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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 October 2016 (06.10.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/161273 A1

(51) International Patent Classification:
A61K 35/17 (2015.01) *C12Q 1/68* (2006.01)
G01N 33/00 (2006.01) *C12N 5/16* (2006.01)

(21) International Application Number:
PCT/US2016/025535

(22) International Filing Date:
1 April 2016 (01.04.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/141,701 1 April 2015 (01.04.2015) US

(71) Applicants: **ADAPTIVE BIOTECHNOLOGIES CORP.** [US/US]; 1551 Eastlake Avenue East, Suite 200, Seattle, Washington 98102 (US). **FRED HUTCHINSON CANCER RESEARCH CENTER** [US/US]; 1100 Fairview Avenue North, Mail Stop J2-110, Seattle, Washington 98109 (US).

(72) Inventors: **ROBINS, Harlan S.**; c/o Adaptive Biotechnologies Corp., 1551 Eastlake Avenue East, Suite 200, Seattle, Washington 98102 (US). **CHAPUIS, Aude Georgiana**; c/o Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Mail Stop J2-110, Seattle, Washington 98109 (US). **SCHMITT, Thomas M.**; c/o Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Mail Stop J2-110, Seattle, Washington 98109 (US). **GREENBERG, Philip**; c/o Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Mail Stop J2-110, Seattle, Washington 98109 (US). **SHERWOOD, Anna**; c/o Adaptive Biotechnologies Corp., 1551 Eastlake Ave E., Suite 200, Seattle, Washington 98102 (US).

(74) Agents: **FULLER, Alyson C.** et al.; Cooley LLP, 1299 Pennsylvania Avenue, NW, 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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, 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))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

WO 2016/161273 A1

(54) Title: METHOD OF IDENTIFYING HUMAN COMPATIBLE T CELL RECEPTORS SPECIFIC FOR AN ANTIGENIC TARGET

(57) Abstract: Methods are provided for identifying T cell receptors that specifically bind a particular antigenic target and can be used as therapeutics against disease.

METHOD OF IDENTIFYING HUMAN COMPATIBLE T CELL RECEPTORS SPECIFIC FOR AN ANTIGENIC TARGET

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application No. 62/141,701 filed April 1, 2015, the contents of which are hereby incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] Recent efforts have established modified T cells as an effective therapeutic against cancer cells. Modified T cells can be engineered to target particular antigens expressed on cancer cells. In particular, genes encoding T cell receptor (TCR) chains that bind specific HLA presented antigens have been inserted into patient or donor-derived T cells (usually CD8 T cells) by a vector to create clones that can kill cells expressing those antigens. T cells are extracted from a cancer patient, the modified TCR genes are inserted into the T cells, and the engineered T cells are then put back into the patient (the T cell might or might not be selected for certain subtypes). The antigenic targets are those that are expressed by certain cancer cell types and not expressed at high levels on other human cells. A classic example of an antigenic target is a peptide from testis specific NY-ESO-1, which is often expressed in a wide range of cancers, such as ovarian carcinoma or melanoma. Other groups have attempted to identify T cells that bind particular peptide epitopes using binding assays and TCR analysis. *See* Hunsucker et al. Cancer Immunol Res; 3(3) March 2015. However, it can be difficult to determine which TCRs bind a particular antigenic target of interest with optimal binding characteristics.

[0003] There is a need for effective and accurate methods of identifying individual T cell receptors that target and bind to particular antigens of interest with optimal binding characteristics in a high throughput and efficient manner.

SUMMARY OF THE INVENTION

[0004] The invention includes methods for identifying an antigen-specific T cell receptor as a therapeutic, comprising steps for dividing one or more samples comprising T cells into a first subset and a second subset, performing immunosequencing of rearranged nucleic acid molecules extracted from the first subset of cells to obtain a first plurality of unique sequence reads, and determining based on the first plurality of unique sequence reads a relative abundance of each unique T cell receptor (TCR) sequence out of a total number of T cells in the first subset. The

method also includes steps for enriching the second subset of cells with multimer molecules comprising an HLA-presented antigen to identify a population of antigen-specific T cells that bind the multimer, performing immunosequencing of rearranged nucleic acid molecules extracted from said sorted population of antigen-specific T cells to obtain a second plurality of unique sequence reads, determining based on the second plurality of unique sequence reads a relative abundance of each unique TCR sequence out of a total number of T cells in the second subset, and determining a relative change in abundance of each antigen-specific TCR sequence based on the relative abundances of the antigen-specific TCR sequence in the first subset and the second subset. The method also includes identifying at least one antigen-specific TCR sequence as a clone for therapeutic use based on its determined relative change in abundance, and identifying a second TCR sequence that pairs with the at least one antigen-specific TCR sequence to form a TCR cognate pair.

[0005] In one embodiment, the enriching is performed by flow cytometry. In another embodiment, the one or more samples are blood samples. In certain embodiments, the one or more samples are tissue samples.

[0006] In some embodiments, identifying at least one antigen-specific TCR sequence as a clone for therapeutic use based on its determined relative change in abundance comprises ranking each of the antigen-specific TCR sequences based on its binding affinity for said HLA-presented antigen.

[0007] In another embodiment, the TCR sequence is a TCRB, TCRA, TCRG, or TCRD sequence.

[0008] The invention includes a method of identifying an antigen-specific T cell as a therapeutic, comprising receiving one or more samples comprising T cells, dividing the one or more samples into a first subset and a second subset, performing immunosequencing of rearranged nucleic acid molecules extracted from the first subset of cells to obtain a first plurality of unique sequence reads, determining based on the first plurality of unique sequence reads a relative abundance of each unique T cell receptor (TCR) sequence out of a total number of T cells in the first subset, and enriching the second subset of cells with multimer molecules comprising an HLA-presented antigen to identify a population of antigen-specific T cells that bind the multimer.

[0009] The method also includes performing immunosequencing of rearranged nucleic acid molecules extracted from said sorted population of antigen-specific T cells to obtain a second

plurality of unique sequence reads, determining based on the second plurality of unique sequence reads a relative abundance of each unique TCR sequence out of a total number of T cells in the second subset, determining a relative change in abundance of each antigen-specific TCR sequence based on the relative abundances of the antigen-specific TCR sequence in the first subset and the second subset, and identifying an antigen-specific TCR sequence based on its relative change in abundance as a clone for therapeutic use for said HLA-presented antigen.

[0010] In one embodiment, enriching is performed by flow cytometry. In another embodiment, the one or more samples are blood samples. In yet another embodiment, the one or more samples are tissue samples.

[0011] In yet another embodiment, identifying an antigen-specific TCR sequence based on its relative change in abundance as a clone for therapeutic use for said HLA-presented antigen comprises ranking each of the antigen-specific TCR sequences based on its binding affinity for said HLA-presented antigen.

[0012] The method further comprises pairing the antigen-specific TCR sequence with a second TCR sequence that forms its cognate pair in the T cell. In some embodiments, the pairing comprises pairing a TCR heavy chain with a TCR light chain. In some embodiments, the TCR heavy chain is a TCRβ or a TCRδ chain. In some embodiments, the TCR light chain is a TCRα or a TCRγ chain. In some embodiments, the pairing comprises pairing a TCRβ chain with a TCRα chain. In some embodiments, the pairing comprising pairing a TCRδ chain with a TCRγ chain.

[0013] In certain embodiments, the invention comprises a method of identifying an antigen-specific T cell as a therapeutic, comprising: dividing one or more samples comprising T cells into a first subset and a second subset, performing immunosequencing of rearranged nucleic acid molecules extracted from the first subset of cells to obtain a first plurality of unique sequence reads, determining based on the first plurality of unique sequence reads a relative abundance of each unique T cell receptor (TCR) sequence out of a total number of T cells in the first subset, enriching the second subset of cells with multimer molecules comprising an HLA-presented antigen to identify a population of antigen-specific T cells that bind the multimer, performing immunosequencing of rearranged nucleic acid molecules extracted from said sorted population of antigen-specific T cells to obtain a second plurality of unique sequence reads, determining based on the second plurality of unique sequence reads a relative abundance of each unique TCR

sequence out of a total number of T cells in the second subset, determining a relative change in abundance of each antigen-specific TCR sequence based on the relative abundances of the antigen-specific TCR sequence in the first subset and the second subset, and identifying the top antigen-specific TCR sequences having the greatest relative change in abundance as therapeutics for said HLA-presented antigen. In certain embodiments, the method provides sorting the first subset with a high concentration of tetramer and sorting the second subset with a lower concentration multimer.

[0014] In some embodiments, the top antigen-specific TCR sequences comprise the top 100 ranked clones in the sample. In other embodiments, the top antigen-specific TCR sequences comprise the top 50 ranked clones in the sample. In another embodiment, the top antigen-specific TCR sequences comprise the top 10 ranked clones in the sample.

BRIEF DESCRPTION OF THE DRAWINGS

[0015] **Fig. 1** shows the generation of antigen-specific T cell lines from two donors that are specific for the WT1 peptide, WT_{37-45} (VLDFAPPGA).

[0016] **Fig. 2** shows frequency of each TCRβ clonotype calculated by dividing a post-sort frequency by the baseline frequency determined prior to sorting.

[0017] **Fig. 3** shows the final TCRs selected for synthesis.

[0018] **Fig. 4** shows Jurkat T cells transduced with lentiviral vectors encoding the TCRα and TCRβ gene sequences from the final TCRs selected for synthesis. A previously characterized TCR (P37) was included for comparison.

[0019] **Fig. 5** shows a comparison between the method of the current invention and conventional techniques.

DETAILED DESCRIPTION OF THE INVENTION

[0020] The practice of the present invention may employ, unless otherwise indicated, conventional techniques and descriptions of molecular biology (including recombinant techniques), bioinformatics, cell biology, and biochemistry, which are within the skill of the art. Such conventional techniques include, but are not limited to, sampling and analysis of blood cells, nucleic acid sequencing and analysis, and the like. Specific illustrations of suitable techniques can be had by reference to the example herein below. However, other equivalent

conventional procedures can, of course, also be used. Such conventional techniques and descriptions can be found in standard laboratory manuals.

Definitions

[0021] Terms used in the claims and specification are defined as set forth below unless otherwise specified.

[0022] As used herein, an adaptive immune receptor (AIR) refers to an immune cell receptor, e.g., a T cell receptor (TCR) or a B cell receptor (BCR) found in mammalian cells. In certain embodiments, the adaptive immune receptor is encoded by a TCRB, TCRG, TCRA, TCRD, IGH, IGK, and IGL gene or gene segment.

[0023] "Primer" means an oligonucleotide, either natural or synthetic that is capable, upon forming a duplex with a polynucleotide template, of acting as a point of initiation of nucleic acid synthesis and being extended from its 3' end along the template so that an extended duplex is formed. Extension of a primer is usually carried out with a nucleic acid polymerase, such as a DNA or RNA polymerase. The sequence of nucleotides added in the extension process is determined by the sequence of the template polynucleotide. Usually primers are extended by a DNA polymerase. Primers usually have a length in the range of from 14 to 40 nucleotides, or in the range of from 18 to 36 nucleotides. Primers are employed in a variety of nucleic amplification reactions, for example, linear amplification reactions using a single primer, or polymerase chain reactions, employing two or more primers. Guidance for selecting the lengths and sequences of primers for particular applications is well known to those of ordinary skill in the art, as evidenced by the following references that are incorporated by reference: Dieffenbach, editor, PCR Primer: A Laboratory Manual, 2nd Edition (Cold Spring Harbor Press, New York, 2003).In some embodiments, as used herein, the term "gene" refers to the segment of DNA involved in producing a polypeptide chain, such as all or a portion of a TCR or Ig polypeptide (*e.g.*, a CDR3-containing polypeptide); it includes regions preceding and following the coding region "leader and trailer" as well as intervening sequences (introns) between individual coding segments (exons), regulatory elements (*e.g.*, promoters, enhancers, repressor binding sites and the like), or recombination signal sequences (RSSs), as described herein.

[0024] The nucleic acids of the present embodiments, also referred to herein as polynucleotides, and including oligonucleotides, can be in the form of RNA or in the form of DNA, including cDNA, genomic DNA, and synthetic DNA. The DNA can be double-stranded or single-

stranded, and if single stranded, can be the coding strand or non-coding (anti-sense) strand. A coding sequence which encodes a TCR or an IG or a region thereof (*e.g.*, a V region, a D segment, a J region, a C region, *etc.*) for use according to the present embodiments can be identical to the coding sequence known in the art for any given TCR or immunoglobulin gene regions or polypeptide domains (*e.g.*, V-region domains, CDR3 domains, etc.), or can be a different coding sequence, which as a result of the redundancy or degeneracy of the genetic code, encodes the same TCR or immunoglobulin region or polypeptide.

[0025] "Polymerase chain reaction," or "PCR," means a reaction for the in vitro amplification of specific DNA sequences by the simultaneous primer extension of complementary strands of DNA. In other words, PCR is a reaction for making multiple copies or replicates of a target nucleic acid flanked by primer binding sites, such reaction comprising one or more repetitions of the following steps: (i) denaturing the target nucleic acid, (ii) annealing primers to the primer binding sites, and (iii) extending the primers by a nucleic acid polymerase in the presence of nucleoside triphosphates. Usually, the reaction is cycled through different temperatures optimized for each step in a thermal cycler instrument. Particular temperatures, durations at each step, and rates of change between steps depend on many factors well-known to those of ordinary skill in the art, e.g. exemplified by the references: McPherson et al, editors, PCR: A Practical Approach and PCR2: A Practical Approach (IRL Press, Oxford, 1991 and 1995, respectively). For example, in a conventional PCR using Taq DNA polymerase, a double stranded target nucleic acid may be denatured at a temperature >90°C, primers annealed at a temperature in the range 50-75°C, and primers extended at a temperature in the range 72-78°C. The term "PCR" encompasses derivative forms of the reaction, including but not limited to, RT-PCR, real-time PCR, nested PCR, quantitative PCR, multiplexed PCR, and the like. Reaction volumes range from a few hundred nanoliters, e.g. 200 nL, to a few hundred μL, e.g. 200 μL. "Reverse transcription PCR," or "RT-PCR," means a PCR that is preceded by a reverse transcription reaction that converts a target RNA to a complementary single stranded DNA, which is then amplified, e.g. Tecott et al., U.S. patent 5,168,038, which patent is incorporated herein by reference. "Real-time PCR" means a PCR for which the amount of reaction product, i.e. amplicon, is monitored as the reaction proceeds. There are many forms of real-time PCR that differ mainly in the detection chemistries used for monitoring the reaction product, e.g. Gelfand et al., U.S. Patent No. 5,210,015 ("taqman"); Wittwer et al., U.S. Patent Nos. 6,174,670 and

6,569,627 (intercalating dyes); Tyagi et al., U.S. Patent No. 5,925,517 (molecular beacons); which are incorporated herein by reference. Detection chemistries for real-time PCR are reviewed in Mackay et al., Nucleic Acids Research, 30: 1292-1305 (2002), which is also incorporated herein by reference. "Nested PCR" means a two-stage PCR wherein the amplicon of a first PCR becomes the sample for a second PCR using a new set of primers, at least one of which binds to an interior location of the first amplicon. As used herein, "initial primers" in reference to a nested amplification reaction mean the primers used to generate a first amplicon, and "secondary primers" mean the one or more primers used to generate a second, or nested, amplicon. "Multiplexed PCR" refers to a PCR wherein multiple target sequences (or a single target sequence and one or more reference sequences) are simultaneously carried out in the same reaction mixture, e.g. Bernard et al., Anal. Biochem., 273: 221-228 (1999)(two-color real-time PCR). Usually, distinct sets of primers are employed for each sequence being amplified. Typically, the number of target sequences in a multiplex PCR is in the range of 2 to 50, or from 2 to 40, or from 2 to 30. "Quantitative PCR" means a PCR designed to measure the abundance of one or more specific target sequences in a sample or specimen. Quantitative PCR includes both absolute quantitation and relative quantitation of such target sequences. Quantitative measurements are made using one or more reference sequences or internal standards that may be assayed separately or together with a target sequence. The reference sequence may be endogenous or exogenous to a sample or specimen, and in the latter case, may comprise one or more competitor templates. Typical endogenous reference sequences include segments of transcripts of the following genes: β-actin, GAPDH, β2-microglobulin, ribosomal RNA, and the like. Techniques for quantitative PCR are well-known to those of ordinary skill in the art, as exemplified in the following references that are incorporated by reference: Freeman et al., Biotechniques, 26: 112-126 (1999); Becker-Andre et al., Nucleic Acids Research, 17: 9437-9447 (1989); Zimmerman et al., Biotechniques, 21: 268-279 (1996); Diviacco et al., Gene, 122: 3013-3020 (1992); Becker-Andre et al., Nucleic Acids Research, 17: 9437-9446 (1989); and the like.

[0026] The term percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (e.g., BLASTP and BLASTN or other algorithms available to persons of skill) or by visual

inspection. Depending on the application, the percent "identity" can exist over a region of the sequence being compared, e.g., over a functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

[0027] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0028] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (*see* generally Ausubel *et al., infra*).

[0029] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/).

[0030] "Activation" or "immune activation" or "activated", especially in reference to T-cells, means a phase of an adaptive immune response that follows the antigen recognition phase (during which antigen-specific lymphocytes bind to antigens) and is characterized by proliferation of lymphocytes and their differentiation into effector cells, e.g. Abbas et al., Cellular and Molecular Immunology, Fourth Edition, (W.B. Saunders Company, 2000). Activation of T cells may be associated with secretion of certain cytokines that are detectable using conventional assays, such as an ELISPOT assay, and may be associated with the expression of characteristic cell surface markers, such as CD25, CD134, CD69, CD137, CD154, or the like, e.g. Gratama et al., Cytometry A, 73A: 971-974 (2008).

[0031] "Aligning" means a method of comparing a test sequence, such as a sequence read, to one or more reference sequences to determine which reference sequence or which portion of a

reference sequence is closest based on some sequence distance measure. An exemplary method of aligning nucleotide sequences is the Smith Waterman algorithm. Distance measures may include Hamming distance, Levenshtein distance, or the like. Distance measures may include a component related to the quality values of nucleotides of the sequences being compared.

[0032] "Amplicon" means the product of a polynucleotide amplification reaction. That is, a clonal population of polynucleotides, which may be single stranded or double stranded, which are replicated from one or more starting sequences. The one or more starting sequences may be one or more copies of the same sequence, or they may be a mixture of different sequences. Preferably, amplicons are formed by the amplification of a single starting sequence. Amplicons may be produced by a variety of amplification reactions whose products comprise replicates of the one or more starting, or target, nucleic acids. In one aspect, amplification reactions producing amplicons are "template-driven" in that base pairing of reactants, either nucleotides or oligonucleotides, have complements in a template polynucleotide that are required for the creation of reaction products. In one aspect, template-driven reactions are primer extensions with a nucleic acid polymerase or oligonucleotide ligations with a nucleic acid ligase. Such reactions include, but are not limited to, polymerase chain reactions (PCRs), linear polymerase reactions, nucleic acid sequence-based amplification (NASBAs), rolling circle amplifications, and the like, disclosed in the following references that are incorporated herein by reference: Mullis et al., U.S. Patent Nos. 4,683,195; 4,965,188; 4,683,202; 4,800,159 (PCR); Gelfand et al., U.S. Patent No. 5,210,015 (real-time PCR with "taqman" probes); Wittwer et al, U.S. Patent No. 6,174,670; Kacian et al., U.S. Patent No. 5,399,491 ("NASBA"); Lizardi, U.S. Patent No. 5,854,033; Aono et al., Japanese Patent Pub. No. JP 4-262799 (rolling circle amplification); and the like. In one aspect, amplicons of the invention are produced by PCRs. An amplification reaction may be a "real-time" amplification if a detection chemistry is available that permits a reaction product to be measured as the amplification reaction progresses, e.g. "real-time PCR" described below, or "real-time NASBA" as described in Leone et al, Nucleic Acids Research, 26: 2150-2155 (1998), and like references. As used herein, the term "amplifying" means performing an amplification reaction. A "reaction mixture" means a solution containing all the necessary reactants for performing a reaction, which may include, but not be limited to, buffering agents to maintain pH at a selected level during a reaction, salts, co-factors, scavengers, and the like.

[0033] "Clonotype" or "clone" means a rearranged or recombined nucleotide sequence of a lymphocyte which encodes an immune receptor or a portion thereof. More particularly, clonotype means a recombined nucleotide sequence of a T cell or B cell which encodes a T cell receptor (TCR) or B cell receptor (BCR), or a portion thereof. In various embodiments, clonotypes may encode all or a portion of a VDJ rearrangement of IgH, a DJ rearrangement of IgH, a VJ rearrangement of IgK, a VJ rearrangement of IgL, a VDJ rearrangement of TCR β, a DJ rearrangement of TCR β, a VJ rearrangement of TCR α, a VJ rearrangement of TCR γ, a VDJ rearrangement of TCR δ, a VD rearrangement of TCR δ, a Kde-V rearrangement, or the like. Clonotypes may also encode translocation breakpoint regions involving immune receptor genes, such as Bcl1-IgH or Bcl1-IgH. In one aspect, clonotypes have sequences that are sufficiently long to represent or reflect the diversity of the immune molecules that they are derived from; consequently, clonotypes may vary widely in length. In some embodiments, clonotypes have lengths in the range of from 25 to 400 nucleotides; in other embodiments, clonotypes have lengths in the range of from 25 to 200 nucleotides.

[0034] "Clonotype profile" means a listing of distinct clonotypes and their relative abundances that are derived from a population of lymphocytes. The population of lymphocytes can be obtained from a tissue sample or a blood sample. The term "clonotype profile" is related to, but more general than, the immunology concept of an immune "repertoire" as described in references, such as the following: Arstila et al, Science, 286: 958-961 (1999); Yassai et al, Immunogenetics, 61: 493-502 (2009); Kedzierska et al, Mol. Immunol., 45(3): 607-618 (2008); and the like. The term "clonotype profile" includes a wide variety of lists and abundances of rearranged immune receptor-encoding nucleic acids, which may be derived from selected subsets of lymphocytes (e.g. tissue-infiltrating lymphocytes, immunophenotypic subsets, or the like), or which may encode portions of immune receptors that have reduced diversity as compared to full immune receptors. In some embodiments, clonotype profiles may comprise abundances or relative frequencies of each of the distinct clonotypes. Another measure of a clonotype profile is the clonality, which is a measurement of the diversity of the clonotypes. In one aspect, a clonotype profile is a set of distinct recombined nucleotide sequences (with their abundances) that encode T cell receptors (TCRs) or B cell receptors (BCRs), or fragments thereof, respectively, in a population of lymphocytes of an individual, wherein the nucleotide sequences of the set have a one-to-one correspondence with distinct lymphocytes or their clonal

subpopulations for substantially all of the lymphocytes of the population. In one aspect, nucleic acid segments defining clonotypes are selected so that their diversity (i.e. the number of distinct nucleic acid sequences in the set) is large enough so that substantially every T cell or B cell or clone thereof in an individual carries a unique nucleic acid sequence of such repertoire. That is, preferably each different clone of a sample has different clonotype. In other aspects of the invention, the population of lymphocytes corresponding to a repertoire may be circulating B cells, or may be circulating T cells, or may be subpopulations of either of the foregoing populations, including but not limited to, CD4+ T cells, or CD8+ T cells, or other subpopulations defined by cell surface markers, or the like. Such subpopulations may be acquired by taking samples from particular tissues, e.g. bone marrow, or lymph nodes, or the like, or by sorting or enriching cells from a sample (such as peripheral blood) based on one or more cell surface markers, size, morphology, or the like. In still other aspects, the population of lymphocytes corresponding to a repertoire may be derived from disease tissues, such as a tumor tissue, an infected tissue, or the like. In one embodiment, a clonotype profile comprising human TCR β chains or fragments thereof comprises a number of distinct nucleotide sequences in the range of from 0.1×10^6 to 1.8×10^6, or in the range of from 0.5×10^6 to 1.5×10^6, or in the range of from 0.8×10^6 to 1.2×10^6. In another embodiment, a clonotype profile comprising human IgH chains or fragments thereof comprises a number of distinct nucleotide sequences in the range of from 0.1×10^6 to 1.8×10^6, or in the range of from 0.5×10^6 to 1.5×10^6, or in the range of from 0.8×10^6 to 1.2×10^6. In a particular embodiment, a clonotype profile of the invention comprises a set of nucleotide sequences encoding substantially all segments of the V(D)J region of an IgH chain. In one aspect, "substantially all" as used herein means every segment having a relative abundance of .001 percent or higher; or in another aspect, "substantially all" as used herein means every segment having a relative abundance of .0001 percent or higher. "Substantially all" can also mean at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% of the unique clones found in a sample. In another particular embodiment, a clonotype profile of the invention comprises a set of nucleotide sequences that encodes substantially all segments of the V(D)J region of a TCR β chain. In another embodiment, a clonotype profile of the invention comprises a set of nucleotide sequences having lengths in the range of from 25-600 nucleotides and including segments of the V, D, and J regions of a TCR β chain. In another embodiment, a clonotype profile of the invention comprises a set of nucleotide sequences having lengths in the

range of from 25-600 nucleotides and including segments of the V, D, and J regions of an IgH chain. In another embodiment, a clonotype profile of the invention comprises a number of distinct nucleotide sequences that is substantially equivalent to the number of lymphocytes expressing a distinct IgH chain. In another embodiment, a clonotype profile of the invention comprises a number of distinct nucleotide sequences that is substantially equivalent to the number of lymphocytes expressing a distinct TCR β chain. In still another embodiment, "substantially equivalent" means that with ninety-nine percent probability, a clonotype profile will include a nucleotide sequence encoding an IgH or TCR β or portion thereof carried or expressed by every lymphocyte of a population of an individual at a frequency of .001 percent or greater. In still another embodiment, "substantially equivalent" means that with ninety-nine percent probability, a repertoire of nucleotide sequences will include a nucleotide sequence encoding an IgH or TCR β or portion thereof carried or expressed by every lymphocyte present at a frequency of .0001 percent or greater.

[0035] "Coalescing" or "clustering" means treating two candidate clonotypes with sequence differences as the same by determining that such differences are due to experimental or measurement error and not due to genuine biological differences. In one aspect, a sequence of a higher frequency candidate clonotype is compared to that of a lower frequency candidate clonotype and if predetermined criteria are satisfied then the number of lower frequency candidate clonotypes is added to that of the higher frequency candidate clonotype and the lower frequency candidate clonotype is thereafter disregarded. That is, the read counts associated with the lower frequency candidate clonotype are added to those of the higher frequency candidate clonotype.

[0036] "Complementarity determining regions" (CDRs) mean regions of an immunoglobulin (i.e., antibody) or T cell receptor where the molecule complements an antigen's conformation, thereby determining the molecule's specificity and contact with a specific antigen. T cell receptors and immunoglobulins each have three CDRs: CDR1 and CDR2 are found in the variable (V) domain, and CDR3 includes some of V, all of diverse (D) (heavy chains only) and joint (J), and some of the constant (C) domains.

[0037] The term "ameliorating" refers to any therapeutically beneficial result in the treatment of a disease state, e.g., a disease state, including prophylaxis, lessening in the severity or progression, remission, or cure thereof.

[0038] The term "in situ" refers to processes that occur in a living cell growing separate from a living organism, e.g., growing in tissue culture.

[0039] The term "in vivo" refers to processes that occur in a living organism.

[0040] The term "mammal" as used herein includes but is not limited to humans, non-human primates, canines, felines, murines, bovines, equines, and porcines.

[0041] Unless specific definitions are provided, the nomenclature utilized in connection with, and the laboratory procedures and techniques of, molecular biology, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques can be used for recombinant technology, molecular biological, microbiological, chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

[0042] Unless the context requires otherwise, throughout the present specification and claims, the word "comprise" and variations thereof, such as, "comprises" and "comprising" are to be construed in an open, inclusive sense, that is, as "including, but not limited to." By "consisting of" is meant including, and typically limited to, whatever follows the phrase "consisting of." By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that no other elements are required and can or cannot be present depending upon whether or not they affect the activity or action of the listed elements.

[0043] It must be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

[0044] As used herein, in particular embodiments, the terms "about" or "approximately" when preceding a numerical value indicates the value plus or minus a range of 5%, 6%, 7%, 8% or 9%, or greater, *etc.* In other embodiments, the terms "about" or "approximately" when preceding a numerical value indicates the value plus or minus a range of 10%, 11%, 12%, 13% or 14%, or greater, *etc.* In yet other embodiments, the terms "about" or "approximately" when preceding a numerical value indicates the value plus or minus a range of 15%, 16%, 17%, 18%, 19% or 20%, or greater, *etc.*

[0045] Reference throughout this specification to "one embodiment" or "an embodiment" or "an aspect" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, the appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics can be combined in any suitable manner in one or more embodiments.

Samples

[0046] Samples used in the methods of the invention can include, any tissue from a subject where there are lymphocytes or a lymphoid infiltrate in the tissue, and the lymphoid infiltrate can be malignant or benign. Samples can be obtained from a bodily fluid from a subject, such as a peripheral blood sample. Other examples of samples include, but not limited to, urine, saliva, internal body fluids, organ tissue, lymph tissue, skin tissue, or a biopsy of a solid tumor.

[0047] In some embodiments, the subject is a mammalian subject, for example, a human subject. In one embodiment, the subject is a healthy subject. In other embodiments, the subject has a disease or condition of interest, such as cancer, autoimmune disease, etc. In another embodiment, samples from the subject are obtained prior to and after a medical event, such as a treatment, immunotherapy, surgery, or vaccination. In yet another embodiment, samples are obtained from the subject and analyzed before and after a stimulation event, such as an enrichment (in vitro stimulation of lymphocytes with an antigen), or a mixed lymphocyte reaction.

[0048] The sample includes T-cells and/or B-cells. T-cells (T lymphocytes) include, for example, cells that express T cell receptors. T-cells include Helper T cells (effector T cells or Th cells), cytotoxic T cells (CTLs), memory T cells, and regulatory T cells. The sample can include one or more expanded clones, including one or more dominant clones (e.g., a top T cell clone), among a number of T cells or a total number of nucleated cells. The sample can include at least 1,000, at least 10,000, at least 100,000, at least 250,000, at least 500,000, at least 750,000, or at least 1,000,000 T-cells.

[0049] B-cells include, for example, plasma B cells, memory B cells, B1 cells, B2 cells, marginal-zone B cells, and follicular B cells. B-cells can express immunoglobulins (Igs, antibodies, B cell receptor). The sample can include one or more expanded clones, including a

dominant clone (e.g., a top B cell clone), among a number of benign B cells or a total number of nucleated cells. The sample can include a single B cell in some applications or more generally at least 1,000, at least 10,000, at least 100,000, at least 250,000, at least 500,000, at least 750,000, or at least 1,000,000 B-cells.

[0050] The sample can include nucleic acid molecules extracted from a cell, for example, DNA (e.g., genomic DNA or mitochondrial DNA) or RNA (e.g., messenger RNA or microRNA). The nucleic acid can be cell-free DNA or RNA. In other embodiments, the sample comprises complementary DNA (cDNA) that has been reverse transcribed from mRNA. In the methods of the provided invention, the amount of RNA or DNA from a subject that can be analyzed includes, for example, as low as a single cell in some applications and as many as 10 million cells or more, translating to a range of DNA of 6 pg-60 μg, and RNA of approximately 1 pg-10 μg.

Cells

[0051] B cells and T cells can be obtained from a biological sample, such as from a variety of tissues, solid tumor samples, and biological fluid samples, including skin tissue, bone marrow, thymus, lymph glands, lymph nodes, peripheral tissues and peripheral blood.

[0052] Any peripheral tissue can be sampled for the presence of B and T cells and is therefore contemplated for use in the methods described herein. Tissues and biological fluids from which adaptive immune cells may be obtained include, but are not limited to skin, epithelial tissues, colon, spleen, a mucosal secretion, oral mucosa, intestinal mucosa, vaginal mucosa or a vaginal secretion, cervical tissue, ganglia, saliva, cerebrospinal fluid (CSF), bone marrow, cord blood, serum, serosal fluid, plasma, lymph, urine, ascites fluid, pleural fluid, pericardial fluid, peritoneal fluid, abdominal fluid, culture medium, conditioned culture medium or lavage fluid. In certain embodiments, adaptive immune cells may be isolated from an apheresis sample. Peripheral blood samples may be obtained by phlebotomy from subjects. Peripheral blood mononuclear cells (PBMC) are isolated by techniques known to those of skill in the art, *e.g.*, by Ficoll-Hypaque® density gradient separation. In certain embodiments, whole PBMCs are used for analysis.

[0053] In other embodiments, the sample comprises solid tumor tissue, a circulating blood mononuclear cell fraction, or cells collected from urinary sediment.

[0054] In certain related embodiments, preparations that comprise predominantly lymphocytes (*e.g.*, T and B cells) or that comprise predominantly T cells or predominantly B cells, may be prepared. In other related embodiments, specific subpopulations of T or B cells may be isolated prior to analysis using the methods described herein. Various methods and commercially available kits for isolating different subpopulations of T and B cells are known in the art and include, but are not limited to, subset selection immunomagnetic bead separation or flow immunocytometric cell sorting using antibodies specific for one or more of any of a variety of known T and B cell surface markers. Illustrative markers include, but are not limited to, one or a combination of CD2, CD3, CD4, CD8, CD14, CD19, CD20, CD25, CD28, CD45RO, CD45RA, CD54, CD62, CD62L, CDw137 (41BB), CD154, GITR, FoxP3, CD54, and CD28. For example, and as is known to the skilled person, cell surface markers, such as CD2, CD3, CD4, CD8, CD14, CD19, CD20, CD45RA, and CD45RO may be used to determine T, B, and monocyte lineages and subpopulations in flow cytometry. Similarly, forward light-scatter, side-scatter, and/or cell surface markers such as CD25, CD62L, CD54, CD137, and CD154 may be used to determine activation state and functional properties of cells.

[0055] Illustrative combinations useful in certain of the methods described herein may include $CD8^+CD45RO^+$ (memory cytotoxic T cells), $CD4^+CD45RO^+$ (memory T helper), $CD8^+CD45RO^-$ ($CD8^+CD62L^+CD45RA^+$ (naïve-like cytotoxic T cells); $CD4^+CD25^+CD62L^{hi}GITR^+FoxP3^+$ (regulatory T cells). Illustrative antibodies for use in immunomagnetic cell separations or flow immunocytometric cell sorting include fluorescently labeled anti-human antibodies, *e.g.*, CD4 FITC (clone M-T466, Miltenyi Biotec), CD8 PE (clone RPA-T8, BD Biosciences), CD45RO ECD (clone UCHL-1, Beckman Coulter), and CD45RO APC (clone UCHL-1, BD Biosciences). Staining of total PBMCs may be done with the appropriate combination of antibodies, followed by washing cells before analysis. Lymphocyte subsets can be isolated by fluorescence activated cell sorting (FACS), *e.g.*, by a BD FACSAria™ cell-sorting system (BD Biosciences) and by analyzing results with FlowJo™ software (Treestar Inc.), and also by conceptually similar methods involving specific antibodies immobilized to surfaces or beads.

Nucleic Acid Extraction

[0056] In some embodiments, total genomic DNA can be extracted from cells by methods known to those of skill in the art. Examples include using the QIAamp® DNA blood Mini Kit

(QIAGEN®). The approximate mass of a single haploid genome is 3 pg. Preferably, at least 100,000 to 200,000 cells are used for analysis of diversity, i.e., about 0.6 to 1.2 μg DNA from diploid T cells. Using PBMCs as a source, the number of T cells can be estimated to be about 30% of total cells.

[0057] In some embodiments, RNA can be extracted from cells in a sample, such as a sample of blood, lymph, tissue, or other sample from a subject known to contain lymphoid cells, using standard methods or commercially available kits known in the art. In other embodiments, cDNA can be transcribed from mRNA obtained from the cells and then used as templates in a multiplex PCR.

[0058] Alternatively, total nucleic acid can be isolated from cells, including both genomic DNA and mRNA. If diversity is to be measured from mRNA in the nucleic acid extract, the mRNA can be converted to cDNA prior to measurement. This can readily be done by methods of one of ordinary skill, for example, using reverse transcriptase according to known procedures.

[0059] In certain embodiments, DNA can be isolated from frozen, OCT embedded or formalin fixed paraffin embedded (FFPE) skin samples. For OCT embedded tissue samples, cryosections can be cut and DNA extraction can be carried extracted using known techniques. For FFPE samples, paraffin is first removed from the tissue scrolls and DNA can then be extracted by known techniques.

Multiplex Quantitative PCR

[0060] "Multiplex PCR" or "multiplexed PCR" refers to a PCR wherein multiple target sequences are simultaneously amplified by a set of primers in the same reaction mixture. Multiplex quantitative PCR is described herein and in Robins *et al.*, 2009 *Blood* 114, 4099; Robins *et al.*, 2010 *Sci. Translat. Med.* 2:47ra64; Robins *et al.*, 2011 *J.* Immunol. *Meth.* doi:10.1016/j.jim.2011.09. 001; Sherwood *et al.* 2011 *Sci. Translat. Med.* 3:90ra61; U.S. 2012/0058902; U.S 2010/0330571; WO/2011/106738 (PCT/US2011/026373); , U.S. 9,279,159; 9,181,590; and 9,181,591, which are each incorporated by reference in its entirety. In one embodiment, a single multiplex PCR method uses a set of forward primers that specifically hybridize to V segments and a set of reverse primers that specifically hybridize to the J segments of a TCR or IG locus, where a single multiplex PCR reaction using the primers allows amplification of all the possible VJ (and VDJ) combinations within a given population of T or B cells.

[0061] A single multiplex PCR system can be used to amplify rearranged adaptive immune cell receptor loci from genomic DNA, preferably from a CDR3 region. In certain embodiments, the CDR3 region is amplified from a TCRA, TCRB, TCRG or TCRD CDR3 region or similarly from an IGH or IGL (lambda or kappa) locus. Compositions are provided that comprise a plurality of V-segment and J-segment primers that are capable of promoting amplification in a multiplex polymerase chain reaction (PCR) of substantially all productively rearranged adaptive immune receptor CDR3-encoding regions in the sample for a given class of such receptors to produce a multiplicity of amplified rearranged DNA molecules from a population of T cells (for TCR) or B cells (for IG) in the sample. In certain embodiments, primers are designed so that each amplified rearranged DNA molecule is less than 600 nucleotides in length, thereby excluding amplification products from non-rearranged adaptive immune receptor loci.

[0062] In some embodiments, the method uses two pools of primers to provide for a highly multiplexed, single tube PCR reaction. A "forward" pool of primers can include a plurality of V-segment oligonucleotide primers used as "forward" primers and a plurality of J-segment oligonucleotide primers used as "reverse" primers. In other embodiments, J-segment primers can be used as "forward" primers, and V-segment can be used "reverse" primers. In some embodiments, an oligonucleotide primer that is specific to (*e.g.*, having a nucleotide sequence complementary to a unique sequence region of) each V-region encoding segment ("V segment) in the respective TCR or IG gene locus can be used. In other embodiments, primers targeting a highly conserved region are used to simultaneously amplify multiple V segments or multiple J segments, thereby reducing the number of primers required in the multiplex PCR. In certain embodiments, the J-segment primers anneal to a conserved sequence in the joining ("J") segment.

[0063] Each primer can be designed such that a respective amplified DNA segment is obtained that includes a sequence portion of sufficient length to identify each J-segment unambiguously based on sequence differences amongst known J-region encoding gene segments in the human genome database, and also to include a sequence portion to which a J-segment specific primer can anneal for resequencing. This design of V- and J-segment specific primers enables direct observation of a large fraction of the somatic rearrangements present in the adaptive immune receptor gene repertoire within an individual. This feature in turn enables rapid comparison of the TCR and/or IG repertoires in individuals pre-transplant and post-transplant, for example.

[0064] In one embodiment, the present disclosure provides a plurality of V-segment primers and a plurality of J-segment primers, wherein the plurality of V-segment primers and the plurality of J-segment primers amplify all or substantially all combinations of the V- and J-segments of a rearranged immune receptor locus. In some embodiments, the method provides amplification of substantially all of the rearranged adaptive immune receptor (AIR) sequences in a lymphoid cell and is capable of quantifying the diversity of the TCR or IG repertoire of at least 10^6, 10^5, 10^4, or 10^3 unique rearranged AIR sequences in a sample. "Substantially all combinations" can refer to at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more of all the combinations of the V- and J-segments of a rearranged immune receptor locus. In certain embodiments, the plurality of V-segment primers and the plurality of J-segment primers amplify all of the combinations of the V- and J-segments of a rearranged adaptive immune receptor locus.

[0065] In general, a multiplex PCR system can use 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25, and in certain embodiments, at least 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, or 39, and in other embodiments 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 65, 70, 75, 80, 85, or more forward primers, in which each forward primer specifically hybridizes to or is complementary to a sequence corresponding to one or more V region segments. The multiplex PCR system also uses at least 2, 3, 4, 5, 6, or 7, and in certain embodiments, 8, 9, 10, 11, 12 or 13 reverse primers, or 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more primers, in which each reverse primer specifically hybridizes to or is complementary to a sequence corresponding to one or more J region segments. In some embodiments, each reverse J primer is specific to a different J gene segment. In other embodiments, there is no common J primer that binds to all J gene segments.

[0066] The V segment and J segment primers have certain characteristics to amplify the total diversity of TCR or IG repertoires. In certain embodiments, the V segment primers have similar melting temperatures within a range of 0.1°C, 0.2°C, 0.3°C, 0.4°C, 0.5°C, 0.6°C, 0.7°C, 0.8°C, 0.9°C, 1.0°C, 1.1°C, 1.2°C, 1.3°C, 1.4°C, 1.5°C, 1.6°C, 1.7°C, 1.8°C, 1.9°C, 2.0°C, 2.1°C, 2.2°C, 2.3°C, 2.4°C, 2.5°C, 2.6°C, 2.7°C, 2.8°C, 2.9°C, 3.0°C, 3.1°C, 3.2°C, 3.3°C, 3.4°C, 3.5°C, 3.6°C, 3.7°C, 3.8°C, 3.9°C, 4.0°C, 4.5°C, 5.0°C. In some embodiments, the J segment primers have similar melting temperatures within a range of 0.1°C, 0.2°C, 0.3°C, 0.4°C, 0.5°C, 0.6°C, 0.7°C, 0.8°C, 0.9°C, 1.0°C, 1.1°C, 1.2°C, 1.3°C, 1.4°C, 1.5°C, 1.6°C, 1.7°C, 1.8°C,

1.9°C, 2.0°C, 2.1°C, 2.2°C, 2.3°C, 2.4°C, 2.5°C, 2.6°C, 2.7°C, 2.8°C, 2.9°C, 3.0°C, 3.1°C, 3.2°C, 3.3°C, 3.4°C, 3.5°C, 3.6°C, 3.7°C, 3.8°C, 3.9°C, 4.0°C, 4.5°C, 5.0°C.

[0067] In certain embodiments, the plurality of V segment and J segment primers are not consensus primers. The V segment and J segment primers are not universal, degenerate primers. In some embodiments, each V segment primer is complementary to a single V segment or a family of V segments. In some embodiments, each J segment primer is complementary to a single J segment or a family of J segments. In other embodiments, each J segment primer is complementary and specific to a single J segment gene.

[0068] In other embodiments, the plurality of V segment and J segment primers sit outside a region of untemplated deletions in the TCR or IG locus. In some embodiments, the 3' end of the V segment primers are complementary to a target region that is at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more nucleotides upstream from the V-RSS. In some embodiments, the 3' end of the J segment primers are complementary to a target region that is at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more nucleotides downstream from the J-RSS.

[0069] Various combinations of V and J segment primers can be used to amplify the full diversity of TCR and IG sequences in a repertoire. For details on the multiplex PCR system, including exemplary primer oligonucleotide sequences for amplifying substantially all TCR and/or IG sequences, see, *e.g.,* Robins *et al.*, 2009 *Blood* 114, 4099; Robins *et al.*, 2010 *Sci. Translat. Med.* 2:47ra64; Robins *et al.*, 2011 *J. Immunol. Meth.* doi:10.1016/j.jim.2011.09. 001; Sherwood *et al.* 2011 *Sci. Translat. Med.* 3:90ra61; U.S. 2012/0058902; U.S 2010/0330571; WO/2011/106738 (PCT/US2011/026373); , U.S. 9,279,159; 9,181,590; and 9,181,591 , which is each incorporated by reference in its entirety.

[0070] Oligonucleotides or polynucleotides that are capable of specifically hybridizing or annealing to a target nucleic acid sequence by nucleotide base complementarity can do so under moderate to high stringency conditions. For purposes of illustration, suitable moderate to high stringency conditions for specific PCR amplification of a target nucleic acid sequence would be between 25 and 80 PCR cycles, with each cycle consisting of a denaturation step (*e.g.,* about 10-30 seconds (s) at greater than about 95°C), an annealing step (*e.g.,* about 10-30s at about 60-68°C), and an extension step (*e.g.,* about 10-60s at about 60-72°C), optionally according to certain embodiments with the annealing and extension steps being combined to provide a two-

step PCR. As would be recognized by the skilled person, other PCR reagents can be added or changed in the PCR reaction to increase specificity of primer annealing and amplification, such as altering the magnesium concentration, optionally adding DMSO, and/or the use of blocked primers, modified nucleotides, peptide-nucleic acids, and the like.

[0071] In certain embodiments, nucleic acid hybridization techniques can be used to assess hybridization specificity of the primers described herein. Hybridization techniques are well known in the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide as provided herein with other polynucleotides include prewashing in a solution of 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-60°C, 5X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC containing 0.1% SDS. One skilled in the art will understand that the stringency of hybridization can be readily manipulated, such as by altering the salt content of the hybridization solution and/or the temperature at which the hybridization is performed. For example, in another embodiment, suitable highly stringent hybridization conditions include those described above, with the exception that the temperature of hybridization is increased, *e.g.*, to 60°C-65°C or 65°C-70°C.

[0072] In certain embodiments, the primers are designed not to cross an intron/exon boundary. The forward primers in certain embodiments anneal to the V segments in a region of relatively strong sequence conservation between V segments so as to maximize the conservation of sequence among these primers. Accordingly, this minimizes the potential for differential annealing properties of each primer, and so that the amplified region between V and J primers contains sufficient TCR or Ig V sequence information to identify the specific V gene segment used. In one embodiment, the J segment primers hybridize with a conserved element of the J segment and have similar annealing strength. In one particular embodiment, the J segment primers anneal to the same conserved framework region motif. In certain embodiments, the J segment primers have a melting temperature range within 10°C, 7.5°C, 5°C, or 2.5°C or less.

[0073] Oligonucleotides (*e.g.*, primers) can be prepared by any suitable method, including direct chemical synthesis by a method such as the phosphotriester method of Narang *et al.*, 1979, *Meth. Enzymol.* 68:90-99; the phosphodiester method of Brown *et al.*, 1979, *Meth. Enzymol.* 68:109-151; the diethylphosphoramidite method of Beaucage *et al.*, 1981, *Tetrahedron Lett.* 22:1859-1862; and the solid support method of U.S. Pat. No. 4,458,066, each incorporated herein by

reference. A review of synthesis methods of conjugates of oligonucleotides and modified nucleotides is provided in Goodchild, 1990, *Bioconjugate Chemistry* 1(3): 165-187, incorporated herein by reference.

[0074] A primer is preferably a single-stranded oligonucleotide. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 6 to 50 nucleotides, 15-50 nucleotides, or in certain embodiments, from 15-35 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template nucleic acid, but must be sufficiently complementary to hybridize with the template. The design of suitable primers for the amplification of a given target sequence is well known in the art and described in the literature cited herein.

[0075] As described herein, primers can incorporate additional features which allow for the detection or immobilization of the primer, but do not alter the basic property of the primer, that of acting as a point of initiation of DNA synthesis. For example, primers can contain an additional nucleic acid sequence at the 5' end, which does not hybridize to the target nucleic acid, but which facilitates cloning, detection, or sequencing of the amplified product. The region of the primer which is sufficiently complementary to the template to hybridize is referred to herein as the hybridizing region.

[0076] As used herein, a primer is "specific" for a target sequence if, when used in an amplification reaction under sufficiently stringent conditions, the primer hybridizes primarily to the target nucleic acid. Typically, a primer is specific for a target sequence if the primer-target duplex stability is greater than the stability of a duplex formed between the primer and any other sequence found in the sample. One of skill in the art will recognize that various factors, such as salt conditions as well as base composition of the primer and the location of the mismatches, will affect the specificity of the primer, and that routine experimental confirmation of the primer specificity will be needed in many cases. Hybridization conditions can be chosen under which the primer can form stable duplexes only with a target sequence. Thus, the use of target-specific primers under suitably stringent amplification conditions enables the selective amplification of those target sequences which contain the target primer binding sites. In other terms, the primers of the invention are each complementary to a target sequence and can include 1, 2, or more

mismatches without reducing complementarity or hybridization of the primer to the target sequence.

[0077] In particular embodiments, primers for use in the methods described herein comprise or consist of a nucleic acid of at least about 15 nucleotides long that has the same sequence as, or is substantially complementary to, a contiguous nucleic acid sequence of the target V or J segment. Longer primers, *e.g.*, those of about 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 65, 70, 75, 80 or more nucleotides long that have the same sequence as, or sequence complementary to, a contiguous sequence of the target V or J segment, will also be of use in certain embodiments. Various mismatches (1, 2, 3, or more) to the target sequence can be contemplated in the primers, while preserving complementarity to the target V or J segment. All intermediate lengths of the aforementioned primers are contemplated for use herein. As would be recognized by the skilled person, the primers can have additional sequence added (*e.g.*, nucleotides that cannot be the same as or complementary to the target V or J segment), such as restriction enzyme recognition sites, adaptor sequences for sequencing, bar code sequences, and the like (see *e.g.*, primer sequences provided herein and in the sequence listing). Therefore, the length of the primers can be longer, such as 55, 56, 57, 58, 59, 60, 65, 70, 75, or 80 or more nucleotides in length or more, depending on the specific use or need.

[0078] For example, in one embodiment, the forward and reverse primers are both modified at the 5' end with the universal forward primer sequence compatible with a DNA sequencing nucleic acid sequence. Such universal primers sequences can be adapted to those used in the Illumina GAII single-end read sequencing system. Exemplary universal primer sequences and sequencing oligonucleotides are provided in U.S. 2012/0058902; U.S.2010/0330571; U.S. 2014/0322716; and U.S. 2015/0299786, which are incorporated by reference in their entireties.

[0079] In some embodiments, the forward and reverse primers are both modified at the 5' end with an adaptor sequence that is not complementary to the V-segment, J-segment, or C-segment (target sequence) and can be a region that is identical to or complementary to a second set of primers or a sequencing oligonucleotide. The adaptor sequence can be complementary to a second set of primers that are used in a second amplification reaction. The second set of primers can include a region complementary to the adaptor sequence and one or more other sequences (barcode sequence, random sequences, or other sequencing oligonucleotide sequences).

[0080] The adaptor sequence can be a universal adaptor oligonucleotide sequences or sequencing platform-specific oligonucleotide sequences that are specific to a single-molecule sequencing technology being employed. Examples of sequencers include the HiSeq™ or GeneAnalyzer™-2 (GA-2) systems (Illumina, Inc., San Diego, CA) or another suitable sequencing suite of instrumentation, reagents and software. Inclusion of such platform-specific adaptor sequences permits direct quantitative sequencing of amplification products. This feature therefore advantageously permits qualitative and quantitative characterization of the composition. In one example, dsDNA amplification products may be generated that have universal adaptor sequences at both ends, so that the adaptor sequences can be used to further incorporate sequencing platform-specific oligonucleotides at each end of each template.

[0081] As would be recognized by the skilled person, in certain embodiments, other modifications may be made to the primers, such as the addition of restriction enzyme sites, fluorescent tags, and the like, depending on the specific application.

[0082] Also contemplated are adaptive immune receptor V-segment or J-segment oligonucleotide primer variants that can share a high degree of sequence identity to the oligonucleotide primers. Thus, in these and related embodiments, adaptive immune receptor V-segment or J-segment oligonucleotide primer variants can have substantial identity to the adaptive immune receptor V-segment or J-segment oligonucleotide primer sequences disclosed herein. For example, such oligonucleotide primer variants can comprise at least 70% sequence identity, preferably at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or higher sequence identity compared to a reference polynucleotide sequence such as the oligonucleotide primer sequences disclosed herein, using the methods described herein (*e.g.*, BLAST analysis using standard parameters). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding ability of an oligonucleotide primer variant to anneal to an adaptive immune receptor segment-encoding polynucleotide by taking into account codon degeneracy, reading frame positioning and the like. Typically, oligonucleotide primer variants will contain one or more substitutions, additions, deletions and/or insertions, preferably such that the annealing ability of the variant oligonucleotide is not substantially diminished relative to that of an adaptive immune receptor V-segment or J-segment oligonucleotide primer sequence that is specifically set forth herein. As also noted elsewhere herein, in preferred embodiments adaptive immune receptor V-segment and J-segment

oligonucleotide primers are designed to be capable of amplifying a rearranged TCR or IGH sequence that includes the coding region for CDR3.

[0083] According to certain embodiments, the primers for use in the multiplex PCR methods of the present disclosure can be functionally blocked to prevent non-specific priming of non-T or B cell sequences. For example, the primers can be blocked with chemical modifications as described in U.S. 2010/0167353.

[0084] In some embodiments, the V- and J-segment primers are used to produce a plurality of amplicons from the multiplex PCR reaction. In certain embodiments, the V-segment primer sand J-segment primers can produce at least 10^6 amplicons representing the diversity of TCR or IG rearranged CDR3 molecules in the sample. In some embodiments, the amplicons range in size from 10, 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500 to 1600 nucleotides in length. In preferred embodiments, the amplicons have a size between 50-600 nucleotides in length.

[0085] According to non-limiting theory, these embodiments exploit current understanding in the art that once an adaptive immune cell (*e.g.*, a T or B lymphocyte) has rearranged its adaptive immune receptor-encoding (*e.g.*, TCR or Ig) genes, its progeny cells possess the same adaptive immune receptor-encoding gene rearrangement, thus giving rise to a clonal population that can be uniquely identified by the presence therein of rearranged (*e.g.*, CDR3-encoding) V- and J-gene segments that can be amplified by a specific pairwise combination of V- and J-specific oligonucleotide primers as herein disclosed.

Amplification Bias Control

[0086] Multiplex PCR assays can result in a bias in the total numbers of amplicons produced from a sample, given that certain primer sets are more efficient in amplification than others. To overcome the problem of such biased utilization of subpopulations of amplification primers, methods can be used that provide a template composition for standardizing the amplification efficiencies of the members of an oligonucleotide primer set, where the primer set is capable of amplifying rearranged DNA encoding a plurality of adaptive immune receptors (TCR or Ig) in a biological sample that comprises DNA from lymphoid cells.

[0087] Since accurate quantification of clones for CTCL detection is critical, an approach can be used to ensure minimal bias in multiplex PCR. *See* Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung M-W, Parsons JM, et al. Using synthetic templates to design an unbiased

multiplex PCR assay. Nature Communications. 2013;4:2680, which is incorporated by reference. For example, each potential VDJ rearrangement of the TCRB locus contains one of thirteen J segments, one of 2 D segments and one of 52 V segments, many of which have disparate nucleotide sequences. In order to amplify all possible VDJ combinations, a single tube, multiplex PCR assay with 45 V forward and 13 J reverse primers was used. To remove potential PCR bias, every possible V-J pair was chemically synthesized as a template with specific barcodes. *Id.* These templates were engineered so as to be recognizable as non-biologic and have universal 3' and 5' ends to permit amplification with universal primers and subsequent quantification by HTS. This synthetic immune system can then be used to calibrate the multiplex PCR assay. Iteratively, the multiplex pool of templates is amplified and sequenced with TCRB V/J-specific primers, and the primer concentrations are adjusted to re-balance PCR amplification. Once the multiplex primer mixture amplifies each V and J template nearly equivalently, residual bias is removed computationally. The parallel procedure for TCRG was described previously in Carlson et al. Nature Communications. 2013; 4:2680.

[0088] In some embodiments, the synthetic templates comprise a template composition of general formula (I):

[0089] 5'-U1-B1-V-B2-X-J-B3-U2-3' (I)

[0090] The constituent template oligonucleotides, of which the template composition is comprised, are diverse with respect to the nucleotide sequences of the individual template oligonucleotides. The individual template oligonucleotides can vary in nucleotide sequence considerably from one another as a function of significant sequence variability among the large number of possible TCR or BCR variable (V) and joining (J) region polynucleotides. Sequences of individual template oligonucleotide species can also vary from one another as a function of sequence differences in U1, U2, B (B1, B2 and B3) and R oligonucleotides that are included in a particular template within the diverse plurality of templates.

[0091] In certain embodiments, V is a polynucleotide comprising at least 20, 30, 60, 90, 120, 150, 180, or 210, and not more than 1000, 900, 800, 700, 600 or 500 contiguous nucleotides of an adaptive immune receptor variable (V) region encoding gene sequence, or the complement thereof, and in each of the plurality of template oligonucleotide sequences V comprises a unique oligonucleotide sequence.

[0092] In some embodiments, J is a polynucleotide comprising at least 15-30, 31-60, 61-90, 91-120, or 120-150, and not more than 600, 500, 400, 300 or 200 contiguous nucleotides of an adaptive immune receptor joining (J) region encoding gene sequence, or the complement thereof, and in each of the plurality of template oligonucleotide sequences J comprises a unique oligonucleotide sequence.

[0093] U1 and U2 can be each either nothing or each comprise an oligonucleotide having, independently, a sequence that is selected from (i) a universal adaptor oligonucleotide sequence, and (ii) a sequencing platform-specific oligonucleotide sequence that is linked to and positioned 5' to the universal adaptor oligonucleotide sequence.

[0094] B1, B2 and B3 can be each either nothing or each comprise an oligonucleotide B that comprises a first and a second oligonucleotide barcode sequence of 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 or 1000 contiguous nucleotides (including all integer values therebetween), wherein in each of the plurality of template oligonucleotide sequences B comprises a unique oligonucleotide sequence in which (i) the first barcode sequence uniquely identifies the unique V oligonucleotide sequence of the template oligonucleotide and (ii) the second barcode sequence uniquely identifies the unique J oligonucleotide sequence of the template oligonucleotide.

[0095] X can be either nothing or comprises a restriction enzyme recognition site that comprises an oligonucleotide sequence that is absent from V, J, U1, U2, B1, B2 and B3.

[0096] The template compositions can also include random (R) sequences of length N. Random sequences R can include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or more random contiguous nucleotides in each template composition and can be unique to each template composition. There can be one or more R sequences in each synthetic template composition. The random sequences may be inserted in various sections between or within the components in the general formula I (5'-U1-B1-V-B2-X-B3-J-B4-U2-3') and be of various lengths in size. For example, the general formula can be 5'-U1-B1-V-R-B2-X-B3-J-B4-U2-3' and R can be at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 55, 60, 70, 80, 90, 100, 200, 300, 300, 500 or more contiguous nucleotides. The random sequence can be used to uniquely identify each specific paired V-J combination or to quantify or estimate the number of molecules in a sample. Each unique random sequence identifies a single molecule comprising a paired V-J combination.

[0097] Methods of the invention include using the template composition for determining non-uniform nucleic acid amplification potential among members of a set of oligonucleotide amplification primers that are capable of amplifying productively rearranged DNA encoding one or a plurality of adaptive immune receptors in a biological sample that comprises DNA from lymphoid cells of a subject. The method can include the steps of: (a) amplifying DNA of a template composition for standardizing amplification efficiency of an oligonucleotide primer set in a multiplex polymerase chain reaction (PCR) that comprises: (i) the template composition (I) described above, wherein each template oligonucleotide in the plurality of template oligonucleotides is present in a substantially equimolar amount; (ii) an oligonucleotide amplification primer set that is capable of amplifying productively rearranged DNA encoding one or a plurality of adaptive immune receptors in a biological sample that comprises DNA from lymphoid cells of a subject.

[0098] The primer set can include: (1) in substantially equimolar amounts, a plurality of V-segment oligonucleotide primers that are each independently capable of specifically hybridizing to at least one polynucleotide encoding an adaptive immune receptor V-region polypeptide or to the complement thereof, wherein each V-segment primer comprises a nucleotide sequence of at least 15 contiguous nucleotides that is complementary to at least one functional adaptive immune receptor V region-encoding gene segment and wherein the plurality of V-segment primers specifically hybridize to substantially all functional adaptive immune receptor V region-encoding gene segments that are present in the template composition, and (2) in substantially equimolar amounts, a plurality of J-segment oligonucleotide primers that are each independently capable of specifically hybridizing to at least one polynucleotide encoding an adaptive immune receptor J-region polypeptide or to the complement thereof, wherein each J-segment primer comprises a nucleotide sequence of at least 15 contiguous nucleotides that is complementary to at least one functional adaptive immune receptor J region-encoding gene segment and wherein the plurality of J-segment primers specifically hybridize to substantially all functional adaptive immune receptor J region-encoding gene segments that are present in the template composition.

[0099] The V-segment and J-segment oligonucleotide primers are capable of promoting amplification in said multiplex polymerase chain reaction (PCR) of substantially all template oligonucleotides in the template composition to produce a multiplicity of amplified template DNA molecules, said multiplicity of amplified template DNA molecules being sufficient to

quantify diversity of the template oligonucleotides in the template composition, and wherein each amplified template DNA molecule in the multiplicity of amplified template DNA molecules is less than 1000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80 or 70 nucleotides in length.

[00100] The method also includes steps of: (b) sequencing all or a sufficient portion of each of said multiplicity of amplified template DNA molecules to determine, for each unique template DNA molecule in said multiplicity of amplified template DNA molecules, (i) a template-specific oligonucleotide DNA sequence and (ii) a relative frequency of occurrence of the template oligonucleotide; and (c) comparing the relative frequency of occurrence for each unique template DNA sequence from said template composition, wherein a non-uniform frequency of occurrence for one or more template DNA sequences indicates non-uniform nucleic acid amplification potential among members of the set of oligonucleotide amplification primers. The amounts for each V-segment and J-segment primer set used in subsequent amplification assays can be adjusted to reduce amplification bias across the primer sets based on the relative frequency of occurrence for each unique template DNA sequence in the template composition.

[00101] Further description about bias control compositions and methods are provided in U.S. 2013/0253842 U.S. 9,150,905, WO 2015/134787, and WO 2013/169957, filed on May 8, 2013, PCT/US2013/045994 (WO/2013/188831), filed on 6/14/2013, which are incorporated by reference in their entireties.

Sequencing

[00102] Sequencing can be performed using any of a variety of available high throughput single molecule sequencing machines and systems. Illustrative sequence systems include sequence-by-synthesis systems, such as the Illumina Genome Analyzer and associated instruments (Illumina, Inc., San Diego, CA), Helicos Genetic Analysis System (Helicos BioSciences Corp., Cambridge, MA), Pacific Biosciences PacBio *RS* (Pacific Biosciences, Menlo Park, CA), or other systems having similar capabilities.

[00103] Sequencing is achieved using a set of sequencing oligonucleotides that hybridize to a defined region within the amplified molecules. In some embodiments, the sequencing oligonucleotides are designed such that the V- and J- encoding gene segments can be uniquely identified by the sequences that are generated, based on the present disclosure and in view of known adaptive immune receptor gene sequences that appear in publicly available databases. Exemplary sequencing oligonucleotides are described in Robins *et al.*, 2009 *Blood* 114, 4099;

Robins *et al.*, 2010 *Sci. Translat. Med.* 2:47ra64; Robins *et al.*, 2011 *J. Immunol. Meth.* doi:10.1016/j.jim.2011.09. 001; Sherwood *et al.* 2011 *Sci. Translat. Med.* 3:90ra61; U.S.2012/0058902 U.S.2010/0330571, WO/2010/151416, WO/2011/106738 (PCT/US2011/026373), U.S. 9,279,159; 9,181,590; and 9,181,591, 2013/0253842, and U.S. 2016/0024493 which are incorporated by reference in their entireties.

[00104] Techniques for sequencing nucleic acid known to those skilled in the art can be used in the methods of the provided invention. DNA sequencing techniques include classic dideoxy sequencing reactions (Sanger method) using labeled terminators or primers and gel separation in slab or capillary, sequencing by synthesis using reversibly terminated labeled nucleotides, pyrosequencing, 454 sequencing, allele specific hybridization to a library of labeled oligonucleotide probes, sequencing by synthesis using allele specific hybridization to a library of labeled clones that is followed by ligation, real time monitoring of the incorporation of labeled nucleotides during a polymerization step, polony sequencing, and SOLiD sequencing. Sequencing of the separated molecules has more recently been demonstrated by sequential or single extension reactions using polymerases or ligases as well as by single or sequential differential hybridizations with libraries of probes. These reactions have been performed on many clonal sequences in parallel including demonstrations in current commercial applications of over 100 million sequences in parallel. These sequencing approaches can thus be used to study the repertoire of T-cell receptor (TCR) and/or B-cell receptor (BCR).

[00105] The sequencing technique used in the methods of the invention can generate least 1000 reads per run, at least 10,000 reads per run, at least 100,000 reads per run, at least 500,000 reads per run, or at least 1,000,000 reads per run. The sequencing technique used in the methods of the invention can generate about 30 bp, about 40 bp, about 50 bp, about 60 bp, about 70 bp, about 80 bp, about 90 bp, about 100 bp, about 110, about 120 bp, about 150 bp, about 200 bp, about 250 bp, about 300 bp, about 350 bp, about 400 bp, about 450 bp, about 500 bp, about 550 bp, or about 600 bp per read. The sequencing technique used in the methods of the invention can generate at least 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, or 600 bp per read.

[00106] Example sequencing methods also include, but are not limited to, true single molecule sequencing (tSMS), 454 sequencing (Roche), SOLiD sequencing (Applied Biosystems), SOLEXA sequencing (Illumina), SMRT Sequencing (Pacific Biosciences), nanopore

sequencing, chemical-sensitive field effect transitor array sequencing, or sequencing by electron microscope, or other high throughput sequencing methods known to those of skill in the art.

[00107] In some embodiments, bias-controlled V segment and J segment gene primers are used to amplify rearranged V(D)J segments to produce a plurality of amplicons for high throughput sequencing at ~20x coverage. Coverage means the number of copies sequenced of each synthetic template.

Processing Sequence Data

[00108] As presently disclosed, there are provided methods for analyzing the sequences of the diverse pool of uniquely rearranged CDR3-encoding regions that are generated using the compositions and methods that are described herein. As described above, amplification bias can be corrected using bias control synthetic templates.

[00109] Corrections can also be made for PCR errors or sequencing errors. In some embodiments, the step of sequencing includes coalescing at least a plurality of sequence reads to form each clonotype. The step of coalescing is a process of combining sequence reads with error rates (for example, from sequencing and/or amplification errors) to produce clonotypes that are correct with a high degree of likelihood, such as with a 99% confidence level.

[00110] In some embodiments, the sequenced reads are filtered for those including CDR3 sequences. Sequencer data processing involves a series of steps to remove errors in the primary sequence of each read, and to compress the data. A complexity filter removes approximately 20% of the sequences that are misreads from the sequencer. Then, sequences were required to have a minimum of a six base match to both one of the TCR or IG J-regions and one of the TCR or IG V-regions. Applying the filter to the control lane containing phage sequence, on average only one sequence in 7-8 million passed these steps. Finally, a nearest neighbor algorithm is used to collapse the data into unique sequences by merging closely related sequences, in order to remove both PCR error and sequencing error.

[00111] In some embodiments, methods are used for estimating the true distribution of specific clonotypes (*e.g.*, a TCR or IG having a uniquely rearranged CDR3 sequence) in blood or in a sample derived from other peripheral tissue or bodily fluid. For example, the ratio of sequences in the PCR product can be derived by working backward from the sequence data before estimating the true distribution of clonotypes (*e.g.*, unique clonal sequences) in the blood. For each sequence observed a given number of times in the data herein, the probability that that

sequence was sampled from a particular size PCR pool is estimated. Because the CDR3 regions sequenced are sampled randomly from a massive pool of PCR products, the number of observations for each sequence are drawn from Poisson distributions. The Poisson parameters are quantized according to the number of T cell genomes that provided the template for PCR. A simple Poisson mixture model both estimates these parameters and places a pairwise probability for each sequence being drawn from each distribution. This is an expectation maximization method, which reconstructs the abundances of each sequence that was drawn from the blood.

[00112] In some embodiments, to estimate the total number of unique adaptive immune receptor CDR3 sequences that are present in a sample, a computational approach employing the "unseen species" formula may be employed (Efron and Thisted, 1976 *Biometrika 63*, 435-447). This approach estimates the number of unique species (*e.g.*, unique adaptive immune receptor sequences) in a large, complex population (*e.g.*, a population of adaptive immune cells such as T cells or B cells), based on the number of unique species observed in a random, finite sample from a population (Fisher et al., 1943 *J. Anim. Ecol.* 12:42-58; Ionita-Laza et al., 2009 *Proc. Nat. Acad. Sci. USA* 106:5008). The method employs an expression that predicts the number of "new" species that would be observed if a second random, finite and identically sized sample from the same population were to be analyzed. "Unseen" species refers to the number of new adaptive immune receptor sequences that would be detected if the steps of amplifying adaptive immune receptor-encoding sequences in a sample and determining the frequency of occurrence of each unique sequence in the sample were repeated an infinite number of times. By way of non-limiting theory, it is operationally assumed for purposes of these estimates that adaptive immune cells (*e.g.*, T cells, B cells) circulate freely in the anatomical compartment of the subject that is the source of the sample from which diversity is being estimated (*e.g.*, blood, lymph, etc.).

[00113] To apply this formula, unique adaptive immune receptors (*e.g.*, TCRβ, TCRα, TCRγ, TCRδ, IgH) clonotypes takes the place of species. The mathematical solution provides that for *S*, the total number of adaptive immune receptors having unique sequences (*e.g.*, TCRβ, TCRγ, IgH "species" or clonotypes, which may in certain embodiments be unique CDR3 sequences), a sequencing experiment observes x_s copies of sequence s. For all of the unobserved clonotypes, x_s equals *0*, and each TCR or Ig clonotype is "captured" in the course of obtaining a random sample (*e.g.*, a blood draw) according to a Poisson process with parameter λ_s. The number of T or B cell

genomes sequenced in the first measurement is defined as *1*, and the number of T or B cell genomes sequenced in the second measurement is defined as *t*.

[00114] Because there are a large number of unique sequences, an integral is used instead of a sum. If $G(\lambda)$ is the empirical distribution function of the parameters $\lambda_1, \ldots, \lambda_S$, and n_x is the number of clonotypes (*e.g.*, unique TCR or Ig sequences, or unique CDR3 sequences) observed exactly x times, then the total number of clonotypes, *i.e.*, the measurement of diversity *E*, is given by the following formula (I):

$$E(n_x) = S\int_0^\infty \left(\frac{e^{-\lambda}\lambda^x}{x!}\right) dG(\lambda) \quad \text{(I)}$$

[00115] Accordingly, formula (I) may be used to estimate the total diversity of species in the entire source from which the identically sized samples are taken. Without wishing to be bound by theory, the principle is that the sampled number of clonotypes in a sample of any given size contains sufficient information to estimate the underlying distribution of clonotypes in the whole source. The value for $\Delta(t)$, the number of *new* clonotypes observed in a second measurement, may be determined, preferably using the following equation (II):

$$\Delta(t) = \sum_x E(n_x)_{msmt1+msmt2} - \sum_x E(n_x)_{msmt1} = S\int_0^\infty e^{-\lambda}\left(1-e^{-\lambda t}\right) dG(\lambda) \quad \text{(II)}$$

in which *msmt1* and *msmt2* are the number of clonotypes from measurements 1 and 2, respectively. Taylor expansion of $1\text{-}e^{-\lambda t}$ and substitution into the expression for $\Delta(t)$ yields:

$$\Delta(t) = E(x_1)t - E(x_2)t^2 + E(x_3)t^3 - \ldots \text{(III)}$$

which can be approximated by replacing the expectations ($E(n_x)$) with the actual numbers sequences observed exactly *x* times in the first sample measurement. The expression for $\Delta(t)$ oscillates widely as *t* goes to infinity, so $\Delta(t)$ is regularized to produce a lower bound for $\Delta(\infty)$, for example, using the Euler transformation (Efron et al., 1976 *Biometrika* 63:435).

[00116] In one example, using the numbers observed in a first measurement of TCRβ sequence diversity in a blood sample, this formula (II) predicted that $1.6*10^5$ new unique sequences should be observed in a second measurement. The actual value of the second measurement was $1.8*10^5$ new TCRβ sequences, which suggested according to non-limiting

theory that the prediction provided a valid lower bound on total TCRβ sequence diversity in the subject from whom the sample was drawn.

[00117] Additional description about the unseen species model and processing sequence data are described in Robins *et al.*, 2009 *Blood* 114, 4099; Robins *et al.*, 2010 *Sci. Translat. Med.* 2:47ra64; Robins *et al.*, 2011 *J. Immunol. Meth.* doi:10.1016/j.jim.2011.09. 001; Sherwood *et al.* 2011 *Sci. Translat. Med.* 3:90ra61; U.S.2012/0058902; U.S.2010/0330571; WO/2011/106738 (PCT/US2011/026373); WO2012/027503 (PCT/US2011/049012); U.S. 9,279,159; 9,181,590; 9,181,591; and 2013/0253842, which are incorporated by reference in their entireties.

[00118] In certain embodiments, after correcting for sequencing errors via a clustering algorithm, CDR3 segments are annotated according to the International ImMunoGeneTics collaboration. *See* Lefranc, M.-P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., Wu, Y., Gemrot, E., Brochet, X., Lane, J., Regnier, L., Ehrenmann, F., Lefranc, G. and Duroux, P. IMGT®, the International ImMunoGeneTics Information system®. Nucl. Acids Res, 37, D1006-D1012 (2009); doi:10.1093/nar/gkn838. PMID: 18978023; Lefranc, M.-P., IMGT, the International ImMunoGeneTics Information System. Cold Spring Harb Protoc. 2011 Jun 1. 2011(6). pii: pdb.top115. doi: 10.1101/pdb.top115. PMID: 21632786.

Clonotype Determination

[00119] In some embodiments, a clonotype is defined when at least two identical sequence reads are obtained. Briefly, after exclusion of low quality reads, sequence data can be analyzed to determine the clonotype sequences including mapping to germline V and J consensus sequences. In one embodiment, the sample index sequences were used to identify which of the sequences originate from which of the pooled samples. Sequences whose index are not a perfect match to one of the indices used in a specific run are excluded. Next the forward read is used to map the J segment. Since all the sequences started from the same position of the J segments, all the J segments started at a predefined sequencing position. The first 25 bp of the J segments are used to map the J segment. Any read with more than 5 high quality mismatches to the known J segments are excluded from further analysis.

[00120] After J segment identification, V segments are mapped. The reverse read is used for this purpose. First, the V primer is mapped and excluded. Thereafter, the next 70 bases of the reverse read are mapped to the known V segments. Reads that do not map to J and V segments are excluded. The next step in mapping involves identifying the frame that related the forward

and reverse reads and this allows a continuous sequence from J to V to be constructed. This is done using the last 15 bases of the forward read which are reliably within the V segment regardless of NDN length. While these bases could be of relatively lower sequence quality as they are at the terminal end of a long read, they can be used to map within a single identified V segment in order to identify the position at which the two reads could be joined. Finally, the known V and J sequences to which the reads map are used to identify the point in the forward read at which the sequences at the junctions diverged from these mapped segments.

[00121] Other methods known to one of skill in the art can be used to identify and remove sequence errors and cluster sequences.

Multimer Binding Assay

[00122] T cells recognize antigens that are displayed by major histocompatibility complex (MHC) on cell surfaces. These antigens may be derived from pathogens that replicate within cells, such as viruses or intracellular bacteria, or from pathogens or their products that cells take up by endocytosis from the extracellular fluid or normal or mutated human proteins. Infected cells display on their surface peptide fragments derived from the pathogens' proteins and can thus be detected by T cells. These foreign, naturally occurring, or mutated peptides are delivered to the cell surface by specialized host-cell glycoproteins, the MHC molecules. The display of a peptide at the cell surface by the MHC molecules is referred to as antigen presentation.

[00123] There are two main classes of MHC molecules: MHC Class I and MHC class II. In humans, these genes are called human leukocyte antigen (HLA) genes. Generally, antigens presented by class I MHC molecules are recognized by TCRs from CD8+ T cells, and antigens presented by class II MHC molecules are generally recognized by TCRs from CD4+ T cells.

[00124] An antigen of interest is chosen for an *in vitro* multimer binding assay to determine the top binding TCRs for that antigen. In some embodiments, the antigen is presented as a peptide in a multimer complex comprising multiple MHC:antigenic peptide complexes. In one embodiment, the multimer is a tetramer, pentamer, dextamer, etc.

[00125] Example antigens can be derived from proteins of interest obtained from pathogens, such as viruses, bacteria, fungi, parasite, from a vaccine, normal human proteins, or mutated human proteins. Examples include peptides derived from hepatitis A, hepatitis B, hepatitis C, human papilloma virus (HPV), human immunodeficiency virus (HIV), herpes simplex virus, or other chronic viruses. In some embodiments, peptides are derived from human proteins that are

overexpressed in cancer or disease. One example is the WT1 gene that is overexpressed in several human cancers.

[00126] In one embodiment, a tetramer is used for testing TCR and HLA:antigen binding. The tetramer itself consists of multiple bound MHC molecules. The need for an MHC tetramer arises from the high dissociation rate of MHC monomers, making monomers difficult to use as a detection strategy. Tetramers however, can bind multiple MHCs at a time to a T cell and so increase the binding avidity and circumvent the problem of dissociation.

[00127] The centerpiece of each tetramer is a streptavidin complex. Streptavidin is a molecule that forms homotetramer complexes, with each monomer having an unusually high affinity for biotin. Exploiting these facts, scientists have bioengineered *E. coli* to produce soluble MHC molecules with a biotinylation protein domain, meaning a part of the MHC can be replaced by covalently bound biotin (via BirA enzyme activity). The biotinylated domain then allows for up to 4 MHC to bind to a fluorescently tagged streptavidin complex with high affinity. The MHC tetramers are typically also labeled with a detectable flurochrome, for example fluorescein isothiocyanate (FITC), phycoerythrin (PE), allophycocyanin (APC) or another fluorophore detectable by flow cytometry. The resulting fluorophore-labeled MHC tetramers are then loaded with the antigenic peptide of interest, forming tetramer:peptide complexes, i.e., an antigen loaded tetramer. The resulting antigen loaded tetramer is then added to a sample of cells and binds to T cells that are specific for both the MHC type and peptide being used in the tetramer (i.e., antigen-loaded MHC class II tetramers typically bind to $CD4^+$ T cells whereas antigen-loaded MHC class I tetramers typically bind to $CD8^+$ T cells). Once the tetramers are bound the sample is washed to remove non-bound tetramers and ligands and the washed sample is then analyzed and/or sorted by flow cytometry. The fluorophore on any bound tetramers can be excited to give a signal, indicating that the tetramer is bound to a T cell, and thus, the bound T cell is specific for the peptide antigen of interest. Ultimately, a signal means that there exists some cell-mediated immune response to the pathogen from which the antigenic peptide is derived, and the strength of the signal gives the strength of the immune response.

[00128] In this manner, T cell receptors that have the highest binding affinity to the MHC:antigenic peptide are enriched in the sample.

[00129] In the context of MHC class I tetramers, the CD8 coreceptor on T cells also interacts with the MHC multimer, contributing to the affinity of the TCR/peptide-MHC interaction. TCRs

that have an exceptionally high affinity for a target peptide-MHC can bind peptide-MHC multimers independent of any interactions with CD8. Thus, peptide-MHC multimers can be generated that contain MHC mutations that abolish CD8 interactions. Using these CD8-independent tetramers, T cells expressing the highest affinity CD8-independent TCRs can be identified and ranked by relative affinity.

[00130] TCR affinity is determined by the on-rate and off-rate the TCR/peptide-MHC interaction. Flow cytometry-based technologies exist that allow the discrimination of T cells that express high affinity TCRs due to fast on-rates or slow off-rates. These technologies can be used in conjunction with the described technology to identify high affinity TCRs based on these parameters.

[00131] Using the multimer binding assay, the T cell receptors that have the highest binding affinity or avidity to a particular antigen can be identified. Binding affinity can be measured by strength of the signal. In some embodiments, the top 100 T cell clones that bind a particular MHC:antigenic peptide are identified as top binders, and can be further tested as therapeutics, according to the methods of the invention. In other embodiments, the top 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 T cell clones that bind a particular MHC:antigenic peptide are identified as top binders, and can be further tested as therapeutics, according to the methods of the invention. In yet other embodiments, the top 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, or 20% of T cell clones that bind a particular MHC:antigenic peptide are identified as top binders, and can be further tested as therapeutics, according to the methods of the invention. In another embodiment, the top 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400 or 500 ranked T cell clones that bind a particular MHC:antigenic peptide are identified as top binders, and can be further tested as therapeutics, according to the methods of the invention.

[00132] In other embodiments, T cell clones with a lower binding affinity (not the best or top binders) to the particular antigen are chosen as clones for use as possible therapeutics.

Clonotype Abundance / Frequency Determination

[00133] Following clonotype determination, relative frequencies of the clonotypes can be analyzed from the samples. For example, the relative frequencies of clonotypes can be determined from an enriched population of clones and a non-enriched population of clones and then the relative frequencies can be compared.

[00134] In one embodiment, the abundance or frequency of clones is measured in a sample obtained from one or more subjects. The sample is then enriched with a multimer comprising MHC:antigenic peptide complexes. The binding affinity (or avidity) of clones (or clonotypes, TCRs) to the MHC:antigenic peptide is measured.

[00135] The pre-enrichment and post-enrichment samples are immunosequenced, as described above. The relative abundance (e.g., frequency of occurrence, number, or ratio) of each clonotype out of the total population of T cells is determined prior to and after enrichment. The change in relative abundance of a particular clonotype before and after an enrichment is determined. For example, if a clonotype makes up .01% of the T cell population prior to sorting, and then the clonotype represents 1% of the T cell population after sorting, this amounts to a 100x enrichment of that particular clonotype, even though the clonotype is still only at 1% of the entire T cell population. In some embodiments, the sample is subjected to more than one round of enrichment/sorting. In some embodiments, the sample is enriched/sorted at least twice, or at least three times or at least 4 times or at least 5 times.

[00136] In some embodiments, the sample will be divided and stained with decreasing concentrations of tetramer before enrichment and identification of top binders to the particular antigen. The relative abundance (e.g., frequency of occurrence, number, or ratio) of each clonotype out of the total population of T cells is determined prior to and after enrichment for each multimer-dilution stained sample and relative enrichment for each clonotype compared at decreasing multimer dilutions.

[00137] In some embodiments, the clones are ranked according to their binding affinity for a particular MHC:peptide (HLA:peptide). In some embodiments, the top ranked clones are identified as the best binders and can be developed as possible therapeutics.

[00138] T cell clones with the greatest change in relative abundance between the pre-enriched and post-enriched samples can be identified as optimally binding TCRs for a particular antigenic target and can be used for development of therapeutics. In other embodiments, a lower binding clone can be chosen in cases where the antigenic peptide of interest is derived from a normal human protein that is over-expressed in cancer or disease.

Pairing Chains of Adaptive Immune Receptors

[00139] In some embodiments, a pairing assay can be performed to match TCR nucleic acid sequences that encode polypeptide pairs of a TCR. Description about methods for determining

pairs of TCR and/or Ig heterodimers are those described in PCT/US2014/030859, filed on March 17, 2014 (WO 2014/145992), PCT/US2013/045994, filed on June 14, 2013 (WO 2013/188831), PCT/US13/028942, filed on March 4, 2012 (WO 2013/134162), US 2014/0322716 filed on July 7, 2014, US 2015/0299786 filed on June 5, 2015, and US 2016/0024493, filed on September 15, 2015 which are incorporated by reference in their entireties. Briefly, the pairing assay is done by distributing a plurality of T cells among a plurality of contains and generating a library of amplicons from the DNA (either genomic DNA or cDNA reverse transcribed from RNA) of the T cells in the plurality of containers by multiplex PCR. The library of amplicons comprises a plurality of first adaptive immune receptor amplicons encoding a first polypeptide comprising a unique variable (V) region encoding sequence and a unique joining (J) region encoding sequence of one TCR chain (e.g. a TCRβ chain), at least one barcode, at least one universal adapter sequence and at least one sequencing platform tag. The library of amplicons similarly contains a plurality of second adaptive immune receptor amplicons encoding a first polypeptide comprising a unique variable (V) region encoding sequence and a unique joining (J) region encoding sequence of one TCR chain (e.g. a TCRα chain), at least one barcode, at least one universal adapter sequence and at least one sequencing platform tag. In some embodiments, the plurality of first and second amplicons is subject to high throughput sequencing to obtain a dataset of a plurality of first and second amplicons. From that data, a container occupancy pattern for each unique first and second amplicons is determined and a for each possible pairing of unique first and second immune receptor amplicons sequence a statistical probability of observing the container occupancy pattern is calculated and the first and second amplicons are paired based on the statistical probability and a false discovery rate. Alternative methods for pairing can also be used. Suitable alternatives include pairing first and second chains of adaptive immune receptors by frequency. This frequency based method can be performed on an unsorted or sorted (i.e. distributed among several discrete containers) sample of T cells. Another alternative approach to pairing comprises linking the two adaptive immune receptor chains (e.g. TCRα and TCRβ) by polymerase cycling amplification, for example. The pairing assay can be performed either prior to or after enrichment of the sample using the multimer assay, as described above.

[00140] The pairing assay allows for identification of cognate pairs of first and second rearranged nucleic acid sequences encoding first and second polypeptides of adaptive immune receptor heterodimers. For example, the pairing assay allows pairing of TCRα and TCRβ

sequences that form TCRαβ heterodimers (or pairing of TCRγ and TCRδ sequences, for example).

[00141] A pair of nucleic acid sequences encoding a cognate TCRαβ receptor and enriched after exposure to a HLA presented antigen can be identified as a therapeutic.

Therapeutic Uses of Identified TCRs

[00142] Once the antigen-specific TCR sequences have been determined, those sequences can be used to construct therapeutic molecules/cells comprising the sequences or a portion thereof. In some embodiments, expression vectors are constructed which may be transduced into autologous or syngeneic T cells which may then be delivered to a patient in need thereof. In some embodiments, the T cell receptor sequence may be used as a basis to construct a chimeric receptor, which may then be transduced into autologous or syngeneic cells and delivered to a patient in need thereof. In certain embodiments, once the TCR sequence has been determined, those T cells harboring this particular TCR may be sorted and expanded *in vitro* and those expanded T cells may then be delivered as a therapeutic. In some embodiments, once identified, the TCR may be used as a soluble TCR for detecting antigen-specific cells. In some embodiments, once identified, the TCR may be used to construct bi-specific soluble TCRs that can bind to cross link CD3 on the surface of T cells.

EXAMPLES

[00143] Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

[00144] The practice of the present invention will employ, unless otherwise indicated, conventional methods of protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company,

1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols A and B(1992).

Example 1: Method of Identifying Human Compatible T Cell Receptors Against An Antigenic Target

[00145] *Generation and sorting ot T cell lines*

[00146] **Generation of T cell lines:** Antigen-specific T cell lines were generated from two donors that were specific for the WT1 peptide $WT1_{37-45}$ (VLDFAPPGA). A total of 10 lines were generated from each donor for a total of 20 different T cell lines. The lines were generated as described in Ho, et al (2006) J. Immunol Methods, 310 (1-2), 40-52. Briefly, dendritic cells (DCs) were derived from the plastic adherent fraction of peripheral blood mononuclear cells (PBMCs) after culture for 2 days (days -2 to day 0) in media supplemented with GM-CSF and IL-4. On day -1, TNF-α, IL-1β, IL-6 and PGE_2 were added. On day 0, DCs were harvested, washed and pulsed with peptide ($WT1_{37-45}$). $CD8^+$ T cells were isolated from PBMCs using anti-CD8 microbeads and stimulated with peptide-pulsed DCs in the presence of IL-21. Cells were restimulated twice between days 10 and 14, with the plastic adherent fraction of irradiated autologous PBMCs as antigen presenting cells, pulsed with the relevant peptide.

[00147] **HLA-A2/peptide tetramer sorting**: The 20 cultured T cell lines were combined prior to cell sorting and stained with either a) an optimized concentration of $WT1_{37-45}$-loaded tetramer determined empirically to achieve optimal resolution by flow cytometry; b) a 1:100 dilution of optimized tetramer; or c) an optimized concentration of $WT1_{37-45}$-loaded tetramer with mutations introduced into the HLA-A2 protein (D227K, T228A), that interfere with CD8 binding and therefore only bind TCRs with sufficiently high affinity to associate with MHC in a CD8-independent manner. A population of total tetramer positive cells (all cells staining tetramer above background levels) was sorted from sample a) for pairing analysis. A more restrictive sort (2% of total) was done for each tetramer-stained fraction for frequency analysis. 2.5-5 x 10^5 cells were sorted for TCRβ sequencing per sample (Figure 1).

[00148] *Determination of relative fold-enrichment*

[00149] TCRβ-seq was also performed on ~2 million T cells from the pre-sort pooled sample, and this provided a baseline frequency for each TCRβ clonotype. The frequency of each clonotype in the restrictive sort populations was also determined by TCRβ-seq, and the fold-increase following cell sorting was determined by dividing the post-sort frequency by the

baseline frequency. Since some T cell clones with lower affinity might be highly enriched in the post-sort fraction due to higher surface expression of the TCR, we included a sort using a 1:100 dilution of tetramer in order to select against cells with a higher equilibrium binding constant (lower affinity) but compensating high level of TCR surface expression. Likewise, CD8-independent tetramer binding is also known to be a characteristic of very high affinity TCRs, with the affinity threshold for CD8-independent tetramer binding estimated to be ~5μM (Holler & Kranz, (2003) Immunity, 18(2):255-264). In order to determine candidate high affinity TCRs to advance for gene synthesis, clonotypes with high relative enrichment in the restrictive sort gates compared to baseline, and that were also more highly enriched in the most restrictive sorts (1:100 tetramer and CD8-independent tetramer sorts) were selected. Several less enriched clonotypes, or clonotypes that showed lower enrichment in the most restrictive sorts were also included for comparison (Figure 2) and a total of 9 TCRs were selected for further synthesis and testing (Figure 3).

[00150] *TCR gene synthesis and generation of Lentiviral constructs:*

[00151] The TCR expression constructs were generated consisting of codon-optimized (GeneArt/Life technologies) TCRα and TCRβ genes derived from candidate HLA-A2-restricted $CD8^+$ T cell clones predicted to have a high affinity for the WT1 peptide $WT1_{37-45}$ (VLDFAPPGA). The TCRα and TCRβ genes were linked by a 2A element from the porcine teschovirus (P2A) to ensure coordinated expression under the control of the murine stem cell virus (MSCV) U3 promoter. The constant domains of each TCRα and TCRβ chain were modified to express complementary cysteine residues at positions 48 (Thr to Cys) and 57 (Ser to Cys), respectively, in order to promote inter-chain pairing of the TCR chains and to discourage mispairing of the introduced TCRs with endogenous TCR chains. Each TCR expression vector consists of the TCR expression construct ligated into the pRRLSIN.cPPT.MSCV/GFP.WPRE lentiviral vector between the AscI and SalI restriction sites, replacing GFP. The pRRLSIN.cPPT.MSCV/GFP.WPRE plasmid is a third-generation, self-inactivating lentiviral vector and was a gift from Richard Morgan.

[00152] Lentivirus was produced from each construct, and used to stably transduce a variant of the Jurkat T cell line that lacks endogenous TCRα and TCRβ gene expression, such that CD3 surface staining correlates with transgenic TCR surface expression. A previously characterized TCR (P37) that is the highest affinity TCR specific for WT137-45 that was identified by

conventional methods was also included for comparison. Each of the selected TCRs were found to be antigen specific by tetramer staining, and at least three of the selected TCRs were higher affinity than our most avid WT137-45 specific TCR. As predicted, the highest affinity TCRs were found to be increasingly enriched in the more stringent sort conditions (Figure 4).

Example 2: Comparison of Current Method to Conventional Techniques

[00153] In order to compare the methodology described herein to the current state of the art, conventional techniques were used to identify the highest affinity TCR from the same set of 20 T cell lines made in Example 1, as follows: Each line was stained separately with tetramer and analyzed by flow cytometry. The lines containing an identifiable population of cells with tetramer staining were subject to limiting dilution cloning and single clones that had the highest tetramer staining were lysed and the antigen-specific TCRα and TCRβ chains were identified by RACE PCR. Using this technique TCR #2 was identified as the highest affinity clone. However, this clone as a significantly lower apparent affinity compared to TCR#1, which was the highest affinity clone identified by the methodology described here (Figure 5), illustrating the superiority of the current technology.

[00154] While the invention has been particularly shown and described with reference to a preferred embodiment and various alternate embodiments, it will be understood by persons skilled in the relevant art that various changes in form and details can be made therein without departing from the spirit and scope of the invention.

[00155] All references, issued patents and patent applications cited within the body of the instant specification are hereby incorporated by reference in their entirety, for all purposes.

CLAIMS

What is claimed is:

1. A method of identifying an antigen-specific T cell receptor cognate pair, comprising:

dividing one or more samples comprising T cells into a first subset and a second subset;

performing immunosequencing of rearranged nucleic acid molecules extracted from the first subset of cells to obtain a first plurality of unique sequence reads;

determining based on the first plurality of unique sequence reads a relative abundance of each unique T cell receptor (TCR) sequence out of a total number of T cells in the first subset;

enriching the second subset of cells with multimer molecules comprising an HLA-presented antigen to identify a population of antigen-specific T cells that bind the multimer;

performing immunosequencing of rearranged nucleic acid molecules extracted from said sorted population of antigen-specific T cells to obtain a second plurality of unique sequence reads;

determining based on the second plurality of unique sequence reads a relative abundance of each unique TCR sequence out of a total number of T cells in the second subset;

determining a relative change in abundance of each antigen-specific TCR sequence based on the relative abundances of the antigen-specific TCR sequence in the first subset and the second subset;

identifying at least one antigen-specific TCR sequence based on its determined relative change in abundance; and

identifying a second TCR sequence that pairs with the at least one antigen-specific TCR sequence to form an antigen-specific TCR cognate pair.

2. The method of claim 1, wherein the enriching is performed by flow cytometry.

3. The method of claim 1, wherein the one or more samples are blood samples.

4. The method of claim 1, wherein the one or more samples are tissue samples.

5. The method of claim 1, further comprising identifying at least one antigen-specific TCR cognate pair sequence is based on its determined relative change in abundance.

6. The method of claim 5, wherein the relative change in abundance is calculated by ranking each of the antigen-specific TCR sequences based on its binding affinity for said HLA-presented antigen.

7. The method of claim 1, wherein said TCR sequence is a TCRB, TCRA, TCRG, or TCRD sequence.

8. A method of identifying an antigen-specific T cell as a therapeutic, comprising:

receiving one or more samples comprising T cells;

dividing the one or more samples into a first subset and a second subset;

performing immunosequencing of rearranged nucleic acid molecules extracted from the first subset of cells to obtain a first plurality of unique sequence reads;

determining based on the first plurality of unique sequence reads a relative abundance of each unique T cell receptor (TCR) sequence out of a total number of T cells in the first subset;

enriching the second subset of cells with multimer molecules comprising an HLA-presented antigen to identify a population of antigen-specific T cells that bind the multimer;

performing immunosequencing of rearranged nucleic acid molecules extracted from said sorted population of antigen-specific T cells to obtain a second plurality of unique sequence reads;

determining based on the second plurality of unique sequence reads a relative abundance of each unique TCR sequence out of a total number of T cells in the second subset;

determining a relative change in abundance of each antigen-specific TCR sequence based on the relative abundances of the antigen-specific TCR sequence in the first subset and the second subset; and

identifying an antigen-specific TCR sequence based on its relative change in abundance as a clone for therapeutic use for said HLA-presented antigen.

9. The method of claim 8, wherein the enriching is performed by flow cytometry.

10. The method of claim 8, wherein the one or more samples are blood samples.

11. The method of claim 8, wherein the one or more samples are tissue samples.

12. The method of claim 8, wherein identifying an antigen-specific TCR sequence based on its relative change in abundance as a clone for therapeutic use for said HLA-presented antigen comprises ranking each of the antigen-specific TCR sequences based on its binding affinity for said HLA-presented antigen.

13. The method of claim 8, further comprising pairing the antigen-specific TCR sequence with a second TCR sequence that forms its cognate pair in the T cell.

14. A method of identifying an antigen-specific T cell as a therapeutic, comprising:

dividing one or more samples comprising T cells into a first subset and a second subset;

performing immunosequencing of rearranged nucleic acid molecules extracted from the first subset of cells to obtain a first plurality of unique sequence reads;

determining based on the first plurality of unique sequence reads a relative abundance of each unique T cell receptor (TCR) sequence out of a total number of T cells in the first subset;

enriching the second subset of cells with multimer molecules comprising an HLA-presented antigen to identify a population of antigen-specific T cells that bind the multimer;

performing immunosequencing of rearranged nucleic acid molecules extracted from said sorted population of antigen-specific T cells to obtain a second plurality of unique sequence reads;

determining based on the second plurality of unique sequence reads a relative abundance of each unique TCR sequence out of a total number of T cells in the second subset;

determining a relative change in abundance of each antigen-specific TCR sequence based on the relative abundances of the antigen-specific TCR sequence in the first subset and the second subset; and

identifying the top antigen-specific TCR sequences having the greatest relative change in abundance as therapeutics for said HLA-presented antigen.

15. The method of claim 14, wherein the top antigen-specific TCR sequences comprise the top 100 ranked clones in the sample.

16. The method of claim 14, wherein the top antigen-specific TCR sequences comprise the top 50 ranked clones in the sample.

17. The method of claim 14, wherein the top antigen-specific TCR sequences comprise the top 10 ranked clones in the sample.

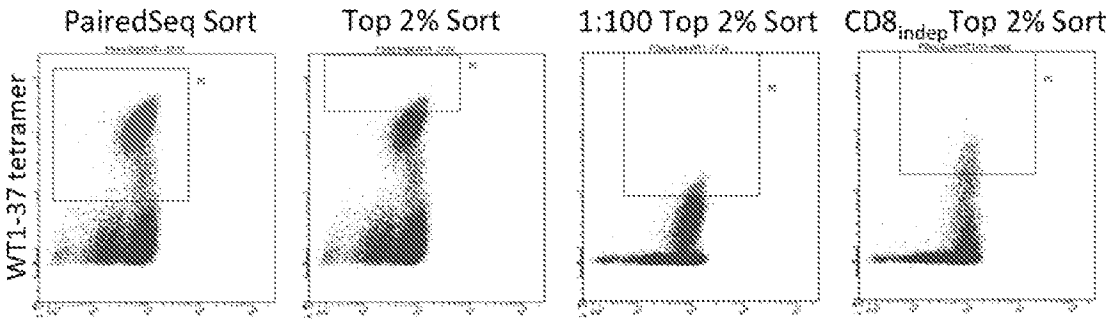

PairedSeq Sort
Top 2% Sort
1:100 Top 2% Sort
CD8indep Top 2% Sort
WT1-37 tetramer

Fig.1.

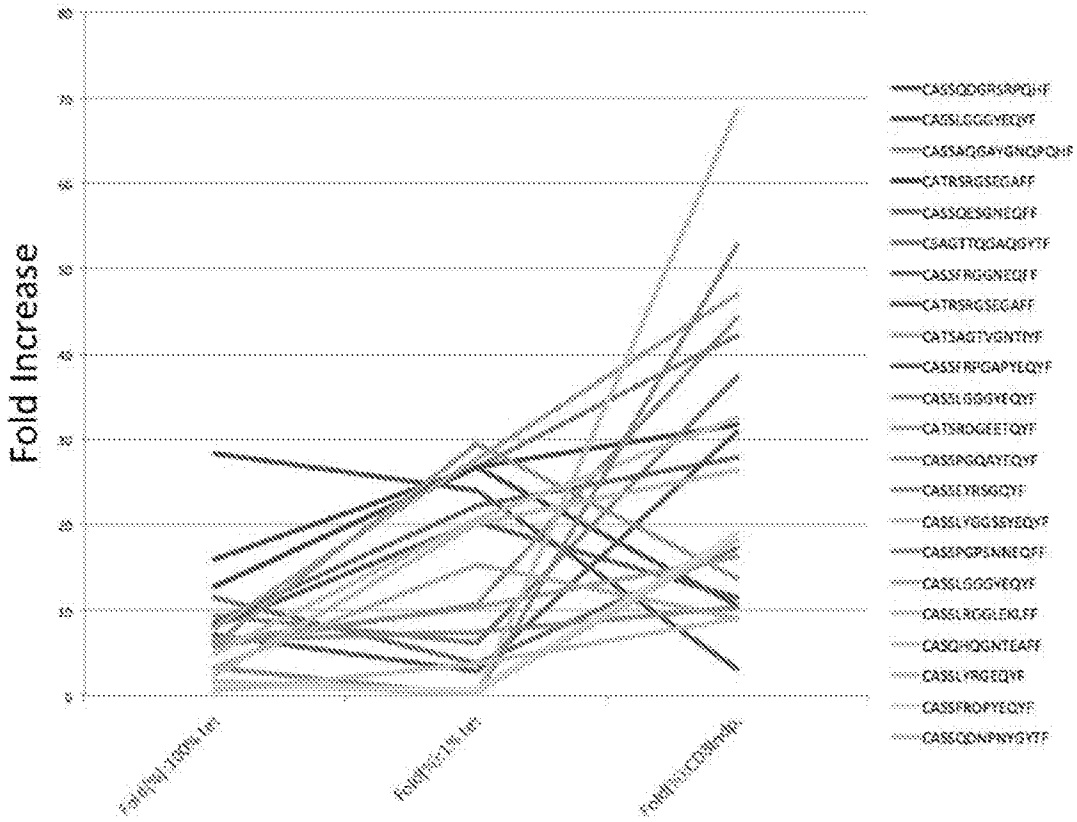

Fold Increase
CASSQDGRGRPQHF
CASSLGGGYEQYF
CASSAQGAYGNQPQHF
CATRSRGSEEAFF
CASSQESGNEQFF
CSAGTTQGAQGYTF
CASSFRGGNEQFF
CATRSRGSEEAFF
CATSAGTVGNTIYF
CASSFRPGAPYEQYF
CASSLGGGYEQYF
CATSRDGEETQYF
CASSPGQAYTQYF
CASSEYRSGDTF
CASSLVGGSEYEQYF
CASSPGPSNNEQFF
CASSLGGGYEQYF
CASSLREGLEKLFF
CASQHQGNTEAFF
CASSLYRGEQYF
CASSFRDPYEQYF
CASSQDNPNYGYTF

Fig. 2.

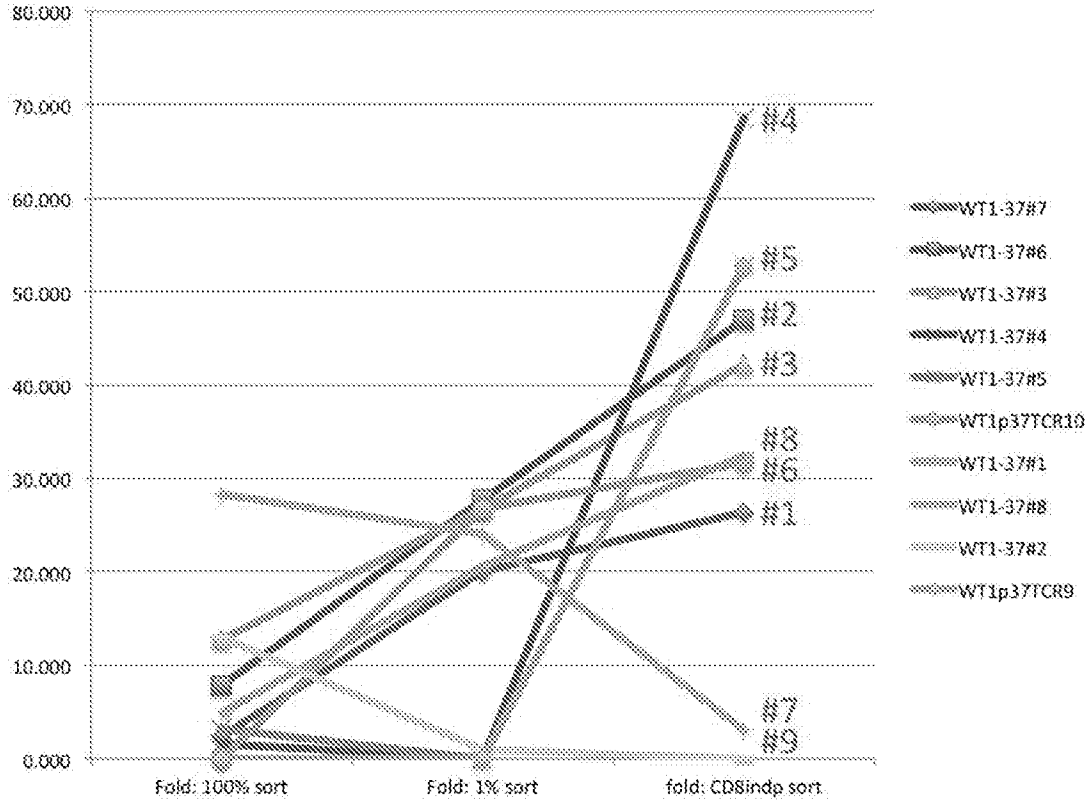

80.000
70.000
60.000
50.000
40.000
30.000
20.000
10.000
0.000
Fold: 100% sort
Fold: 1% sort
fold: CD8indp sort
#4
#5
#2
#3
#8
#6
#1
#7
#9
WT1-37#7
WT1-37#6
WT1-37#3
WT1-37#4
WT1-37#5
WT1p37TCR10
WT1-37#1
WT1-37#8
WT1-37#2
WT1p37TCR9

Fig. 3.

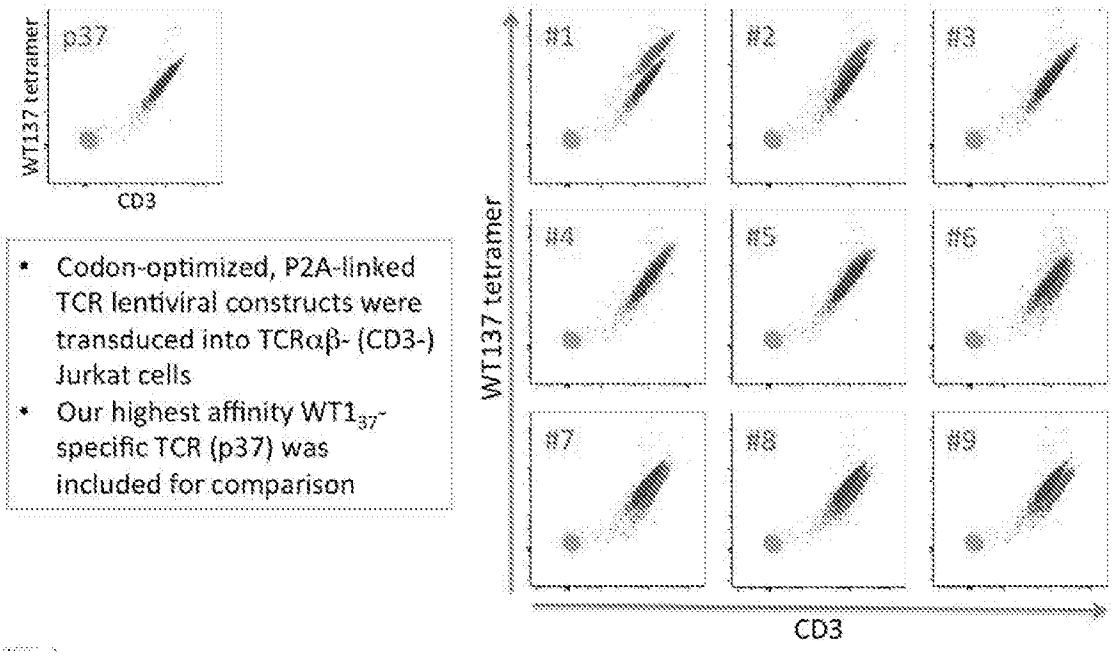

p37
WT137 tetramer
CD3
• Codon-optimized, P2A-linked TCR lentiviral constructs were transduced into TCRαβ- (CD3-) Jurkat cells
• Our highest affinity $WT1_{37}$-specific TCR (p37) was included for comparison
#1
#2
#3
#4
#5
#6
#7
#8
#9
WT137 tetramer
CD3

Fig. 4.

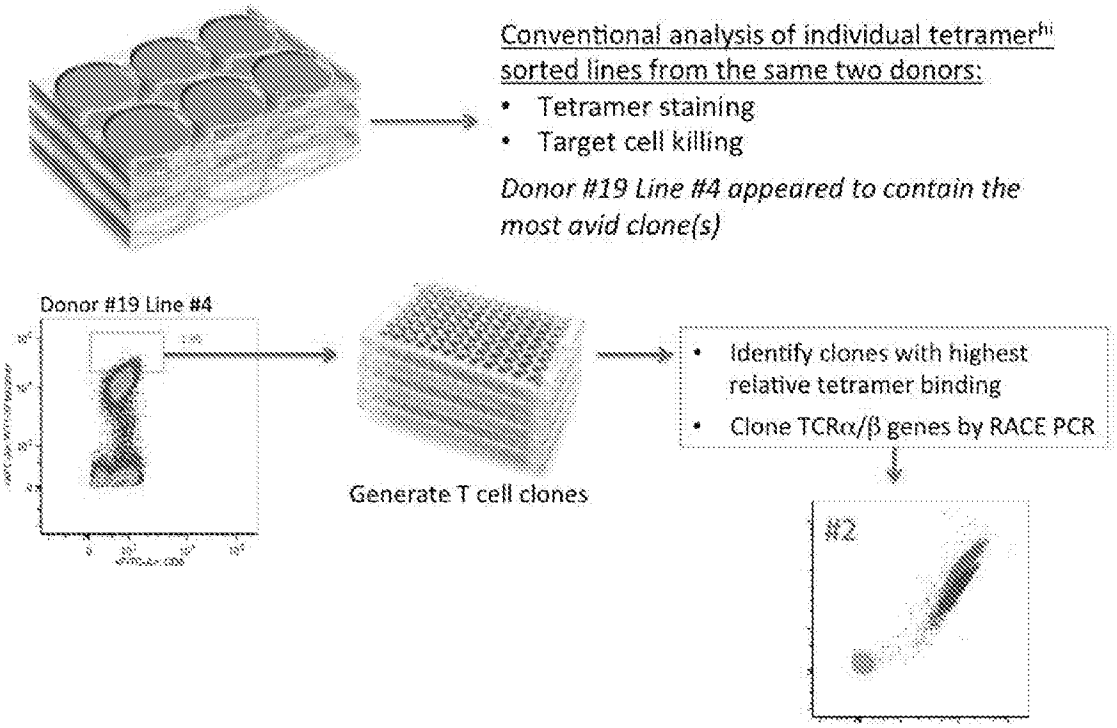

Conventional analysis of individual tetramer^hi sorted lines from the same two donors:
• Tetramer staining
• Target cell killing
Donor #19 Line #4 appeared to contain the most avid clone(s)
Donor #19 Line #4
Generate T cell clones
• Identify clones with highest relative tetramer binding
• Clone TCRα/β genes by RACE PCR
#2

Fig. 5.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 16/25535

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☒ furnished subsequent to the international filing date for the purposes of international search only:

      ☒ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☒ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 16/25535

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A61K 35/17; G01N 33/00; C12Q 1/68; C12N 5/16 (2016.01)
CPC - C12N 15/1034; G06F 19/22; C12Q 2600/106; C12Q 1/6883
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)
IPC(8): A61K 35/17; G01N 33/00; C12Q 1/68; C12N 5/16 (2016.01)
CPC: C12N 15/1034; G06F 19/22; C12Q 2600/106; C12Q 1/6883

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC: C12N 15/1034; G06F 19/22; C12Q 2600/106; C12Q 1/6883 (text search)
USPC: 424/154.1; 702/20; 506/4; 435/6.17, 7.24, 343.2, 91.2 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PatBase; Google Patents; Google Scholar
Search terms: T-cell receptor (TCR), alpha and beta chains, cognate pair, library, antigen specific T-cells, HLA-presented antigen specific, Next Generation Sequencing (NGS); clone rank and score, therapeutic

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ------- Y	KLINGER et al. Combining Next-Generation Sequencing and Immune Assays: A Novel Method for Identification of Antigen-Specific T Cells. PLoS One 19 September 2013 Vol 8 No 9 Pages e74231. Especially pg 1 col 2para 2, pg 2 col 1 para 2, pg 2 col 2 para 2, pg 3 col 1 para 1, pg 3 col 2 para 3, pg 3 fig 1, pg 6 col 2 para 2 continued to pg 7 col 1 para 1; pg 6 table 1.	8-12, 14-17 ---------- 1-7, 13
Y	US 2015/0031043 A1 (Sequenta Inc.) 29 January 2015 (29.01.2015). Especially para [0006], [0007], [0016],[0028]	1-7, 13
Y	WO 2014/145992 A1 (Adaptive Biotechnologies Corporation) 18 September 2014 (18.09.2014). Especially claim 1.	1-7, 13

☐ Further documents are listed in the continuation of Box C. ☐

* 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
07 June 2016 (07.06.2016)	11 JUL 2016
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)