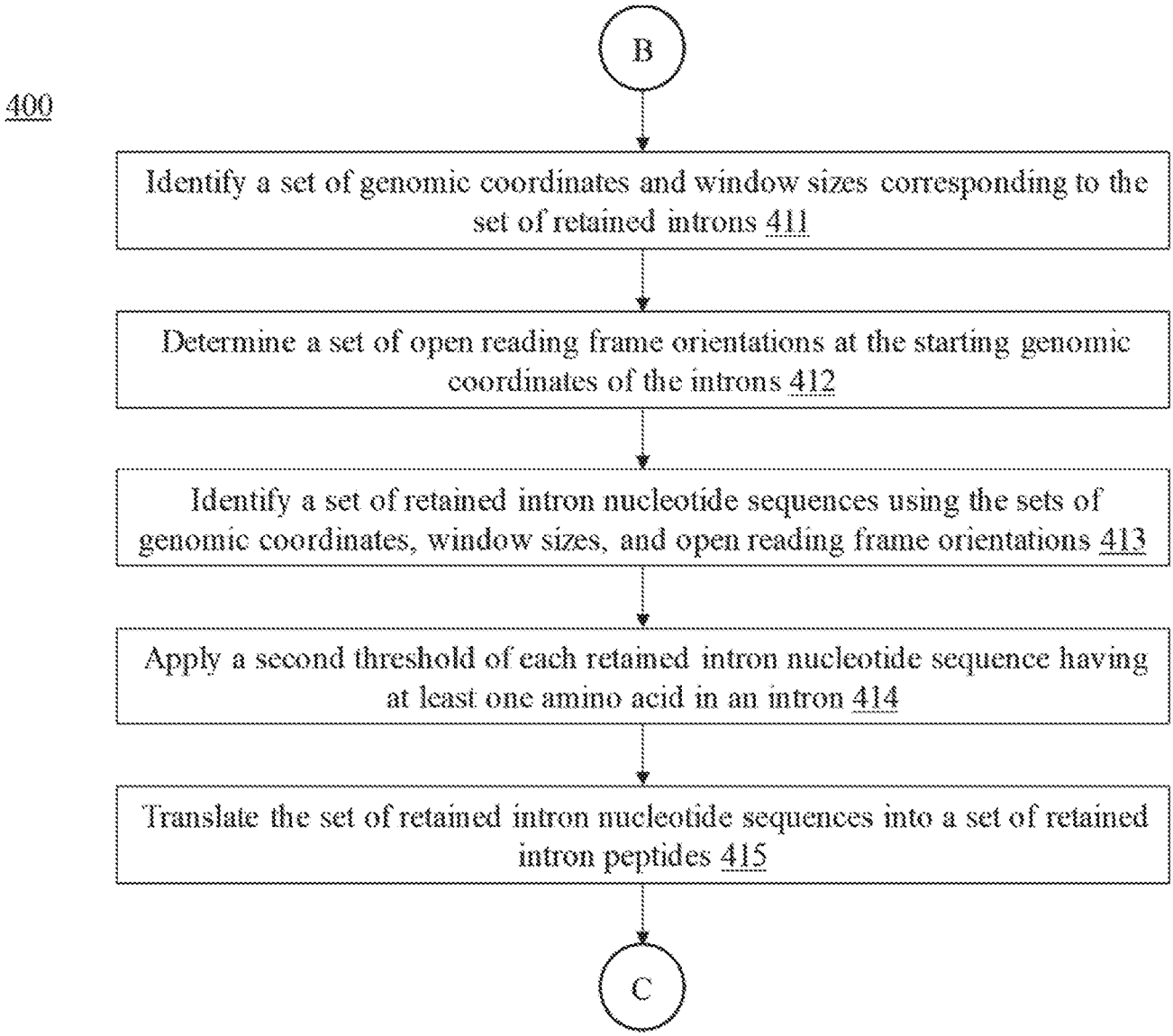
400
B
Identify a set of genomic coordinates and window sizes corresponding to the set of retained introns 411
Determine a set of open reading frame orientations at the starting genomic coordinates of the introns 412
Identify a set of retained intron nucleotide sequences using the sets of genomic coordinates, window sizes, and open reading frame orientations 413
Apply a second threshold of each retained intron nucleotide sequence having at least one amino acid in an intron 414
Translate the set of retained intron nucleotide sequences into a set of retained intron peptides 415
C

FIG. 4B

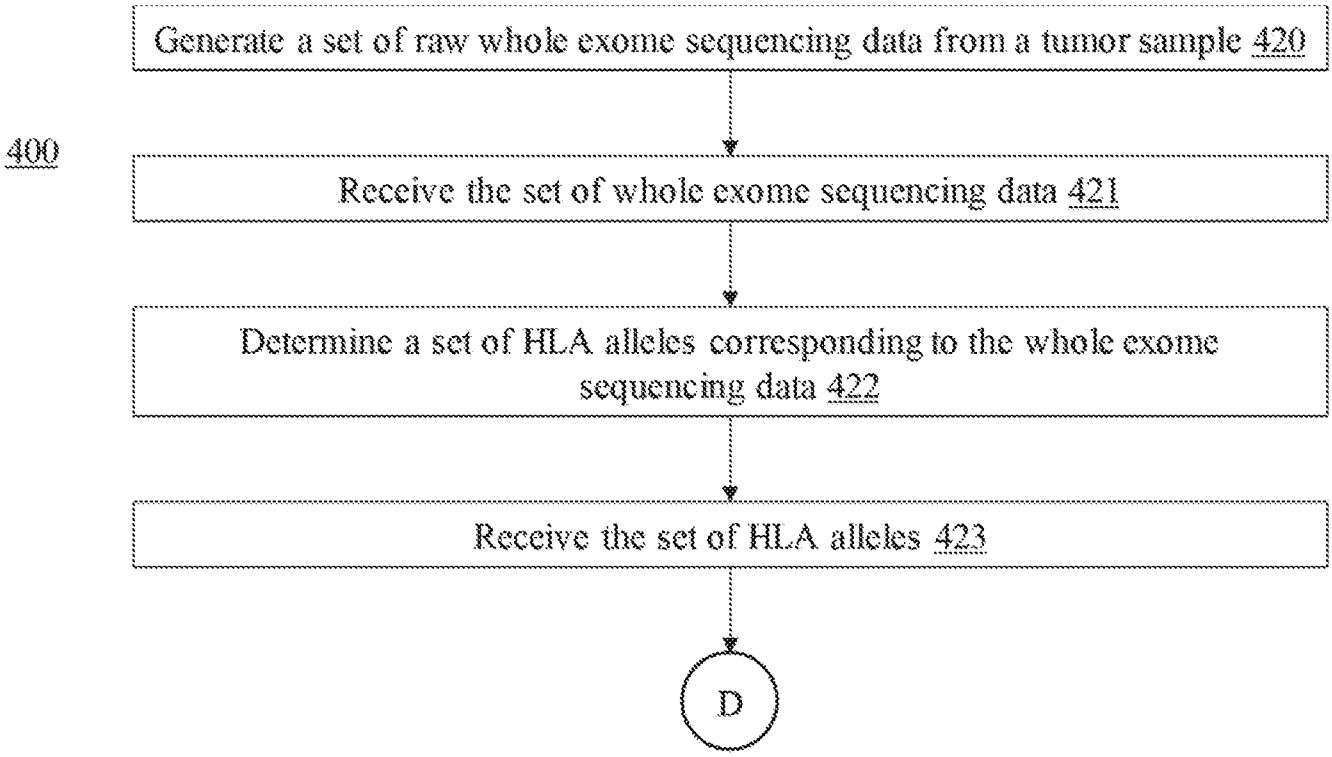
400
Generate a set of raw whole exome sequencing data from a tumor sample 420
Receive the set of whole exome sequencing data 421
Determine a set of HLA alleles corresponding to the whole exome sequencing data 422
Receive the set of HLA alleles 423
D

FIG. 4C

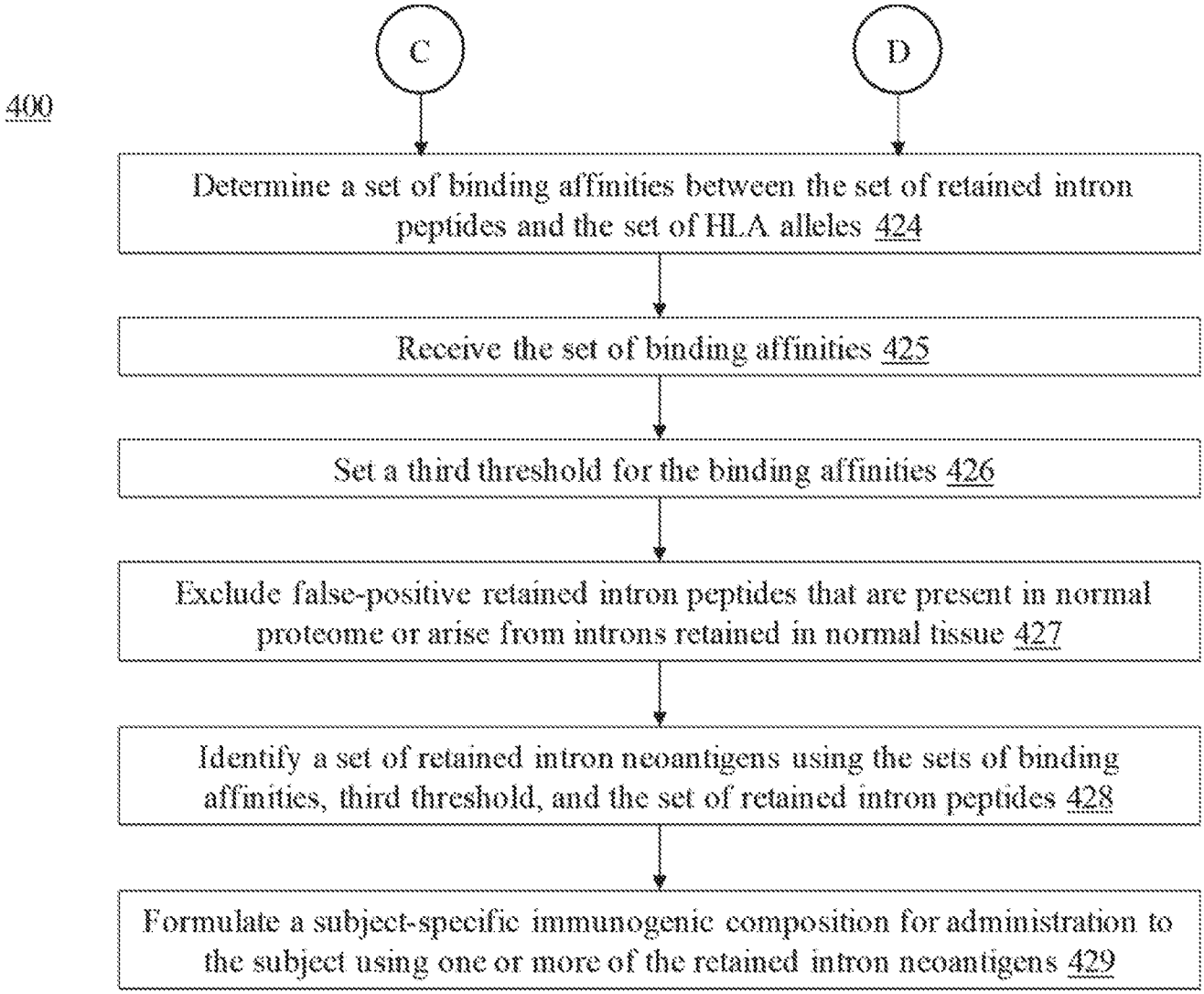
400
C
D
Determine a set of binding affinities between the set of retained intron peptides and the set of HLA alleles 424
Receive the set of binding affinities 425
Set a third threshold for the binding affinities 426
Exclude false-positive retained intron peptides that are present in normal proteome or arise from introns retained in normal tissue 427
Identify a set of retained intron neoantigens using the sets of binding affinities, third threshold, and the set of retained intron peptides 428
Formulate a subject-specific immunogenic composition for administration to the subject using one or more of the retained intron neoantigens 429

FIG. 4D

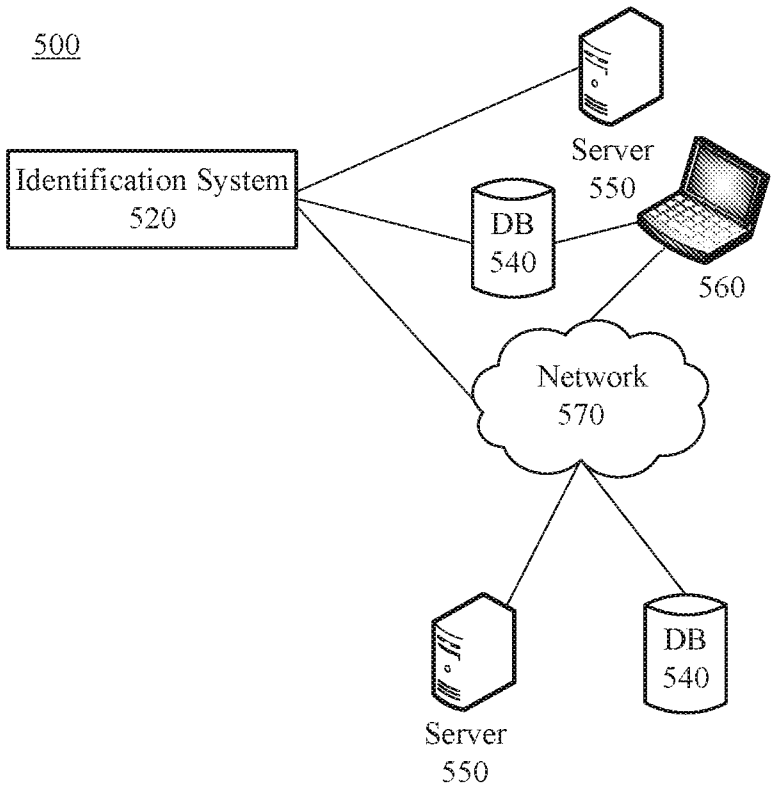
500
Identification System
520
Server
550
DB
540
560
Network
570
Server
550
DB
540

FIG. 5A

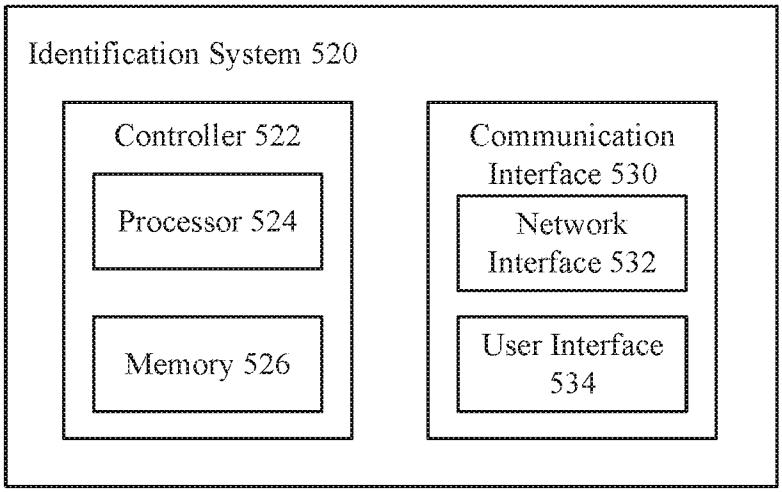
Identification System 520
Controller 522
Processor 524
Memory 526
Communication Interface 530
Network Interface 532
User Interface 534

FIG. 5B

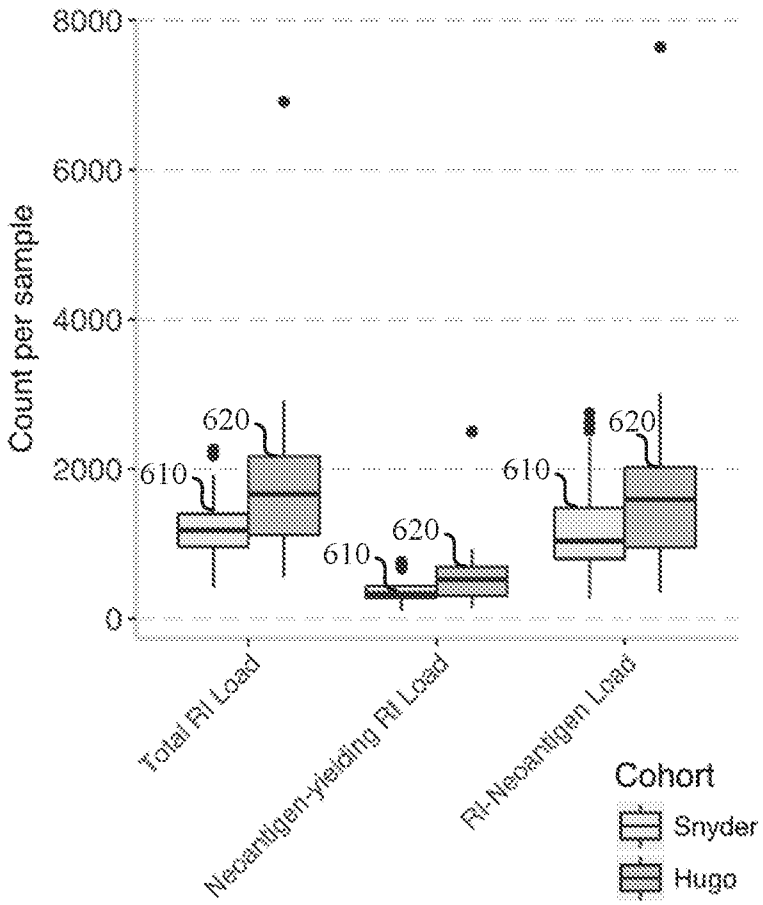
8000
6000
4000
2000
0
Count per sample
610
620
610
620
610
620
Total RI Load
Neoantigen-yielding RI Load
RI-Neoantigen Load
Cohort
Snyder
Hugo

FIG. 6

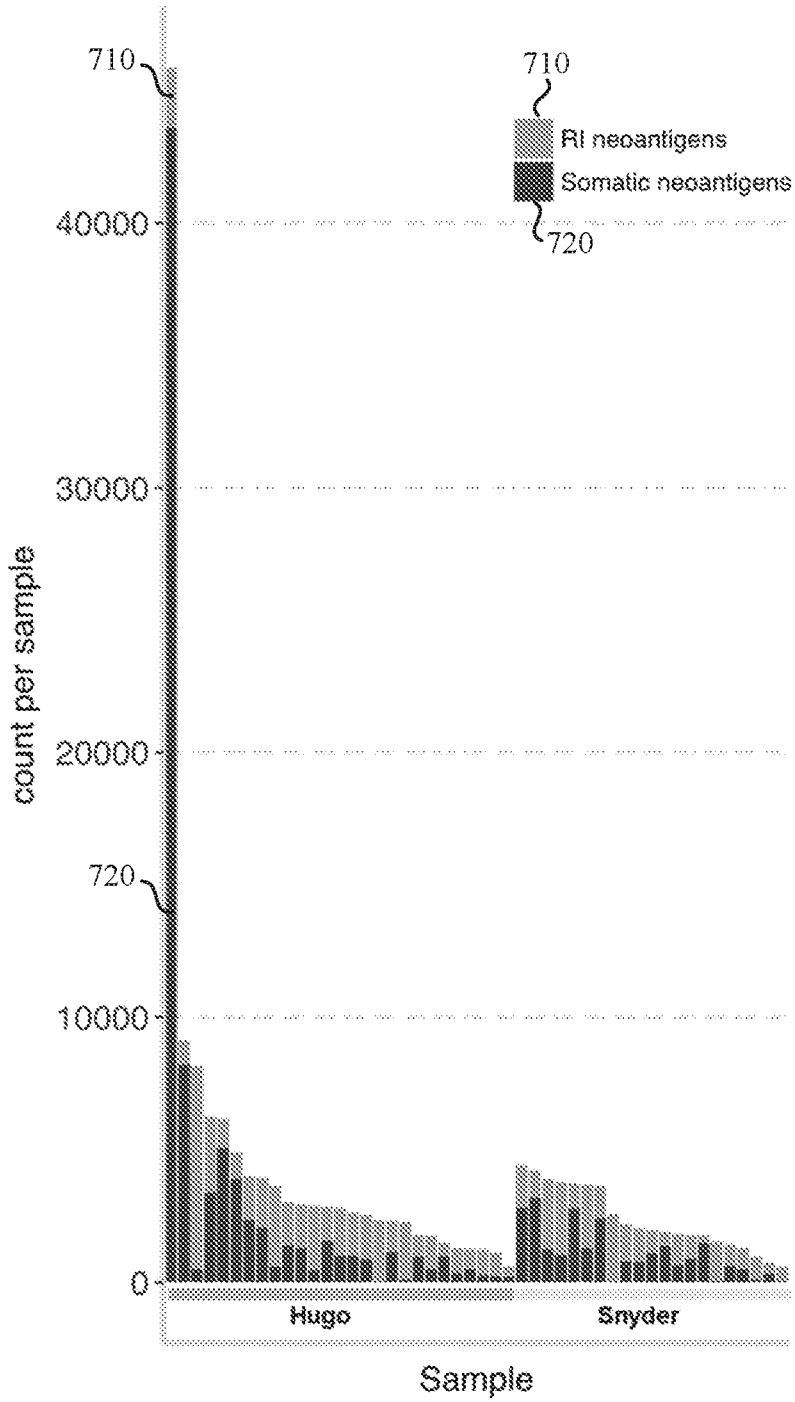
710
710
RI neoantigens
Somatic neoantigens
720
40000
30000
20000
720
10000
0
count per sample
Hugo
Snyder
Sample

FIG. 7

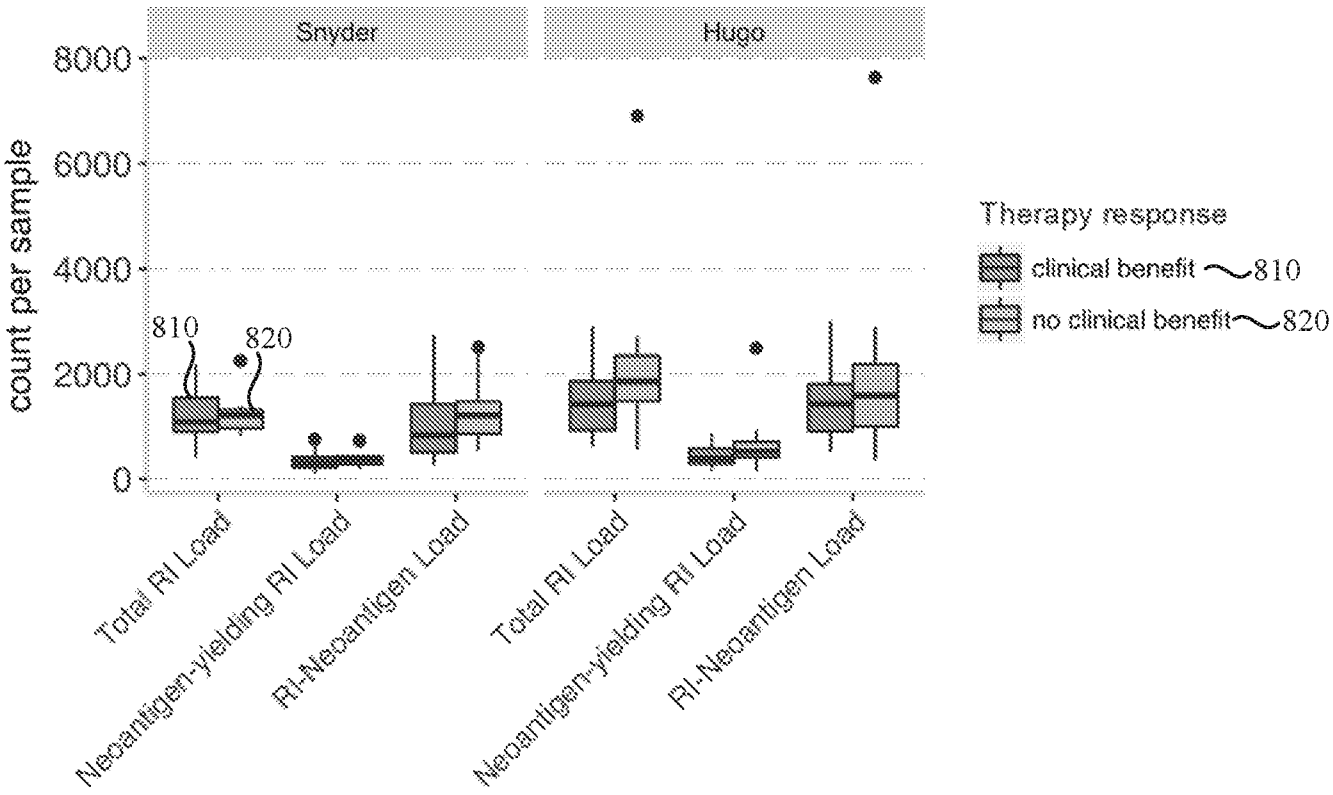
Snyder
Hugo
8000
6000
4000
2000
0
count per sample
810
820
Total RI Load
Neoantigen-yielding RI Load
RI-Neoantigen Load
Total RI Load
Neoantigen-yielding RI Load
RI-Neoantigen Load
Therapy response
clinical benefit 810
no clinical benefit 820

FIG. 8

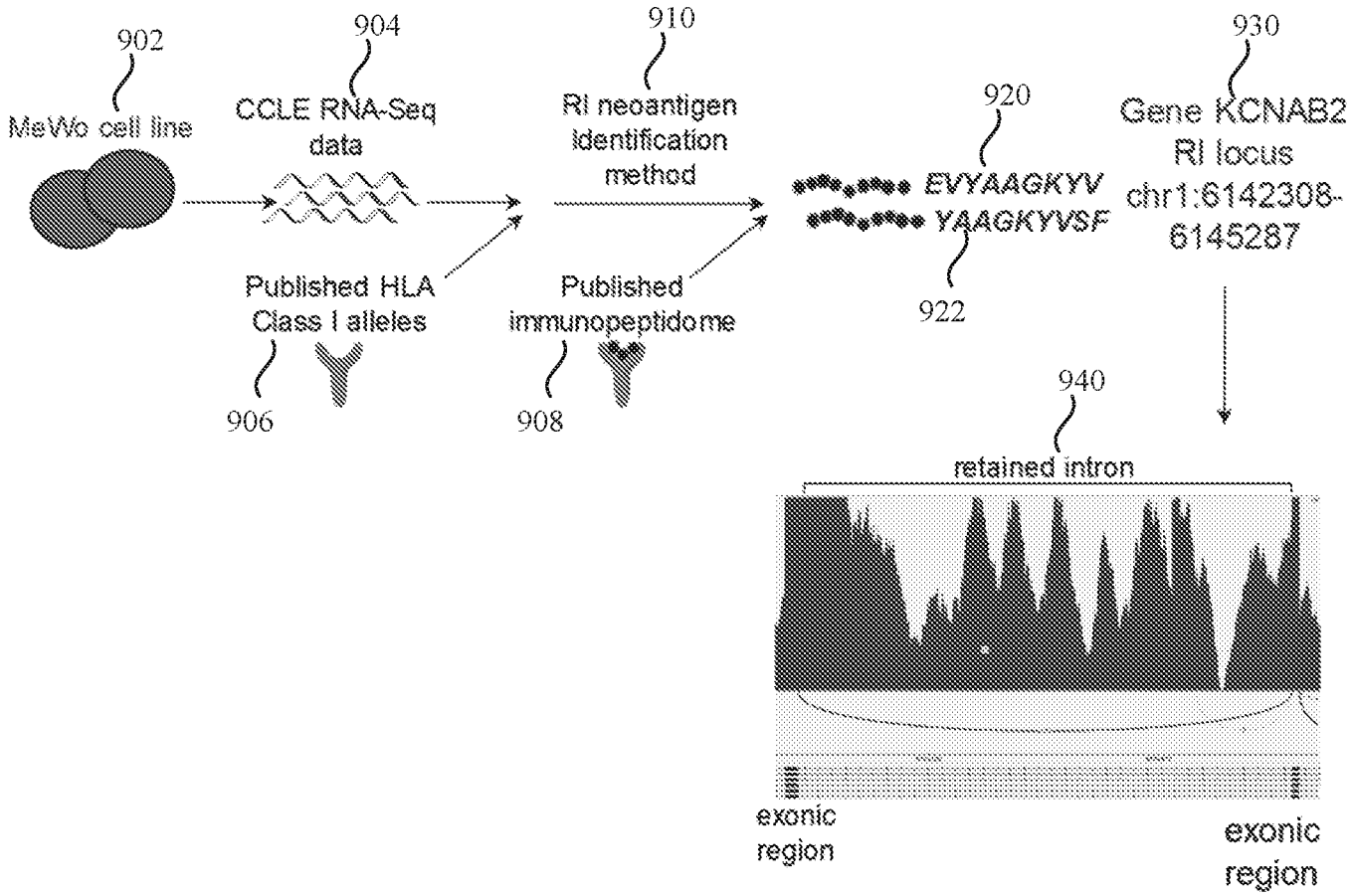
902
MeWo cell line
904
CCLE RNA-Seq data
910
RI neoantigen identification method
920
EVYAAGKYV
YAAGKYVSF
922
930
Gene KCNAB2
RI locus
chr1:6142308-6145287
Published HLA Class I alleles
906
Published immunopeptidome
908
940
retained intron
exonic region
exonic region

FIG. 9

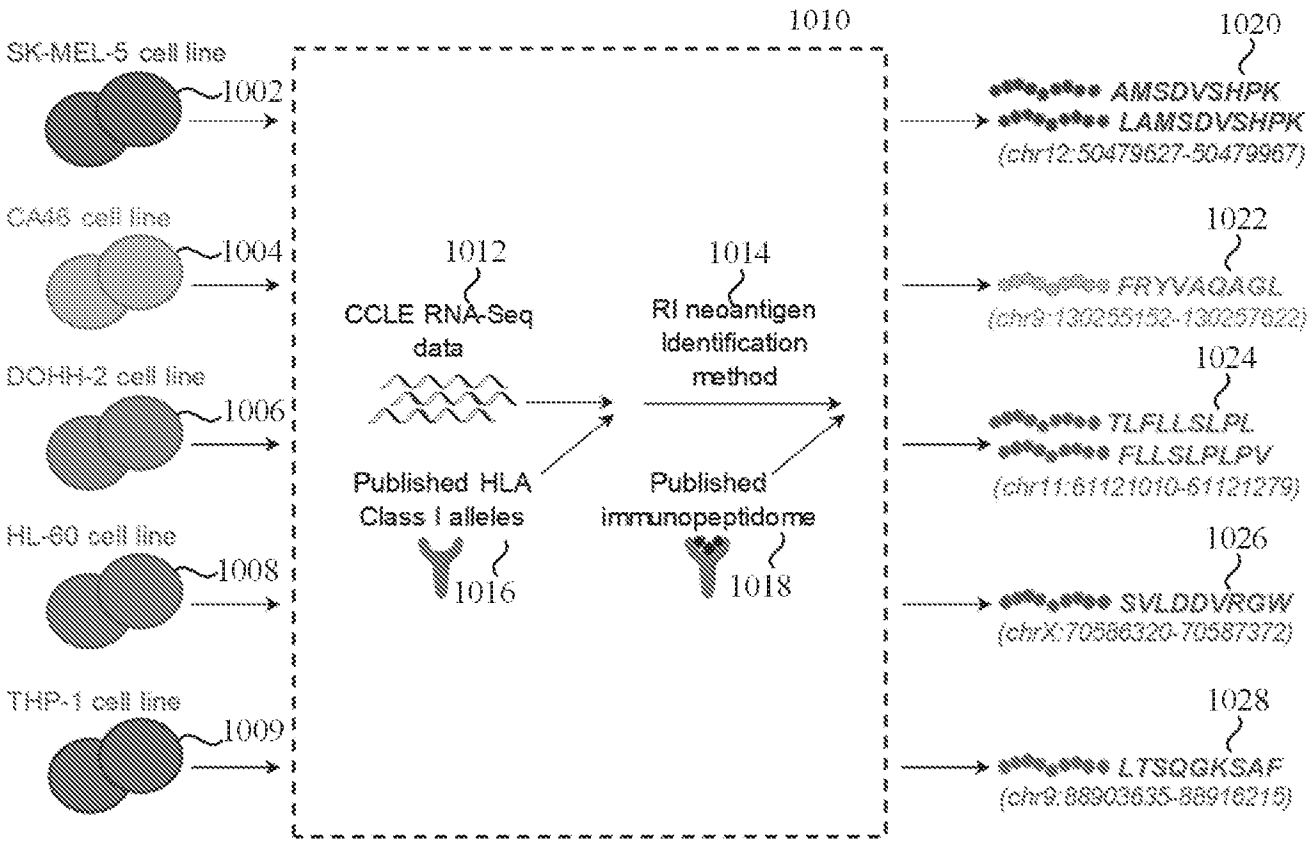
SK-MEL-5 cell line
1002
CA46 cell line
1004
DOHH-2 cell line
1006
HL-60 cell line
1008
THP-1 cell line
1009
1010
1012
CCLE RNA-Seq data
1014
RI neoantigen identification method
Published HLA Class I alleles
1016
Published immunopeptidome
1018
1020
AMSDVSHPK
LAMSDVSHPK
(chr12:50479627-50479967)
1022
FRYVAQAGL
(chr9:130255152-130257622)
1024
TLFLLSLPL
FLLSLPLPV
(chr11:61121010-61121279)
1026
SVLDDVRGW
(chrX:70586320-70587372)
1028
LTSQGKSAF
(chr9:88903635-88916215)

FIG. 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/024905

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6809 C12Q1/6881 C12Q1/6886 G01N33/52
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, EMBASE, WPI Data, BIOSIS, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BECKY SIMON: "Intron retention tied to checkpoint inhibitor responders", BC EXTRA, , 7 March 2017 (2017-03-07), XP009506637, Retrieved from the Internet: URL:https://www.biocentury.com/bc-extra/clinical-news/2017-03-07/intron-retention-tied-checkpoint-inhibitor-responders [retrieved on 2018-07-09] abstract -----	1-25
X	WO 2011/143656 A2 (GEN HOSPITAL CORP [US]; DANA FARBER CANCER INST INC [US]; HACOHEN NIR) 17 November 2011 (2011-11-17) paragraph [0070] claim 1; figure 3 -----	1-25
	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
12 July 2018	26/07/2018

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Reuter, Uwe

2

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/024905

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	M. RAJASAGI ET AL: "Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia", BLOOD, vol. 124, no. 3, 2 June 2014 (2014-06-02), pages 453-462, XP055322841, US ISSN: 0006-4971, DOI: 10.1182/blood-2014-04-567933 page 455; figures 1,2c abstract	1-25
X	WO 2014/168874 A2 (BROAD INST INC [US]; DANA FARBER CANCER INST INC [US]; GEN HOSPITAL CO) 16 October 2014 (2014-10-16) claim 2; figures 1,11c	1-25

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2018/024905

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2011143656 A2	17-11-2011	AU 2011252795 A1	08-11-2012
		CA 2797868 A1	17-11-2011
		CN 103180730 A	26-06-2013
		CN 105648056 A	08-06-2016
		EP 2569633 A2	20-03-2013
		EP 3023788 A1	25-05-2016
		ES 2564841 T3	29-03-2016
		JP 5948319 B2	06-07-2016
		JP 2013530943 A	01-08-2013
		JP 2016156828 A	01-09-2016
		KR 20130119845 A	01-11-2013
		US 2011293637 A1	01-12-2011
		US 2016008447 A1	14-01-2016
		US 2016331822 A1	17-11-2016
		US 2018055922 A1	01-03-2018
		WO 2011143656 A2	17-11-2011
WO 2014168874 A2	16-10-2014	AU 2014251207 A1	05-11-2015
		BR 112015025460 A2	10-10-2017
		CA 2908434 A1	16-10-2014
		CN 105377292 A	02-03-2016
		EP 2983702 A2	17-02-2016
		JP 2016518355 A	23-06-2016
		KR 20150143597 A	23-12-2015
		US 2016101170 A1	14-04-2016
		WO 2014168874 A2	16-10-2014