



(51) International Patent Classification:
A61K 39/00 (2006.01)

(21) International Application Number:
PCT/US2014/047909

(22) International Filing Date:
23 July 2014 (23.07.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/858,839 26 July 2013 (26.07.2013) US

(71) Applicant: **SEQUENTA, INC.** [US/US]; 400 East Jamie Court, Suite 301, South San Francisco, CA 94080 (US).

(72) Inventors: **KLINGER, Mark**; 400 East Jamie Court, Suite 301, South San Francisco, CA 94080 (US). **FAHAM, Malek**; 400 East Jamie Court, Suite 301, South San Francisco, CA 94080 (US).

(74) Agents: **McEVOY, Michael, T.** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (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, LT, 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, 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).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: CANCER VACCINATION WITH ANTIGEN EVOLUTION

(57) Abstract: The invention provides methods for treating with cancer vaccines patients whose cancers undergo clonal evolution. The invention makes use of a series of cancer vaccines to stimulate a patient's immune system to mount both a humoral and cellular immune response against cancer cells as cancer-specific antigens on the cancer cells change by clonal evolution. Vaccines used in the invention are derived from antigens unique to the cancer. In one aspect of the invention, such unique antigens are determined by generating sequence-based profiles of cancer related nucleic acids. In some embodiments, cancer antigens may be identified in sequence-based profiles of exon sequences from a sample suspected of containing cancer cells; in other embodiments in which lymphoid or myeloid cancers are being treated, cancer antigens may be identified in sequence-based clonotype profiles.



WO 2015/013461 A2

CANCER VACCINATION WITH ANTIGEN EVOLUTION

[0001] This application claims the benefit of U.S. Provisional Application No. 61/858,839, filed July 26, 2013, which application is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

[0002] One of the hallmarks of cancer is its ability to evade destruction by the immune system, Hanahan and Weinberg, Cell, 144: 646-674 (2011). Yet evidence of immunosurveillance and immunoediting of cancerous cells suggests that efficient and effective cancer therapies may be attainable by informed manipulation of the immune system, e.g. Schreiber et al, Science, 331: 1565-1570 (2011); Brody et al, J. Clin. Oncol., 29: 1864-1875 (2011); Klebanoff et al, Immunological Reviews, 239: 27-44 (2011). Results of such approaches to date have been inconclusive, but tantalizing, which is due in part to the complexity and still limited understanding of many features of cancer and the immune system, including such features as exhaustion of tumor-reactive T cell populations, immunosuppression by regulatory T cells in tumors, mutability of tumor antigens, and the like, e.g. Turcotte et al, Adv. Surg. 45: 341-360 (2011).

[0003] Another hallmark of cancer is its genetic instability which frequently leads to evolution of cancer cell populations that have unique cancer antigens. Immunotherapies based on such antigens have been an active area of research and development, e.g. serological analysis of recombinant cDNA expression libraries (SEREX), Sahin et al, Proc. Natl. Acad. Sci., 92: 11810-11813 (1995); and related techniques, e.g. Sahin et al, International Patent Publication WO 2012/15964; Castle et al, Cancer Research, 72(5): OF1-OF11 (2012); Li et al, Cancers, 3: 4191-4211 (2011), and the like. Along similar lines, many myeloid and lymphoid cancers have unique surface markers or “clonotypes” associated with their immune receptors, which have been proposed as targets for cancer vaccines, namely so-called idiotypic vaccines, e.g. Kwak et al, New England J. Medicine, 327(17): 1209-1215 (1992). Unfortunately, developing cancer vaccines under either circumstance has been challenging because of on-going clonal evolution in the cancer cell population which not infrequently modifies the target epitopes of the cancer vaccines.

[0004] Recently, many diagnostic and prognostic applications have been developed that use large-scale DNA sequencing as sequencing techniques have become more efficient and convenient, e.g. Faham and Willis, U.S. patent publication 2010/0151471; Freeman et al, Genome Research, 19: 1817-1824 (2009); Boyd et al, Sci. Transl. Med., 1(12): 12ra23 (2009);

He et al, Oncotarget (March 8, 2011); Palomaki et al, Genetics in Medicine (online publication 2 February 2012).

[0005] In view of the potential impact of effective cancer vaccines, it would be highly desirable if the newly developed sequencing technologies could be used in a new treatment approach for developing and using vaccines on patients which overcame the deficiencies of current vaccines because of clonal evolution.

SUMMARY OF THE INVENTION

[0006] The present invention is drawn to methods for treating a cancer patient with a cancer vaccine when the patient's cancer is undergoing clonal evolution. The invention is exemplified in a number of implementations and applications, some of which are summarized below and throughout the specification.

[0007] In one aspect, the invention is directed to a method of controlling a cancer of a patient comprising the following steps: (a) obtaining from the patient a sample comprising cancer cells; (b) generating an exome profile and/or an expression profile from nucleic acid sequences from the cancer cells; (c) determining from the exome profile and/or expression profile a presence, absence and/or level of one or more patient-specific exons and/or transcripts correlated with the cancer and phylogenic exons or transcripts thereof; (d) formulating a peptide vaccine composition whenever the levels of any of the one or more patient-specific exons and/or transcripts correlated with the cancer or phylogenic clonotypes thereof exceed a predetermined value; and (e) administering the peptide vaccine composition to the patient to control the cancer.

[0008] In another aspect, the invention is directed to a method of controlling a myeloid or lymphoid proliferative disorder of a patient comprising the following steps: (a) obtaining from the patient a sample comprising T-cells and/or B-cells; (b) generating a clonotype profile from one or more recombined nucleic acid sequences from T-cell receptor genes and/or immunoglobulin genes; (c) determining from the clonotype profile a presence, absence and/or level of one or more patient-specific clonotypes correlated with the myeloid or lymphoid proliferative disorder and phylogenic clonotypes thereof; (d) formulating a peptide vaccine composition whenever the levels of any of the one or more patient-specific clonotypes correlated with the myeloid or lymphoid proliferative disorder or phylogenic clonotypes thereof exceed a predetermined value; and (e) administering an effective amount of the peptide vaccine composition to the patient to control the myeloid or lymphoid proliferative disorder.

[0009] These above-characterized aspects, as well as other aspects, of the present invention are exemplified in a number of illustrated implementations and applications, some of which are

shown in the figures and characterized in the claims section that follows. However, the above summary is not intended to describe each illustrated embodiment or every implementation of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention is obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0011] Fig. 1A illustrates regions of an immunoglobulin G molecule from which peptide vaccines may be made.

[0012] Fig. 1B illustrates the use of immune receptor chain frequencies to assemble antibody fragments for idiotypic cancer vaccines.

[0013] Figs. 1C-1E show a two-staged PCR scheme for amplifying and sequencing IgH or TCR β genes.

[0014] Figs. 2A-2B illustrate different embodiments for determining a clonotype based on sequence reads of an amplicon produced by the method illustrated in Figs. 1A-1C.

[0015] FIG. 3A illustrates a PCR scheme for generating three sequencing templates from an IgH chain in a single reaction. Figs. 3B-3C illustrates a PCR scheme for generating three sequencing templates from an IgH chain in three separate reactions after which the resulting amplicons are combined for a secondary PCR to add P5 and P7 primer binding sites. Fig. 3D illustrates the locations of sequence reads generated for an IgH chain.

DETAILED DESCRIPTION OF THE INVENTION

[0016] 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 such as *Genome Analysis: A Laboratory Manual Series* (Vols. I-IV); *PCR Primer: A Laboratory Manual*; and *Molecular Cloning: A Laboratory Manual* (all from Cold Spring Harbor Laboratory Press); Bot et al,

Editors, *Cancer Vaccines: From Research To Clinical Practice* (Informa Healthcare, London, 2011); Morse et al, Editors, *Handbook of Cancer Vaccines* (Humana Press, Totowa, NJ, 2004); Kast, Editor, *Peptide-Based Cancer Vaccines* (Landes Bioscience, 2000); and the like.

[0017] In one aspect, the invention is directed to the problem of therapeutic resistance in cancers that undergo clonal evolution and acquire characteristics that permit them to evade control by either drugs and/or the immune system. Generally, the invention includes methods with steps for monitoring nucleotide sequences of cancer-related nucleic acids to detect clonal evolution, including genomic DNA or expression products thereof, such as mRNAs or cDNAs; formulating a cancer vaccine responsive to the current genetic status of the cancer; and administering an effective amount of the cancer vaccine to the cancer patient. Such steps may be repeated for controlling a cancer in response to its clonal evolution. In some embodiments, an additional step of monitoring an immune response of the patient may be included to determine the effectiveness of the current vaccine. If no immune response is detected, or if an immune response level is lower than a predetermined value, then the step of formulating a vaccine may be repeated. In such a repeated step, a different cancer vaccine may be formulated, and the difference may include the use of different peptides, adjuvants, carriers, or the like. In some embodiments, a step of monitoring may include generating profiles of exome sequences encoding antigens or expressed sequences, such as mRNAs, encoding antigens. In other embodiments, a step of monitoring may include generating clonotype profiles.

[0018] In another aspect, the invention is directed to controlling myloid or lymphoid neoplasms by steps of monitoring clonotype profiles for clonal evolution of clonotypes correlated with the neoplasm; formulating a responsive anti-idiotypic vaccine whenever clonal evolution is detected; and administering an effective amount of the formulated vaccine to the cancer patient. As above, such steps may be repeated for controlling a cancer in response to its clonal evolution and/or lack of immune responsiveness, and in some embodiments, additional steps of monitoring an immune response of the patient may be included to determine the effectiveness of the current vaccine in stimulating an immune response.

[0019] In the latter aspect, clonotype profiles employed for monitoring may vary widely. In one embodiment, clonotype profiles may comprise one or more segments that cover substantially all of the nucleic acids encoding variable regions of immunoglobulins and/or T cell receptors (TCR). In some embodiments, as used herein, such a “segment” may be a region of a gene or expression product that is from 30 to 300 nucleotides in length and encode at least a portion of a variable region of an immunoglobulin or TCR. In some embodiments, if a

cancer is restricted to B cells, clonotype profiles may be limited to nucleic acids encoding variable regions of immunoglobulins. Likewise, in some embodiments, if a cancer is restricted to T cells, clonotype profiles may be restricted to nucleic acids encoding variable regions of T cell receptors. In still other embodiments, clonotype profiles may include one or more segments that encode one or more of the complementary determining regions (CDRs), such as CDR1, CDR2 or CDR3. In some embodiments, rearranged nucleic acids of clonotypes may be 25-200 nucleotide segments 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. In another embodiment, rearranged nucleic acids of clonotypes may be 25-200 nucleotide segments of 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 , or a VD rearrangement of TCR δ . In still other embodiments, rearranged nucleic acids of clonotypes may be 25-200 nucleotide segments of a VDJ rearrangement of TCR β . Additional specific segments that may make up clonotype profiles are further described below. In some embodiments, clonotype profiles may comprise a plurality of segments of the sequences indicated above or described more fully below. In some embodiments, such plurality may be in the range of from 2 to 100; in other embodiments, such plurality may be in the range of from 2 to 50; in other embodiments, such plurality may be in the range of from 2 to 10; in other embodiments, such plurality may be in the range of from 2 to 5.

[0020] In some embodiments, nucleic acid segments making up clonotype profiles may encode specific portions of variable regions of IgG molecules, such as those illustrated in Fig. 1A. Fig. 1A illustrates various functional domains of an IgG antibody, including CDRs (black regions)(1300) of heavy chain variable region (1304) and CDRs (black regions) (1302) of light chain variable region (1306) of antibody (1308), which has Fab fragment encompassed by dashed rectangle (1311). The other heavy and light chain variable regions of antibody (1308) are indicated as (1303) and (1305), respectively, and “scaffold” or “framework” regions surrounding CDRs of light chain variable region (1305) are shown on projection (1309) of light chain variable region (1305). The positions of the CDRs and their individual residue in light and heavy chain variable regions are conventionally indicated by various numbering schemes, such as the Kabat, Chothia, Abhinandan numbering schemes, or the like, which permit those of ordinary skill in the art to understand the precise locations of mutants in CDRs and framework regions of antibody-derived binding compounds. Descriptions of such numbering schemes are

described in Martin, chapter 2, Kontermann and Dubel (eds.) Antibody Engineering, Vol. 2 (Springer-Verlag, Berlin, 2010).

[0021] In one aspect, the invention comprises a method of controlling a myeloid or lymphoid proliferative disorder of a patient by the following steps: (a) obtaining from the patient a sample comprising T-cells and/or B-cells; (b) generating a clonotype profile from one or more recombinant nucleic acid sequences from T-cell receptor genes and/or immunoglobulin genes; (c) determining from the clonotype profile a presence, absence and/or level of one or more patient-specific clonotypes correlated with the myeloid or lymphoid proliferative disorder and phylogenetic clonotypes thereof; (d) formulating an anti-idiotypic vaccine composition whenever the levels of any of the one or more patient-specific clonotypes correlated with the myeloid or lymphoid proliferative disorder or phylogenetic clonotypes thereof exceed a predetermined value; and (e) administering an effective amount of the anti-idiotypic vaccine composition to the patient to control the myeloid or lymphoid proliferative disorder. In some embodiments, as described more fully below, the step of generating a clonotype profile may include steps of (i) amplifying one or more segments of recombinant nucleic acid sequences from T-cell receptor genes and/or immunoglobulin genes (or their expression products, such as mRNAs or cDNAs) to form one or more amplicons, and (ii) sequencing the one or more amplicons to form the clonotype profile. In some embodiments, the method may include an additional step of measuring an immune response after administration of a peptide vaccine composition. Patient-specific clonotypes that are correlated with a lymphoid or myeloid proliferative disorder may be determined by an initial diagnostic assay, which for example, may comprise a patient-specific polymerase chain reaction (PCR) or a sequence-based clonotype profile from a patient sample, e.g. as disclosed in the following exemplary references, incorporated herein by reference: Faham and Willis, U.S. patent 8,236,503 and U.S. patent publication US2011/0207134A1; Van Dongen et al, U.S. patent publication US2006/0234234A1; Pilarski et al, U.S. patent 6,416,948; and the like.

[0022] Further description of the various steps of the invention are given below.

Cancer Antigens

[0023] An object of the invention is to provide a series of vaccines to stimulate a patient's immune system to mount both a humoral and cellular immune response against cancer cells as cancer-specific antigens on the cancer cells change by clonal evolution. Preferably, vaccines of the invention are derived from antigens unique to the cancer. In one aspect of the invention, such unique antigens are determined by generating sequence-based profiles of cancer related nucleic acids. In some embodiments, cancer antigens may be identified in sequence-based

profiles of exon sequences from a sample suspected of containing cancer cells. In one embodiment, such sequence-based profiles (“exome profiles”) include substantially every exon of a patient. In other embodiments, exome profiles may include sequences encoding only a subset of exons, such as those known to be mutated in cancers. In some embodiments, exome profiles may include exons of cancer genes, e.g. cancer genes listed in Futreal et al, *Nature Reviews Cancer*, 4: 177-183 (2004) or Higgins et al, *Nucleic Acids Research*, 35 (suppl 1): D721-D726 (2007); or updated lists thereof. In some embodiments, cancer antigens may be identified in profiles of expression products of a sample from a patient which includes cancer cells. In some embodiments, where a cancer is a myeloid or lymphoid cancer, cancer antigens may be identified from clonotype profiles that include sequences encoding idiotypic sequences of immunoglobulins or TCRs, wherein cancer antigens are the idiotypic regions of such immunoglobulins or TCRs whose clonotypes are correlated with the myeloid or lymphoid cancer. As used herein, the term “profile” in reference to a molecular species in a sample means a listing or a representation of every different kind of such species in the sample together with its abundance or frequency.

[0024] Cancer-specific antigens may arise in a wide variety of genes, including, but not limited to, the genes in Table I.

Table I

Exemplary Cancer Genes

ABL1	AKT1	ALK	APC	ATM
BRAF	CDH1	CSF1R	CTNNB1	EGFR
ERBB2	ERBB4	FBXW7	FGFR1	FGFR2
FGFR3	FLT3	GNA11	GNAQ	GNAS
HNF1A	HRAS	IDH1	JAK2	JAK3
KDR	KIT	KRAS	MET	MLH1
MPL	NOTCH1	NPM1	NRAS	PGGFRA
PIK3CA	PTEN	PTPN11	RB1	RET
SMAD4	SMO	SRC	STK	TP53
VHL				

[0025] Once such cancer antigens are identified cancer vaccine compositions may be formulated based on the DNA or amino acid sequences of the antigens using conventional cancer vaccine techniques.

Vaccine Compositions

[0026] In one aspect, vaccine compositions of the invention are provided in order to stimulate a cytotoxic cellular immune response to a cancer. In particular, vaccine compositions of the

invention are aimed at the induction of therapeutic CD8-positive T-cell response against cancer cells. In some embodiments, steps of monitoring, formulating and administering are preceded by a step of reducing tumor burden by conventional induction therapy, such as surgery, radiotherapy or chemotherapy. As further described below, the kind of cancer vaccine employed with the invention may vary widely, and may include peptide vaccines, peptide-pulsed antigen-presenting cells (APCs), APCs in vitro transfected with cancer-antigen encoding nucleic acids, such as mRNAs, DNA vaccines, antibody fragment vaccines (which are processed by APCs after administration), and the like. Generally, in accordance with the invention, such peptides and antibody fragments are determined from sequence-based profiles of nucleic acids encoding a set of proteins that include cancer correlated proteins, such as a clonotype profile that includes a clonotype correlated with a lymphoma (i.e. one which is expressed by the lymphoma cancer cells). Once candidate peptides are identified, a further step of selecting peptide epitopes from them may be carried out. Peptide epitopes may be selected using in vitro assays or computer algorithms, such as disclosed in the following references which are incorporated herein by reference: Kokolus, U.S. patent 6,780,598; Sieker et al, *Curr. Protein Pept. Sci.*, 10(3): 286-296 (2009); Nielsen et al, *Protein Science*, 12(5): 1007-1017 (2003); Zhang et al, *Nucleic Acids Research*, 33(web server issue): W172-179 (July 1, 2005); and the like.

[0027] Vaccines that contain an immunogenically effective amount of one or more peptides, such as idiotypic peptides, may be prepared using known techniques. Once appropriately immunogenic epitopes have been defined, they can be sorted and delivered by various means, herein referred to as vaccine compositions. Such vaccine compositions can include, for example, lipopeptides (e.g., Vitiello, A. et al., *J. Clin. Invest.* 95:341, 1995), peptide compositions encapsulated in poly(DL-lactide-co-glycolide) ("PLG") microspheres (see, e.g., Eldridge, et al., *Molec. Immunol.* 28:287-294, 1991; Alonso et al., *Vaccine* 12:299-306, 1994; Jones et al., *Vaccine* 13:675-681, 1995), peptide compositions contained in immune stimulating complexes (ISCMS) (see, e.g. Takahashi et al., *Nature* 344:873-875, 1990; Hu et al., *Clin Exp Immunol.* 113:235-243, 1998), multiple antigen peptide systems (MAPs) (see e.g., Tam, J. P., *Proc. Natl. Acad. Sci. U.S.A.* 85:5409-5413, 1988; Tam, J. P., *J. Immunol. Methods* 196:17-32, 1996), viral delivery vectors (Perkus, M. E. et al., In: *Concepts in vaccine development*, Kaufmann, S. H. E., ed., p. 379, 1996; Chakrabarti, S. et al., *Nature* 320:535, 1986; Hu, S. L. et al., *Nature* 320:537, 1986; Kieny, M.-P. et al., *AIDS Bio/Technology* 4:790, 1986; Top, F. H. et al, *J. Infect. Dis.* 124:148, 1971; Chanda, P. K. et al., *Virology* 175:535, 1990), particles of viral or synthetic origin (e.g., Kofler, N. et al., *J. Immunol. Methods.* 192:25, 1996; Eldridge, J. H. et al., *Sem. Hematol* 30:16, 1993; Falo, L. D., Jr. et al., *Nature Med.* 7:649, 1995), adjuvants (Warren, H. S.,

Vogel, F. R., and Chedid, L. A. *Annu. Rev. Immunol.* 4:369, 1986; Gupta, R. K. et al., *Vaccine* 11:293, 1993), liposomes (Reddy, R. et al., *J. Immunol.* 148:1585, 1992; Rock, K. L., *Immunol. Today* 17:131, 1996), or, naked or particle absorbed cDNA (Ulmer, J. B. et al., *Science* 259:1745, 1993; Robinson, H. L., Hunt, L. A., and Webster, R. G., *Vaccine* 11:957, 1993; Shiver, J. W. et al., In: *Concepts in vaccine development*, Kaufmann, S. H. E., ed., p. 423, 1996; Cease, K. B., and Berzofsky, J. A., *Annu. Rev. Immunol.* 12:923, 1994 and Eldridge, J. H. et al., *Sem. Hematol.* 30:16, 1993, which are incorporated herein by reference in their entirety, but also for their teaching regarding vaccine compositions).

[0028] Vaccines of the invention include nucleic acid-mediated modalities. DNA or RNA encoding one or more of the peptides of the invention can also be administered to a patient. This approach is described, for instance, in Wolff et. al., *Science* 247:1465 (1990) as well as U.S. Pat. Nos. 5,580,859; 5,589,466; 5,804,566; 5,739,118; 5,736,524; 5,679,647; WO 98/04720 (which are incorporated herein by reference in their entirety, but also for their teaching regarding vaccine compositions); and in more detail below. Examples of DNA-based delivery technologies include “naked DNA”, facilitated (bupivacaine, polymers, peptide-mediated) delivery, cationic lipid complexes, and particle-mediated (“gene gun”) or pressure-mediated delivery (see, e.g., U.S. Pat. No. 5,922,687, which is incorporated herein by reference in its entirety, but also for its teaching regarding DNA vaccines).

[0029] For therapeutic or prophylactic immunization purposes, the peptides of the invention can also be expressed by viral or bacterial vectors. Examples of expression vectors include attenuated viral hosts, such as vaccinia or fowlpox. As an example of this approach, vaccinia virus is used as a vector to express nucleotide sequences that encode the peptides of the invention. Upon introduction into a host bearing a tumor, the recombinant vaccinia virus expresses the immunogenic peptide, and thereby elicits a host cytotoxic T lymphocyte (CTL) and/or helper T lymphocyte (HTL) response. Vaccinia vectors and methods useful in immunization protocols are described in, e.g., U.S. Pat. No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover et al., *Nature* 351:456–460 (1991). A wide variety of other vectors useful for therapeutic administration or immunization of the peptides of the invention, e.g. adeno and adeno-associated virus vectors, retroviral vectors, *Salmonella typhi* vectors, detoxified anthrax toxin vectors, and the like, will be apparent to those skilled in the art from the description herein.

[0030] Carriers that can be used with vaccines of the invention are well known in the art, and include, e.g., thyroglobulin, albumins such as human serum albumin, tetanus toxoid, polyamino

acids such as poly L-lysine, poly L-glutamic acid, influenza, hepatitis B virus core protein, and the like. The vaccines can contain a physiologically tolerable (i.e., acceptable) diluent such as water, or saline, preferably phosphate buffered saline. The vaccines also typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are examples of materials well known in the art. Additionally, as disclosed herein, CTL responses can be primed by conjugating peptides of the invention to lipids, such as tripalmitoyl-S-glycerylcysteinylserine (P₃CSS).

[0031] As described more fully below, a vaccine of the invention can also include antigen-presenting cells, such as dendritic cells, as a vehicle to present peptides of the invention. Vaccine compositions can be created in vitro, following dendritic cell mobilization and harvesting, whereby loading of dendritic cells occurs in vitro. For example, dendritic cells are transfected, e.g., with a minigene in accordance with the invention. The dendritic cell can then be administered to a patient to elicit immune responses in vivo.

[0032] The vaccine may be prepared as an injectable, either as liquid solution or suspension; solid form suitable for solution in, or suspension in, liquid prior to injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof.

[0033] In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminium hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween 80 emulsion.

[0034] Further examples of adjuvants and other agents include aluminium hydroxide, aluminium phosphate, aluminium potassium sulphate (alum), beryllium sulphate, silica, kaolin, carbon, water-in-oil emulsions, oil-in-water emulsions, muramyl dipeptide, bacterial endotoxin, lipid X, *Corynebacteriu parvum* (*Propionobacterium acnes*), *Bordetella pertussis*, polyribonucleotides, sodium alginate, lanolin, lysolecithin, vitamin A, saponin, liposomes, levamisole, DEAE-

dextran, blocked copolymers, biodegradable microspheres, immunostimulatory complexes (ISCOMs) or other synthetic adjuvants. Such adjuvants are available commercially from various sources, for example, Merck Adjuvant 65 (Merck and Company, Inc., Rahway, N.J.) or Freund's Incomplete Adjuvant and Complete Adjuvant (Difco Laboratories, Detroit, Mich.). Typically, adjuvants such as AMPHIGENT™ (oil-in-water), ALHYDROGEL™ (aluminium hydroxide), or a mixture of AMPHIGENT™ and ALHYDROGEL™ adjuvants may be used. The proportion of immunogen and adjuvant can be varied over a broad range so long as both are present in effective amounts. For example, aluminium hydroxide can be present in an amount of about 0.5% of the vaccine mixture (Al₂O₃ basis). Conveniently, the vaccines are formulated to contain a final concentration of immunogen in the range of from 0.2 to 200 µg/ml, preferably 5 to 50 µg/ml, most preferably 15 µg/ml.

[0035] After formulation, the vaccine may be incorporated into a sterile container which is then sealed and stored at a low temperature, for example 4° C., or it may be freeze-dried.

Lyophilisation permits long-term storage in a stabilised form.

[0036] The vaccine may be administered in a convenient manner such as by the oral, intravenous (where water soluble), intramuscular, subcutaneous, intranasal, intradermal or suppository routes or implanting (e.g. using slow release molecules). The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1% to 2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10% to 95% of active ingredient, preferably 25% to 70%. Where the vaccine composition is lyophilised, the lyophilised material may be reconstituted prior to administration, e.g. as a suspension. Reconstitution is preferably effected in buffer

A. Peptide Vaccines

[0037] As indicated above, peptides in accordance with the invention can be prepared synthetically, by recombinant DNA technology or chemical synthesis, or from natural sources such as native tumors or pathogenic organisms. Peptide epitopes may be synthesized individually

(monoepitopes) or as polyepitopic peptides. Although the peptide will preferably be substantially free of other naturally occurring host cell proteins and fragments thereof, in some embodiments the peptides may be synthetically conjugated to native fragments or particles.

[0038] The peptides in accordance with the invention can be a variety of lengths, and either in their neutral (uncharged) forms or in forms which are salts. The peptides in accordance with the invention are either free of modifications such as glycosylation, side chain oxidation, or phosphorylation; or they contain these modifications, subject to the condition that modifications do not destroy the biological activity of the peptides as described herein. The peptides of the invention can be prepared in a wide variety of ways. For the preferred relatively short size, the peptides can be synthesized in solution or on a solid support in accordance with conventional techniques. Various automatic synthesizers are commercially available and can be used in accordance with known protocols. (See, for example, Stewart & Young, *SOLID PHASE PEPTIDE SYNTHESIS*, 2D. ED., Pierce Chemical Co., 1984, which is incorporated herein by reference in its entirety, but also for its teaching regarding peptide synthesis methods). Further, individual peptide epitopes can be joined using chemical ligation to produce larger peptides that are still within the bounds of the invention.

[0039] Alternatively, recombinant DNA technology can be employed wherein a nucleotide sequence which encodes an immunogenic peptide of interest is inserted into an expression vector, transformed or transfected into an appropriate host cell and cultivated under conditions suitable for expression. These procedures are generally known in the art, as described generally in Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989), which is incorporated herein by reference in its entirety, but also for its teaching regarding recombinant expression of peptides. Thus, recombinant polypeptides which comprise one or more peptide sequences of the invention can be used to present the appropriate T cell epitope.

[0040] The nucleotide coding sequence for peptide epitopes of the preferred lengths contemplated herein can be synthesized by chemical techniques, for example, the phosphotriester method of Matteucci, et al., *J. Am. Chem. Soc.* 103:3185 (1981)), which is incorporated herein by reference in its entirety, but also for its teaching regarding recombinant nucleic acids for expression of peptides. Peptide analogs can be made simply by substituting the appropriate and desired nucleic acid base(s) for those that encode the native peptide sequence; exemplary nucleic acid substitutions are those that encode an amino acid defined by the motifs herein. The coding sequence can then be provided with appropriate linkers and ligated into expression vectors

commonly available in the art, and the vectors used to transform suitable hosts to produce the desired fusion protein. A number of such vectors and suitable host systems are now available. For expression of the fusion proteins, the coding sequence will be provided with operably linked start and stop codons, promoter and terminator regions and usually a replication system to provide an expression vector for expression in the desired cellular host. For example, promoter sequences compatible with bacterial hosts are provided in plasmids containing convenient restriction sites for insertion of the desired coding sequence. The resulting expression vectors are transformed into suitable bacterial hosts. Of course, yeast, insect or mammalian cell hosts may also be used, employing suitable vectors and control sequences.

[0041] It is often preferable that the peptide epitope be as small as possible while still maintaining substantially all of the immunologic activity of the native protein. In some embodiments, peptides for vaccines of the invention may have a length in the range of about 8 to about 30 amino acid residues; in other embodiments, peptides for vaccines of the invention may have a length in the range of about 8 to about 13 amino acid residues; in still other embodiments, peptides for vaccines of the invention may have a length in the range of about 8 to about 10 amino acid residues. HLA class II binding peptide epitope may be optimized to a length of about 6 to about 11 amino acids in length. Preferably, the peptide epitope are commensurate in size with endogenously processed pathogen-derived peptides or tumor cell peptides that are bound to the relevant HLA molecules, however, the identification and preparation of peptides of other lengths can also be carried out using the techniques described herein.

B. Vaccines With Antigen Presenting Cells (APCs)

[0042] APCs, especially dendritic cells (DCs) may be used with peptides of the invention in a vaccine formulation, such as described in the following references incorporated herein by reference: U.S. patents 6,121,044; 6,210,662; 7,060,279; and the like.

[0043] Isolation of Dendritic Cells and T-Lymphocytes. Buffy coats may be prepared from blood from healthy donors. Cells may be harvested from the leukopacs, diluted to 60 mL using Ca^{+} / Mg^{+} free phosphate buffered saline (D-PBS; Gibco Laboratories, Grand Island, N.Y.) and layered over two 15 mL columns of FEP solution, or alternatively, Lymphoprep (Nycomed Laboratories, Oslo, Norway), in 50 mL centrifuge tubes. The tubes are centrifuged at $1000\times g$ for 35 minutes at room temperature. The centrifuge run is allowed to stop without braking and the peripheral blood mononuclear cells (PBMC), present at the interface, may be harvested. PBMC are resuspended in D-PBS, centrifuged once at $650\times g$ for 10 minutes and twice more at $200\times g$ for 5 minutes to remove platelets. Platelet-depleted PBMC are resuspended in 60 mL of D-PBS,

layered on top of two columns of 15 mL of MDP (about 50% “PERCOLL”) and centrifuged at 650× g for 25 minutes at 4° C. without braking. The MDP interface (primarily monocytes) and MDP pellet cells (primarily lymphocytes) may be harvested and washed with D-PBS by centrifugation at room temperature (once at 650× g for 10 minutes and twice thereafter at 200× g for 5 minutes).

APCs may also be transfected with in vitro transcribed RNA encoding peptides (“peptide RNA”) of the invention, for example, as taught by Coughlin et al, *Blood*, 103(6): 2046-2054 (2004). Briefly, peptide RNA may be generated by inserting peptide-encoding DNA into a conventional in vitro transcription vector, such as pGEM4Z, or the like; generating messenger RNA by in vitro transcription; purifying the transcribed RNA by conventional techniques, e.g. phenol/chloroform extraction and/or column purification (RNAeasy, Qiagen); and electroporating the mRNA into the APCs. The peptide-encoding DNA may be DNA that encodes an entire antigen. The electroporated, transcribed and translated nucleic acid is then processed by the APCs into peptides which are presented to T cells.

C. Antibody Fragment Vaccines

[0044] In some embodiments, idiotypic vaccines may be made by first re-constructing immune receptors or fragments thereof, such as Fab fragments or TCRs, from their constituent chains. Such reconstruction may be carried out by matching the frequencies of the different constituent chains in their respective clonotype profiles, e.g. as described for immunoglobulin G molecules (IgGs) by Reddy et al, *Nature Biotechnology*, 28(9): 965-969 (2010); and for TCRs below.

[0045] As illustrated in Fig. 1A, nucleic acid (which may be DNA or RNA) is extracted from a sample containing T cells (100), after which in separate reaction volumes, primers (102) specific for a nucleic acids encoding TCR α 's (or a portion thereof) and primers (104) specific for nucleic acids encoding TCR β 's (or a portion thereof) are combined under conditions that allow the respective nucleic acid populations to be amplified, e.g. by a two-stage polymerase chain reaction (PCR), such as disclosed by Faham and Willis (cited above). Guidance and disclosures for selecting such primers and carrying out such reactions are described extensively in the molecular immunology literature and below (for TCR β and IgH) and in references such as, Yao et al, *Cellular and Molecular Immunology*, 4: 215-220 (2007)(for TCR α), the latter reference being incorporated herein by reference. In one embodiment, amplicons (106) and (108) produced by a two-stage PCR are ready for sequence analysis using a commercially available next generation sequencer, such as MiSeq Personal Sequencer (Illumina, San Diego, CA). After nucleotide sequences have been determined (107) and (109), databases or tables (110 and 112, respectively) are obtained. Like sequences may be counted and frequency versus sequence plots

(114 and 116) are constructed. Reconstituted TCRs may be determined by matching (118) TCR α 's and TCR β 's with identical frequencies or with frequencies having the same rank ordering. Clearly, this embodiment of the method works most efficiently when frequencies of different TCR α 's and TCR β 's are not too close together, i.e. are distinct, even taking into account experimental error.

[0046] Once a pair of clonotype sequences having equal (or equally ranked) frequencies are identified full length sequences encoding each chain may be reconstructed from the known constant and variable regions using conventional techniques for genetic manipulation and expression, e.g. Walchli et al, PLoSOne, 6(11): e27930 (2011); or the like.

D. DNA Vaccines

[0047] Therapeutic quantities of plasmid DNA can be produced for example, by fermentation in *E. coli*, followed by purification. Aliquots from the working cell bank are used to inoculate growth medium, and grown to saturation in shaker flasks or a bioreactor according to well known techniques. Plasmid DNA can be purified using standard bioseparation technologies such as solid phase anion-exchange resins supplied by QIAGEN, Inc. (Valencia, Calif.). If required, supercoiled DNA can be isolated from the open circular and linear forms using gel electrophoresis or other methods.

[0048] Purified plasmid DNA can be prepared for injection using a variety of formulations. The simplest of these is reconstitution of lyophilized DNA in sterile phosphate-buffer saline (PBS). This approach, known as "naked DNA," is currently being used for intramuscular (IM) administration in clinical trials. To maximize the immunotherapeutic effects of minigene DNA vaccines, an alternative method for formulating purified plasmid DNA may be desirable. A variety of methods have been described, and new techniques may become available. Cationic lipids can also be used in the formulation (see, e.g., as described by WO 93/24640; Mannino & Gould-Fogerite, *BioTechniques* 6(7): 682 (1988); U.S. Pat. No. 5,279,833; WO 91/06309; and Felgner, et al., *Proc. Nat'l Acad. Sci. USA* 84:7413 (1987), which are incorporated herein by reference in their entirety, but also for their teaching regarding DNA vaccine compositions. In addition, glycolipids, fusogenic liposomes, peptides and compounds referred to collectively as protective, interactive, non-condensing compounds (PINC) could also be complexed to purified plasmid DNA to influence variables such as stability, intramuscular dispersion, or trafficking to specific organs or cell types.

Administering Vaccines

[0049] Administration of Vaccines for Therapeutic or Prophylactic Purposes. The peptides of the present invention and pharmaceutical and vaccine compositions of the invention are useful for administration to mammals, particularly humans, to treat and/or prevent and/or control a cancer, particularly a myeloid or lymphoid neoplasm. Vaccine compositions containing the peptides of the invention are administered to a patient suffering from a cancer. In therapeutic applications, peptide and/or nucleic acid compositions are administered to a patient in an amount sufficient to elicit an effective CTL and/or HTL response to the cancer antigen, or more particularly, the cancer idiotype, and to cure or at least partially arrest or slow symptoms and/or complications. An amount adequate to accomplish this is defined as “therapeutically effective dose.” Amounts effective for this use will depend on, e.g., the particular composition administered, the manner of administration, the stage and severity of the disease being treated, the weight and general state of health of the patient, and the judgment of the prescribing physician.

[0050] The vaccine compositions of the invention may also be used purely as prophylactic agents. Generally the dosage for an initial prophylactic immunization generally occurs in a unit dosage range where the lower value is about 1, 5, 50, 500, or 1000 µg of peptide and the higher value is about 10,000; 20,000; 30,000; or 50,000 µg. Dosage values for a human typically range from about 500 µg to about 50,000 µg per 70 kilogram patient. This is followed by boosting dosages of between about 1.0 µg to about 50,000 µg of peptide administered at defined intervals from about four weeks to six months after the initial administration of vaccine. The immunogenicity of the vaccine may be assessed by measuring the specific activity of CTL and HTL obtained from a sample of the patient's blood.

[0051] As noted above, peptides comprising CTL and/or HTL epitopes of the invention induce immune responses when presented by HLA molecules and contacted with a CTL or HTL specific for an epitope comprised by the peptide. The manner in which the peptide is contacted with the CTL or HTL is not critical to the invention. For instance, the peptide can be contacted with the CTL or HTL either in vivo or in vitro. If the contacting occurs in vivo, the peptide itself can be administered to the patient, or other vehicles, e.g., DNA vectors encoding one or more peptides, viral vectors encoding the peptide(s), liposomes and the like, can be used, as described herein. When the peptide is contacted in vitro, the vaccinating agent can comprise a population of cells, e.g., peptide-pulsed dendritic cells, or TAA-specific CTLs, which have been induced by pulsing antigen-presenting cells in vitro with the peptide. Such a cell population is subsequently administered to a patient in a therapeutically effective dose.

[0052] The peptides or DNA encoding them can be administered individually or as fusions of one or more peptide sequences.

[0053] For pharmaceutical compositions, the immunogenic peptides of the invention, or DNA encoding them, are generally administered to an individual suffering from a cancer. The peptides or DNA encoding them can be administered individually or as fusions of one or more peptide sequences.

[0054] The dosage for an initial therapeutic immunization generally occurs in a unit dosage range where the lower value is about 1, 5, 50, 500, or 1000 µg of peptide and the higher value is about 10,000; 20,000; 30,000; or 50,000 µg. Dosage values for a human typically range from about 500 µg to about 50,000 µg per 70 kilogram patient. Boosting dosages of between about 1.0 µg to about 50000 µg of peptide pursuant to a boosting regimen over weeks to months may be administered depending upon the patient's response and condition as determined by measuring the specific activity of CTL and HTL obtained from the patient's blood. The peptides and compositions of the present invention may be employed in serious disease states, that is, life-threatening or potentially life threatening situations. In such cases, as a result of the minimal amounts of extraneous substances and the relative nontoxic nature of the peptides in preferred compositions of the invention, it is possible and may be felt desirable by the treating physician to administer substantial excesses of these peptide compositions relative to these stated dosage amounts.

[0055] The pharmaceutical compositions for therapeutic treatment are intended for parenteral, topical, oral, intrathecal or local administration. Preferably, the pharmaceutical compositions are administered parentally, e.g. intravenously, subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration which comprise a solution of the immunogenic peptides dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, e.g. water, buffered water, 0.8% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well-known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH-adjusting and buffering agents, tonicity adjusting agents, wetting agents, preservatives, and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

[0056] The concentration of peptides of the invention in the pharmaceutical formulations can vary widely, i.e., from less than about 0.1%, usually at or at least about 2% to as much as 20% to 50% or more by weight, and will be selected primarily by fluid volumes, viscosities, etc., in accordance with the particular mode of administration selected.

[0057] A human unit dose form of the peptide composition is typically included in a pharmaceutical composition that comprises a human unit dose of an acceptable carrier, preferably an aqueous carrier, and is administered in a volume of fluid that is known by those of skill in the art to be used for administration of such compositions to humans (see, e.g., *Remington's Pharmaceutical Sciences*, 17th Edition, A. Gennaro, Editor, Mack Publishing Co., Easton, Pa., 1985).

[0058] The peptides of the invention may also be administered via liposomes, which serve to target the peptides to a particular tissue, such as lymphoid tissue, or to target selectively to infected cells, as well as to increase the half-life of the peptide composition. Liposomes include emulsions, foams, micelles, insoluble monolayers, liquid crystals, phospholipid dispersions, lamellar layers and the like. In these preparations, the peptide to be delivered is incorporated as part of a liposome, alone or in conjunction with a molecule which binds to a receptor prevalent among lymphoid cells, such as monoclonal antibodies which bind to the CD45 antigen, or with other therapeutic or immunogenic compositions. Thus, liposomes either filled or decorated with a desired peptide of the invention can be directed to the site of lymphoid cells, where the liposomes then deliver the peptide compositions. Liposomes for use in accordance with the invention are formed from standard vesicle-forming lipids, which generally include neutral and negatively charged phospholipids and a sterol, such as cholesterol. The selection of lipids is generally guided by consideration of, e.g., liposome size, acid lability and stability of the liposomes in the blood stream. A variety of methods are available for preparing liposomes, as described in, e.g. Szoka, et al., *Ann. Rev. Biophys. Bioeng.* 9:467 (1980), and U.S. Pat. Nos. 4,235,871, 4,501,728, 4,837,028, and 5,019,369.

[0059] For targeting cells of the immune system, a ligand to be incorporated into the liposome can include, e.g., antibodies or fragments thereof specific for cell surface determinants of the desired immune system cells. A liposome suspension containing a peptide may be administered intravenously, locally, topically, etc. in a dose which varies according to, inter alia, the manner of administration, the peptide being delivered, and the stage of the disease being treated.

[0060] For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium

saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10–95% of active ingredient, that is, one or more peptides of the invention, and more preferably at a concentration of 25%–75%.

[0061] For aerosol administration, the immunogenic peptides are preferably supplied in finely divided form along with a surfactant and propellant. Typical percentages of peptides are 0.01%–20% by weight, preferably 1%–10%. The surfactant must, of course, be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. The surfactant may constitute 0.1%–20% by weight of the composition, preferably 0.25–5%. The balance of the composition is ordinarily propellant. A carrier can also be included, as desired, as with, e.g., lecithin for intranasal delivery.

Methods of Monitoring Immune Response

[0062] Assays to Detect T-Cell Responses. The HLA binding peptides can be tested for the ability to elicit a T-cell response. The preparation and evaluation of motif-bearing peptides are described in PCT publications WO 94/20127 and WO 94/03205), which are incorporated herein by reference in their entirety, but also for their teaching regarding motif-bearing peptides. Briefly, peptides comprising epitopes from a particular antigen are synthesized and tested for their ability to bind to the appropriate HLA proteins. These assays may involve evaluating the binding of a peptide of the invention to purified HLA class I molecules in relation to the binding of a radioiodinated reference peptide. Alternatively, cells expressing empty class I molecules (i.e. lacking peptide therein) may be evaluated for peptide binding by immunofluorescent staining and flow microfluorimetry. Other assays that may be used to evaluate peptide binding include peptide-dependent class I assembly assays and/or the inhibition of CTL recognition by peptide competition. Those peptides that bind to the class I molecule, typically with an affinity of 500 nM or less, are further evaluated for their ability to serve as targets for CTLs derived from infected or immunized individuals, as well as for their capacity to induce primary in vitro or in vivo CTL responses that can give rise to CTL populations capable of reacting with selected target cells associated with a disease. Corresponding assays are used for evaluation of HLA class II binding peptides. HLA class II motif-bearing peptides that are shown to bind, typically at an affinity of 1000 nM or less, are further evaluated for the ability to stimulate HTL responses.

[0063] Conventional assays utilized to detect T cell responses include proliferation assays, lymphokine secretion assays, direct cytotoxicity assays, and limiting dilution assays. For example, antigen-presenting cells that have been incubated with a peptide can be assayed for the ability to induce CTL responses in responder cell populations. Antigen-presenting cells can be normal cells such as peripheral blood mononuclear cells or dendritic cells. Alternatively, mutant non-human mammalian cell lines that are deficient in their ability to load class I molecules with internally processed peptides and that have been transfected with the appropriate human class I gene, may be used to test for the capacity of the peptide to induce in vitro primary CTL responses.

[0064] Peripheral blood mononuclear cells (PBMCs) may be used as the responder cell source of CTL precursors. The appropriate antigen-presenting cells are incubated with peptide, after which the peptide-loaded antigen-presenting cells are then incubated with the responder cell population under optimized culture conditions. Positive CTL activation can be determined by assaying the culture for the presence of CTLs that kill radio-labeled target cells, both specific peptide-pulsed targets as well as target cells expressing endogenously processed forms of the antigen from which the peptide sequence was derived.

[0065] More recently, a method has been devised which allows direct quantification of antigen-specific T cells by staining with Fluorescein-labelled HLA tetrameric complexes (Altman, J. D. et al., *Proc. Natl. Acad. Sci. USA* 90:10330, 1993; Altman, J. D. et al., *Science* 274:94, 1996), which are incorporated herein by reference in their entirety, but also for their teaching regarding quantification of T cells). Other relatively recent technical developments include staining for intracellular lymphokines, and interferon release assays or ELISPOT assays. Tetramer staining, intracellular lymphokine staining and ELISPOT assays all appear to be at least 10-fold more sensitive than more conventional assays (Lalvani, A. et al., *J. Exp. Med.* 186:859, 1997; Dunbar, P. R. et al., *Curr. Biol.* 8:413, 1998; Murali-Krishna, K. et al., *Immunity* 8:177, 1998, which are incorporated herein by reference in their entirety, but also for their teaching regarding quantification of T cells).

[0066] HTL activation may also be assessed using such techniques known to those in the art such as T cell proliferation and secretion of lymphokines, e.g. IL-2 (see, e.g. Alexander et al., *Immunity* 1:751–761, 1994).

Clonotype Profiles

[0067] Guidance for generating clonotype profiles for those of ordinary skill in the art is provided in Faham and Willis, U.S. patent 8,236,503 and U.S. patent publication

US2011/0207134; and Warren et al, International patent publication WO 2011/106738; which are incorporated herein by reference. Additionally, in the sections below, in one aspect, steps for generating clonotype profiles for use in the present invention are disclosed.

[0068] In one aspect, methods of the invention may be used with treatment of solid tumors. In another aspect, methods of the invention may be used with treatment of lymphoid and myeloid proliferative disorders. In another aspect, methods of the invention are applicable to lymphomas and leukemias. In another aspect, methods of the invention are applicable lymphomas or leukemias, such as follicular lymphoma, chronic lymphocytic leukemia (CLL), acute lymphocytic leukemia (ALL), chronic myelogenous leukemia (CML), acute myelogenous leukemia (AML), Hodgkins's and non-Hodgkin's lymphomas, multiple myeloma (MM), monoclonal gammopathy of undetermined significance (MGUS), mantle cell lymphoma (MCL), diffuse large B cell lymphoma (DLBCL), myelodysplastic syndromes (MDS), T cell lymphoma, or the like.

Samples

[0069] Nucleic acid profiles of the invention may be obtained from samples containing cancer cells. In particular, for clonotype profiles in connection with treating myeloid or lymphoid cancers, samples include T-cells and/or B-cells. In one aspect a sample of T cells includes at least 1,000 T cells; but more typically, a sample includes at least 10,000 T cells, and more typically, at least 100,000 T cells. In another aspect, a sample includes a number of T cells in the range of from 1000 to 1,000,000 cells. A sample of immune cells may also comprise B cells. B-cells can express immunoglobulins (antibodies, B cell receptor). As above, in one aspect a sample of B cells includes at least 1,000 B cells; but more typically, a sample includes at least 10,000 B cells, and more typically, at least 100,000 B cells. In another aspect, a sample includes a number of B cells in the range of from 1000 to 1,000,000 B cells.

[0070] Samples used in the methods of the invention can come from a variety of tissues, including, for example, tumor tissue, blood and blood plasma, lymph fluid, cerebrospinal fluid surrounding the brain and the spinal cord, synovial fluid surrounding bone joints, and the like. In one embodiment, the sample is a blood sample. The blood sample can be about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, or 5.0 mL. The sample can be a tumor biopsy. The biopsy can be from, for example, from a tumor of the brain, liver, lung, heart, colon, kidney, or bone marrow. Any biopsy technique used by those skilled in the art can be used for isolating a sample from a subject. For example, a biopsy can be an open biopsy, in which general anesthesia is used. The biopsy can be a closed biopsy, in which a smaller cut is made

than in an open biopsy. The biopsy can be a core or incisional biopsy, in which part of the tissue is removed. The biopsy can be an excisional biopsy, in which attempts to remove an entire lesion are made. The biopsy can be a fine needle aspiration biopsy, in which a sample of tissue or fluid is removed with a needle.

[0071] The sample can be a biopsy, e.g., a skin biopsy. The biopsy can be from, for example, brain, liver, lung, heart, colon, kidney, or bone marrow. Any biopsy technique used by those skilled in the art can be used for isolating a sample from a subject. For example, a biopsy can be an open biopsy, in which general anesthesia is used. The biopsy can be a closed biopsy, in which a smaller cut is made than in an open biopsy. The biopsy can be a core or incisional biopsy, in which part of the tissue is removed. The biopsy can be an excisional biopsy, in which attempts to remove an entire lesion are made. The biopsy can be a fine needle aspiration biopsy, in which a sample of tissue or fluid is removed with a needle.

[0072] The sample can include nucleic acid, for example, DNA (e.g., genomic DNA) or RNA (e.g., messenger RNA). The nucleic acid can be cell-free DNA or RNA, e.g. extracted from the circulatory system, Vlassov et al, Curr. Mol. Med., 10: 142-165 (2010); Swarup et al, FEBS Lett., 581: 795-799 (2007). 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 (e.g., a calibration test) and as many as 10 million of cells or more translating to a range of DNA of 6pg-60ug, and RNA of approximately 1pg-10ug.

[0073] In one aspect, a sample of lymphocytes for generating a clonotype profile is sufficiently large that substantially every T cell or B cell with a distinct clonotype is represented therein. In one embodiment, a sample is taken that contains with a probability of ninety-five percent every clonotype of a population present at a frequency of .001 percent or greater. In another embodiment, a sample is taken that contains with a probability of ninety-nine percent every clonotype of a population present at a frequency of .0001 percent or greater. In one embodiment, a sample of B cells or T cells includes at least a half million cells, and in another embodiment such sample includes at least one million cells.

[0074] Whenever a source of material from which a sample is taken is scarce, such as, clinical study samples, or the like, DNA from the material may be amplified by a non-biasing technique, such as whole genome amplification (WGA), multiple displacement amplification (MDA); or like technique, e.g. Hawkins et al, Curr. Opin. Biotech., 13: 65-67 (2002); Dean et al, Genome Research, 11: 1095-1099 (2001); Wang et al, Nucleic Acids Research, 32: e76 (2004); Hosono et al, Genome Research, 13: 954-964 (2003); and the like.

[0075] Blood samples are of particular interest and may be obtained using conventional techniques, e.g. Innis et al, editors, PCR Protocols (Academic Press, 1990); or the like. For example, white blood cells may be separated from blood samples using convention techniques, e.g. RosetteSep kit (Stem Cell Technologies, Vancouver, Canada). Blood samples may range in volume from 100 μ L to 10 mL; in one aspect, blood sample volumes are in the range of from 200 μ L to 2 mL. DNA and/or RNA may then be extracted from such blood sample using conventional techniques for use in methods of the invention, e.g. DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). Optionally, subsets of white blood cells, e.g. lymphocytes, may be further isolated using conventional techniques, e.g. fluorescently activated cell sorting (FACS)(Becton Dickinson, San Jose, CA), magnetically activated cell sorting (MACS)(Miltenyi Biotec, Auburn, CA), or the like.

[0076] Since the identifying recombinations are present in the DNA of each individual's adaptive immunity cells as well as their associated RNA transcripts, either RNA or DNA can be sequenced in the methods of the provided invention. A recombined sequence from a T-cell or B-cell encoding a T cell receptor or immunoglobulin molecule, or a portion thereof, is referred to as a clonotype. The DNA or RNA can correspond to sequences from T-cell receptor (TCR) genes or immunoglobulin (Ig) genes that encode antibodies. For example, the DNA and RNA can correspond to sequences encoding α , β , γ , or δ chains of a TCR. In a majority of T-cells, the TCR is a heterodimer consisting of an α -chain and β -chain. The TCR α chain is generated by VJ recombination, and the β chain receptor is generated by V(D)J recombination. For the TCR β chain, in humans there are 48 V segments, 2 D segments, and 13 J segments. Several bases may be deleted and others added (called N and P nucleotides) at each of the two junctions. In a minority of T-cells, the TCRs consist of γ and δ delta chains. The TCR γ chain is generated by VJ recombination, and the TCR δ chain is generated by V(D)J recombination (Kenneth Murphy, Paul Travers, and Mark Walport, *Janeway's Immunology* 7th edition, Garland Science, 2007, which is herein incorporated by reference in its entirety).

[0077] The DNA and RNA analyzed in the methods of the invention can correspond to sequences encoding heavy chain immunoglobulins (IgH) with constant regions (α , δ , ϵ , γ , or μ) or light chain immunoglobulins (IgK or IgL) with constant regions λ or κ . Each antibody has two identical light chains and two identical heavy chains. Each chain is composed of a constant (C) and a variable region. For the heavy chain, the variable region is composed of a variable (V), diversity (D), and joining (J) segments. Several distinct sequences coding for each type of these segments are present in the genome. A specific VDJ recombination event occurs during

the development of a B-cell, marking that cell to generate a specific heavy chain. Diversity in the light chain is generated in a similar fashion except that there is no D region so there is only VJ recombination. Somatic mutation often occurs close to the site of the recombination, causing the addition or deletion of several nucleotides, further increasing the diversity of heavy and light chains generated by B-cells. The possible diversity of the antibodies generated by a B-cell is then the product of the different heavy and light chains. The variable regions of the heavy and light chains contribute to form the antigen recognition (or binding) region or site. Added to this diversity is a process of somatic hypermutation which can occur after a specific response is mounted against some epitope.

[0078] As mentioned above, in accordance with the invention, primers may be selected to generate amplicons of subsets of recombined nucleic acids extracted from lymphocytes. Such subsets may be referred to herein as “somatically rearranged regions.” Somatically rearranged regions may comprise nucleic acids from developing or from fully developed lymphocytes, where developing lymphocytes are cells in which rearrangement of immune genes has not been completed to form molecules having full V(D)J regions. Exemplary incomplete somatically rearranged regions include incomplete IgH molecules (such as, molecules containing only D-J regions), incomplete TCR δ molecules (such as, molecules containing only D-J regions), and inactive IgK (for example, comprising Kde-V regions).

[0079] Adequate sampling of the cells is an important aspect of interpreting the repertoire data, as described further below in the definitions of “clonotype” and “repertoire.” For example, starting with 1,000 cells creates a minimum frequency that the assay is sensitive to regardless of how many sequencing reads are obtained. Therefore one aspect of this invention is the development of methods to quantitate the number of input immune receptor molecules. This has been implemented this for TCR β and IgH sequences. In either case the same set of primers are used that are capable of amplifying all the different sequences. In order to obtain an absolute number of copies, a real time PCR with the multiplex of primers is performed along with a standard with a known number of immune receptor copies. This real time PCR measurement can be made from the amplification reaction that will subsequently be sequenced or can be done on a separate aliquot of the same sample. In the case of DNA, the absolute number of rearranged immune receptor molecules can be readily converted to number of cells (within 2 fold as some cells will have 2 rearranged copies of the specific immune receptor assessed and others will have one). In the case of cDNA the measured total number of rearranged molecules in the real time sample can be extrapolated to define the total number of these molecules used in another amplification reaction of the same sample. In addition, this method can be combined with a

method to determine the total amount of RNA to define the number of rearranged immune receptor molecules in a unit amount (say 1 μg) of RNA assuming a specific efficiency of cDNA synthesis. If the total amount of cDNA is measured then the efficiency of cDNA synthesis need not be considered. If the number of cells is also known then the rearranged immune receptor copies per cell can be computed. If the number of cells is not known, one can estimate it from the total RNA as cells of specific type usually generate comparable amount of RNA. Therefore from the copies of rearranged immune receptor molecules per 1 μg one can estimate the number of these molecules per cell.

[0080] One disadvantage of doing a separate real time PCR from the reaction that would be processed for sequencing is that there might be inhibitory effects that are different in the real time PCR from the other reaction as different enzymes, input DNA, and other conditions may be utilized. Processing the products of the real time PCR for sequencing would ameliorate this problem. However low copy number using real time PCR can be due to either low number of copies or to inhibitory effects, or other suboptimal conditions in the reaction.

[0081] Another approach that can be utilized is to add a known amount of unique immune receptor rearranged molecules with a known sequence, i.e. known amounts of one or more internal standards, to the cDNA or genomic DNA from a sample of unknown quantity. By counting the relative number of molecules that are obtained for the known added sequence compared to the rest of the sequences of the same sample, one can estimate the number of rearranged immune receptor molecules in the initial cDNA sample. (Such techniques for molecular counting are well-known, e.g. Brenner et al, U.S. patent 7,537,897, which is incorporated herein by reference). Data from sequencing the added unique sequence can be used to distinguish the different possibilities if a real time PCR calibration is being used as well. Low copy number of rearranged immune receptor in the DNA (or cDNA) would create a high ratio between the number of molecules for the spiked sequence compared to the rest of the sample sequences. On the other hand, if the measured low copy number by real time PCR is due to inefficiency in the reaction, the ratio would not be high.

Amplification of Nucleic Acid Populations

[0082] Amplicons of target populations of nucleic acids may be generated by a variety of amplification techniques. In one aspect of the invention, multiplex PCR is used to amplify members of a mixture of nucleic acids, particularly mixtures comprising recombined immune molecules such as T cell receptors, or portions thereof. Guidance for carrying out multiplex PCRs of such immune molecules is found in the following references, which are incorporated by

reference: Morley, U.S. patent 5,296,351; Gorski, U.S. patent 5,837,447; Dau, U.S. patent 6,087,096; Von Dongen et al, U.S. patent publication 2006/0234234; European patent publication EP 1544308B1; and the like.

[0083] After amplification of DNA from the genome (or amplification of nucleic acid in the form of cDNA by reverse transcribing RNA), the individual nucleic acid molecules can be isolated, optionally re-amplified, and then sequenced individually. Exemplary amplification protocols may be found in van Dongen et al, *Leukemia*, 17: 2257-2317 (2003) or van Dongen et al, U.S. patent publication 2006/0234234, which is incorporated by reference. Briefly, an exemplary protocol is as follows: Reaction buffer: ABI Buffer II or ABI Gold Buffer (Life Technologies, San Diego, CA); 50 μ L final reaction volume; 100 ng sample DNA; 10 pmol of each primer (subject to adjustments to balance amplification as described below); dNTPs at 200 μ M final concentration; $MgCl_2$ at 1.5 mM final concentration (subject to optimization depending on target sequences and polymerase); Taq polymerase (1-2 U/tube); cycling conditions: preactivation 7 min at 95°C; annealing at 60°C; cycling times: 30s denaturation; 30s annealing; 30s extension. Polymerases that can be used for amplification in the methods of the invention are commercially available and include, for example, Taq polymerase, AccuPrime polymerase, or Pfu. The choice of polymerase to use can be based on whether fidelity or efficiency is preferred.

[0084] Real time PCR, picogreen staining, nanofluidic electrophoresis (e.g. LabChip) or UV absorption measurements can be used in an initial step to judge the functional amount of amplifiable material.

[0085] In one aspect, multiplex amplifications are carried out so that relative amounts of sequences in a starting population are substantially the same as those in the amplified population, or amplicon. That is, multiplex amplifications are carried out with minimal amplification bias among member sequences of a sample population. In one embodiment, such relative amounts are substantially the same if each relative amount in an amplicon is within five fold of its value in the starting sample. In another embodiment, such relative amounts are substantially the same if each relative amount in an amplicon is within two fold of its value in the starting sample. As discussed more fully below, amplification bias in PCR may be detected and corrected using conventional techniques so that a set of PCR primers may be selected for a predetermined repertoire that provide unbiased amplification of any sample.

[0086] In regard to many repertoires based on TCR or BCR sequences, a multiplex amplification optionally uses all the V segments. The reaction is optimized to attempt to get amplification that

maintains the relative abundance of the sequences amplified by different V segment primers. Some of the primers are related, and hence many of the primers may “cross talk,” amplifying templates that are not perfectly matched with it. The conditions are optimized so that each template can be amplified in a similar fashion irrespective of which primer amplified it. In other words if there are two templates, then after 1,000 fold amplification both templates can be amplified approximately 1,000 fold, and it does not matter that for one of the templates half of the amplified products carried a different primer because of the cross talk. In subsequent analysis of the sequencing data the primer sequence is eliminated from the analysis, and hence it does not matter what primer is used in the amplification as long as the templates are amplified equally.

[0087] In one embodiment, amplification bias may be avoided by carrying out a two-stage amplification (as described in Faham and Willis, cited above) wherein a small number of amplification cycles are implemented in a first, or primary, stage using primers having tails non-complementary with the target sequences. The tails include primer binding sites that are added to the ends of the sequences of the primary amplicon so that such sites are used in a second stage amplification using only a single forward primer and a single reverse primer, thereby eliminating a primary cause of amplification bias. In some embodiments, the primary PCR will have a small enough number of cycles (e.g. 5-10) to minimize the differential amplification by the different primers. The secondary amplification is done with one pair of primers, which minimizes differential amplification. In some embodiments, a small percent, e.g. one percent, of the primary PCR is taken directly to the secondary PCR. In some embodiments, a total of thirty-five cycles (equivalent to ~28 cycles without the 100 fold dilution step) allocated between a first stage and a second stage are usually sufficient to show a robust amplification irrespective of whether the cycles are divided as follows: 1 cycle primary and 34 secondary; or 25 primary and 10 secondary.

[0088] Briefly, the scheme of Faham and Willis (cited above) for amplifying IgH-encoding or TCR β encoding nucleic acids (RNA) is illustrated in Figs. 1A-1C. Similar amplification schemes are readily for other immune receptor segments, e.g. Van Dongen et al, Leukemia, 17: 2257-2317 (2003), such as, incomplete IgH rearrangements, IgK, Kde, IgL, TCR γ , TCR δ , Bcl1-IgH, Bcl2-IgH, and the like. Nucleic acids (1200) are extracted from lymphocytes in a sample and combined in a PCR with a primer (1202) specific for C region (1203) and primers (1212) specific for the various V regions (1206) of the immunoglobulin or TCR genes. Primers (1212) each have an identical tail (1214) that provides a primer binding site for a second stage of amplification. As mentioned above, primer (1202) is positioned adjacent to junction (1204)

between the C region (1203) and J region (1210). In the PCR, amplicon (1216) is generated that contains a portion of C-encoding region (1203), J-encoding region (1210), D-encoding region (1208), and a portion of V-encoding region (1206). Amplicon (1216) is further amplified in a second stage using primer P5 (1222) and primer P7 (1220), which each have tails (1225 and 1221/1223, respectively) designed for use in an Illumina DNA sequencer. Tail (1221/1223) of primer P7 (1220) optionally incorporates tag (1221) for labeling separate samples in the sequencing process. Second stage amplification produces amplicon (1230) which may be used in an Illumina DNA sequencer.

Generating Sequence Reads

[0089] Any high-throughput technique for sequencing nucleic acids can be used in the method of the invention. Preferably, such technique has a capability of generating in a cost-effective manner a volume of sequence reads from which at least 1000 clonotypes can be determined, and preferably, from which at least 10,000 to 1,000,000 clonotypes can be determined. DNA sequencing techniques include, but are not limited to, sequence by ligation, nanopore sequencing, and sequencing by synthesis (including chemistries with reversibly terminated labeled nucleotides, with enzymatic detection of base incorporation (pyrosequencing), and with reaction-byproduct detection of base incorporation. In one aspect of the invention, high-throughput methods of sequencing are employed that comprise a step of spatially isolating individual molecules on a solid surface where they are sequenced in parallel. Such solid surfaces may include nonporous surfaces (such as in Solexa sequencing, e.g. Bentley et al, Nature, 456: 53-59 (2008) or Complete Genomics sequencing, e.g. Drmanac et al, Science, 327: 78-81 (2010)), arrays of wells, which may include bead- or particle-bound templates (such as with 454, e.g. Margulies et al, Nature, 437: 376-380 (2005) or Ion Torrent sequencing, U.S. patent publication 2010/0137143 or 2010/0304982), micromachined membranes (such as with SMRT sequencing, e.g. Eid et al, Science, 323: 133-138 (2009)), or bead arrays (as with SOLiD sequencing or polony sequencing, e.g. Kim et al, Science, 316: 1481-1414 (2007)). In another aspect, such methods comprise amplifying the isolated molecules either before or after they are spatially isolated on a solid surface. Prior amplification may comprise emulsion-based amplification, such as emulsion PCR, or rolling circle amplification. Of particular interest is Solexa-based sequencing where individual template molecules are spatially isolated on a solid surface, after which they are amplified in parallel by bridge PCR to form separate clonal populations, or clusters, and then sequenced, as described in Bentley et al (cited above) and in manufacturer's instructions (e.g. TruSeqTM Sample Preparation Kit and Data Sheet, Illumina, Inc., San Diego, CA, 2010); and further in the following references: U.S. patents 6,090,592;

6,300,070; 7,115,400; and EP0972081B1; which are incorporated by reference. In one embodiment, individual molecules disposed and amplified on a solid surface form clusters in a density of at least 10^5 clusters per cm^2 ; or in a density of at least 5×10^5 per cm^2 ; or in a density of at least 10^6 clusters per cm^2 .

[0090] In one aspect, a sequence-based clonotype profile of an individual is obtained using the following steps: (a) obtaining a nucleic acid sample from T-cells and/or B-cells of the individual; (b) spatially isolating individual molecules derived from such nucleic acid sample, the individual molecules comprising at least one template generated from a nucleic acid in the sample, which template comprises a somatically rearranged region or a portion thereof, each individual molecule being capable of producing at least one sequence read; (c) sequencing said spatially isolated individual molecules; and (d) determining abundances of different sequences of the nucleic acid molecules from the nucleic acid sample to generate the clonotype profile. In one embodiment, each of the somatically rearranged regions comprise a V region and a J region. In another embodiment, individual molecules comprise nucleic acids selected from the group consisting of complete IgH molecules, incomplete IgH molecules, complete IgK complete, IgK inactive molecules, TCR β molecules, TCR γ molecules, complete TCR δ molecules, and incomplete TCR δ molecules. In another embodiment, the step of sequencing comprises generating the sequence reads having monotonically decreasing quality scores. In another embodiment, the above method comprises the following steps: (a) obtaining a nucleic acid sample from T-cells and/or B-cells of the individual; (b) spatially isolating individual molecules derived from such nucleic acid sample, the individual molecules comprising nested sets of templates each generated from a nucleic acid in the sample and each containing a somatically rearranged region or a portion thereof, each nested set being capable of producing a plurality of sequence reads each extending in the same direction and each starting from a different position on the nucleic acid from which the nested set was generated; (c) sequencing said spatially isolated individual molecules; and (d) determining abundances of different sequences of the nucleic acid molecules from the nucleic acid sample to generate the clonotype profile. In one embodiment, the step of sequencing includes producing a plurality of sequence reads for each of the nested sets. In another embodiment, each of the somatically rearranged regions comprise a V region and a J region, and each of the plurality of sequence reads starts from a different position in the V region and extends in the direction of its associated J region.

[0091] In one aspect, for each sample from an individual, the sequencing technique used in the methods of the invention generates at least 1000 sequence reads per run; in another aspect, such technique generates at least 10,000 sequence reads per run; in another aspect, such technique

generates at least 100,000 sequence reads per run; in another aspect, such technique generates at least 500,000 sequence reads per run; and in another aspect, such technique generates at least 1,000,000 sequence reads per run. In still another aspect, such technique generates between 10,000 to 1,000,000 sequence reads per run per individual sample. The sequence reads thus generated are used to determine clonotypes and from the resulting collection of clonotypes, a clonotype profile. In some embodiments, each clonotype is determined from at least 1 sequence read; in other embodiments, each clonotype is determined from at least 5 sequence reads; in other embodiments, each clonotype is determined from at least 10 sequence reads.

Generating Clonotypes from Sequence Data

[0092] Constructing clonotypes from sequence read data is disclosed in Faham and Willis (cited above), which is incorporated herein by reference. Briefly, constructing clonotypes from sequence read data depends in part on the sequencing method used to generate such data, as the different methods have different expected read lengths and data quality. In one approach, a Solexa sequencer is employed to generate sequence read data for analysis. In one embodiment, a sample is obtained that provides at least $0.5\text{-}1.0 \times 10^6$ lymphocytes to produce at least 1 million template molecules, which after optional amplification may produce a corresponding one million or more clonal populations of template molecules (or clusters). For most high throughput sequencing approaches, including the Solexa approach, such over sampling at the cluster level is desirable so that each template sequence is determined with a large degree of redundancy to increase the accuracy of sequence determination. For Solexa-based implementations, preferably the sequence of each independent template is determined 10 times or more. For other sequencing approaches with different expected read lengths and data quality, different levels of redundancy may be used for comparable accuracy of sequence determination. Those of ordinary skill in the art recognize that the above parameters, e.g. sample size, redundancy, and the like, are design choices related to particular applications.

[0093] In one aspect, clonotypes of IgH chains or TCR β chains (illustrated in Fig. 2A) are determined by at least one sequence read starting in its C region and extending in the direction of its associated V region (referred to herein as a “C read” (2304)) and at least one sequence read starting in its V region and extending in the direction of its associated J region (referred to herein as a “V read” (2306)). Such reads may or may not have an overlap region (2308) and such overlap may or may not encompass the NDN region (2315) as shown in Fig. 2A. Overlap region (2308) may be entirely in the J region, entirely in the NDN region, entirely in the V region, or it may encompass a J region-NDN region boundary or a V region-NDN region boundary, or both

such boundaries (as illustrated in Fig. 2A). Typically, such sequence reads are generated by extending sequencing primers, e.g. (2302) and (2310) in Fig. 2A, with a polymerase in a sequencing-by-synthesis reaction, e.g. Metzger, *Nature Reviews Genetics*, 11: 31-46 (2010); Fuller et al, *Nature Biotechnology*, 27: 1013-1023 (2009). The binding sites for primers (2302) and (2310) are predetermined, so that they can provide a starting point or anchoring point for initial alignment and analysis of the sequence reads. In one embodiment, a C read is positioned so that it encompasses the D and/or NDN region of the IgH chain and includes a portion of the adjacent V region, e.g. as illustrated in Figs. 2A and 2B. In one aspect, the overlap of the V read and the C read in the V region is used to align the reads with one another. In other embodiments, such alignment of sequence reads is not necessary, so that a V read may only be long enough to identify the particular V region of a clonotype. This latter aspect is illustrated in Fig. 2B. Sequence read (2330) is used to identify a V region, with or without overlapping another sequence read, and another sequence read (2332) traverses the NDN region and is used to determine the sequence thereof. Portion (2334) of sequence read (2332) that extends into the V region is used to associate the sequence information of sequence read (2332) with that of sequence read (2330) to determine a clonotype. For some sequencing methods, such as base-by-base approaches like the Solexa sequencing method, sequencing run time and reagent costs are reduced by minimizing the number of sequencing cycles in an analysis. Optionally, as illustrated in Fig. 2A, amplicon (2300) is produced with sample tag (2312) to distinguish between clonotypes originating from different biological samples, e.g. different patients. Sample tag (2312) may be identified by annealing a primer to primer binding region (2316) and extending it (2314) to produce a sequence read across tag (2312), from which sample tag (2312) is decoded.

[0094] In one aspect of the invention, sequences of clonotypes may be determined by combining information from one or more sequence reads, for example, along the V(D)J regions of the selected chains. In another aspect, sequences of clonotypes are determined by combining information from a plurality of sequence reads. Such pluralities of sequence reads may include one or more sequence reads along a sense strand (i.e. “forward” sequence reads) and one or more sequence reads along its complementary strand (i.e. “reverse” sequence reads). When multiple sequence reads are generated along the same strand, separate templates are first generated by amplifying sample molecules with primers selected for the different positions of the sequence reads. This concept is illustrated in Fig. 3A where primers (3404, 3406 and 3408) are employed to generate amplicons (3410, 3412, and 3414, respectively) in a single reaction. Such amplifications may be carried out in the same reaction or in separate reactions. In one aspect,

whenever PCR is employed, separate amplification reactions are used for generating the separate templates which, in turn, are combined and used to generate multiple sequence reads along the same strand. This latter approach is preferable for avoiding the need to balance primer concentrations (and/or other reaction parameters) to ensure equal amplification of the multiple templates (sometimes referred to herein as “balanced amplification” or “unbias amplification”). The generation of templates in separate reactions is illustrated in Figs. 3B-3C. There a sample containing IgH (3400) is divided into three portions (3472, 3474, and 3476) which are added to separate PCRs using J region primers (3401) and V region primers (3404, 3406, and 3408, respectively) to produce amplicons (3420, 3422 and 3424, respectively). The latter amplicons are then combined (3478) in secondary PCR (3480) using P5 and P7 primers to prepare the templates (3482) for bridge PCR and sequencing on an Illumina GA sequencer, or like instrument.

[0095] Sequence reads of the invention may have a wide variety of lengths, depending in part on the sequencing technique being employed. For example, for some techniques, several trade-offs may arise in its implementation, for example, (i) the number and lengths of sequence reads per template and (ii) the cost and duration of a sequencing operation. In one embodiment, sequence reads are in the range of from 20 to 200 nucleotides; in another embodiment, sequence reads are in a range of from 30 to 200 nucleotides; in still another embodiment, sequence reads are in the range of from 30 to 120 nucleotides. In one embodiment, 1 to 4 sequence reads are generated for determining the sequence of each clonotype; in another embodiment, 2 to 4 sequence reads are generated for determining the sequence of each clonotype; and in another embodiment, 2 to 3 sequence reads are generated for determining the sequence of each clonotype. In the foregoing embodiments, the numbers given are exclusive of sequence reads used to identify samples from different individuals. The lengths of the various sequence reads used in the embodiments described below may also vary based on the information that is sought to be captured by the read; for example, the starting location and length of a sequence read may be designed to provide the length of an NDN region as well as its nucleotide sequence; thus, sequence reads spanning the entire NDN region are selected. In other aspects, one or more sequence reads that in combination (but not separately) encompass a D and /or NDN region are sufficient.

[0096] In another aspect of the invention, sequences of clonotypes are determined in part by aligning sequence reads to one or more V region reference sequences and one or more J region reference sequences, and in part by base determination without alignment to reference sequences, such as in the highly variable NDN region. A variety of alignment algorithms may be applied to the sequence reads and reference sequences. For example, guidance for selecting

alignment methods is available in Batzoglou, *Briefings in Bioinformatics*, 6: 6-22 (2005), which is incorporated by reference. In one aspect, whenever V reads or C reads (as mentioned above) are aligned to V and J region reference sequences, a tree search algorithm is employed, e.g. as described generally in Gusfield (cited above) and Cormen et al, *Introduction to Algorithms*, Third Edition (The MIT Press, 2009).

[0097] The construction of IgH clonotypes from sequence reads is characterized by at least two factors: i) the presence of somatic mutations which makes alignment more difficult, and ii) the NDN region is larger so that it is often not possible to map a portion of the V segment to the C read. In one aspect of the invention, this problem is overcome by using a plurality of primer sets for generating V reads, which are located at different locations along the V region, preferably so that the primer binding sites are nonoverlapping and spaced apart, and with at least one primer binding site adjacent to the NDN region, e.g. in one embodiment from 5 to 50 bases from the V-NDN junction, or in another embodiment from 10 to 50 bases from the V-NDN junction. The redundancy of a plurality of primer sets minimizes the risk of failing to detect a clonotype due to a failure of one or two primers having binding sites affected by somatic mutations. In addition, the presence of at least one primer binding site adjacent to the NDN region makes it more likely that a V read will overlap with the C read and hence effectively extend the length of the C read. This allows for the generation of a continuous sequence that spans all sizes of NDN regions and that can also map substantially the entire V and J regions on both sides of the NDN region. Embodiments for carrying out such a scheme are illustrated in Figs. 3A and 3D. In Fig. 3A, a sample comprising IgH chains (3400) are sequenced by generating a plurality amplicons for each chain by amplifying the chains with a single set of J region primers (3401) and a plurality (three shown) of sets of V region (3402) primers (3404, 3406, 3408) to produce a plurality of nested amplicons (e.g., 3410, 3412, 3416) all comprising the same NDN region and having different lengths encompassing successively larger portions (3411, 3413, 3415) of V region (3402). Members of a nested set may be grouped together after sequencing by noting the identity (or substantial identity) of their respective NDN, J and/or C regions, thereby allowing reconstruction of a longer V(D)J segment than would be the case otherwise for a sequencing platform with limited read length and/or sequence quality. In one embodiment, the plurality of primer sets may be a number in the range of from 2 to 5. In another embodiment the plurality is 2-3; and still another embodiment the plurality is 3. The concentrations and positions of the primers in a plurality may vary widely. Concentrations of the V region primers may or may not be the same. In one embodiment, the primer closest to the NDN region has a higher concentration than the other primers of the plurality, e.g. to insure that amplicons containing the NDN region are

represented in the resulting amplicon. In a particular embodiment where a plurality of three primers is employed, a concentration ratio of 60:20:20 is used. One or more primers (e.g. 3435 and 3437 in Fig. 3D) adjacent to the NDN region (3444) may be used to generate one or more sequence reads (e.g. 3434 and 3436) that overlap the sequence read (3442) generated by J region primer (3432), thereby improving the quality of base calls in overlap region (3440). Sequence reads from the plurality of primers may or may not overlap the adjacent downstream primer binding site and/or adjacent downstream sequence read. In one embodiment, sequence reads proximal to the NDN region (e.g. 3436 and 3438) may be used to identify the particular V region associated with the clonotype. Such a plurality of primers reduces the likelihood of incomplete or failed amplification in case one of the primer binding sites is hypermutated during immunoglobulin development. It also increases the likelihood that diversity introduced by hypermutation of the V region will be captured in a clonotype sequence. A secondary PCR may be performed to prepare the nested amplicons for sequencing, e.g. by amplifying with the P5 (3401) and P7 (3404, 3406, 3408) primers as illustrated to produce amplicons (3420, 3422, and 3424), which may be distributed as single molecules on a solid surface, where they are further amplified by bridge PCR, or like technique.

[0098] Somatic Hypermutations. In one embodiment, IgH-based clonotypes that have undergone somatic hypermutation are determined as follows. A somatic mutation is defined as a sequenced base that is different from the corresponding base of a reference sequence (of the relevant segment, usually V, J or C) and that is present in a statistically significant number of reads. In one embodiment, C reads may be used to find somatic mutations with respect to the mapped J segment and likewise V reads for the V segment. Only pieces of the C and V reads are used that are either directly mapped to J or V segments or that are inside the clonotype extension up to the NDN boundary. In this way, the NDN region is avoided and the same 'sequence information' is not used for mutation finding that was previously used for clonotype determination (to avoid erroneously classifying as mutations nucleotides that are really just different recombined NDN regions). For each segment type, the mapped segment (major allele) is used as a scaffold and all reads are considered which have mapped to this allele during the read mapping phase. Each position of the reference sequences where at least one read has mapped is analyzed for somatic mutations. In one embodiment, the criteria for accepting a non-reference base as a valid mutation include the following: 1) at least N reads with the given mutation base, 2) at least a given fraction N/M reads (where M is the total number of mapped reads at this base position) and 3) a statistical cut based on the binomial distribution, the average Q score of the N reads at the mutation base as well as the number (M-N) of reads with a non-

mutation base. Preferably, the above parameters are selected so that the false discovery rate of mutations per clonotype is less than 1 in 1000, and more preferably, less than 1 in 10000.

[0099] It is expected that PCR error is concentrated in some bases that were mutated in the early cycles of PCR. Sequencing error is expected to be distributed in many bases even though it is totally random as the error is likely to have some systematic biases. It is assumed that some bases will have sequencing error at a higher rate, say 5% (5 fold the average). Given these assumptions, sequencing error becomes the dominant type of error. Distinguishing PCR errors from the occurrence of highly related clonotypes will play a role in analysis. Given the biological significance to determining that there are two or more highly related clonotypes, a conservative approach to making such calls is taken. The detection of enough of the minor clonotypes so as to be sure with high confidence (say 99.9%) that there are more than one clonotype is considered. For example of clonotypes that are present at 100 copies/1,000,000, the minor variant is detected 14 or more times for it to be designated as an independent clonotype. Similarly, for clonotypes present at 1,000 copies/1,000,000 the minor variant can be detected 74 or more times to be designated as an independent clonotype. This algorithm can be enhanced by using the base quality score that is obtained with each sequenced base. If the relationship between quality score and error rate is validated above, then instead of employing the conservative 5% error rate for all bases, the quality score can be used to decide the number of reads that need to be present to call an independent clonotype. The median quality score of the specific base in all the reads can be used, or more rigorously, the likelihood of being an error can be computed given the quality score of the specific base in each read, and then the probabilities can be combined (assuming independence) to estimate the likely number of sequencing error for that base. As a result, there are different thresholds of rejecting the sequencing error hypothesis for different bases with different quality scores. For example for a clonotype present at 1,000 copies/1,000,000 the minor variant is designated independent when it is detected 22 and 74 times if the probability of error were 0.01 and 0.05, respectively.

[00100] In the presence of sequencing errors, each genuine clonotype is surrounded by a ‘cloud’ of reads with varying numbers of errors with respect to the its sequence. The “cloud” of sequencing errors drops off in density as the distance increases from the clonotype in sequence space. A variety of algorithms are available for converting sequence reads into clonotypes. In one aspect, coalescing of sequence reads (that is, merging candidate clonotypes determined to have one or more sequencing errors) depends on at least three factors: the number of sequences obtained for each of the clonotypes being compared; the number of bases at which they differ; and the sequencing quality score at the positions at which they are discordant. A likelihood ratio

may be constructed and assessed that is based on the expected error rates and binomial distribution of errors. For example, two clonotypes, one with 150 reads and the other with 2 reads with one difference between them in an area of poor sequencing quality will likely be coalesced as they are likely to be generated by sequencing error. On the other hand two clonotypes, one with 100 reads and the other with 50 reads with two differences between them are not coalesced as they are considered to be unlikely to be generated by sequencing error. In one embodiment of the invention, the algorithm described below may be used for determining clonotypes from sequence reads. In one aspect of the invention, sequence reads are first converted into candidate clonotypes. Such a conversion depends on the sequencing platform employed. For platforms that generate high Q score long sequence reads, the sequence read or a portion thereof may be taken directly as a candidate clonotype. For platforms that generate lower Q score shorter sequence reads, some alignment and assembly steps may be required for converting a set of related sequence reads into a candidate clonotype. For example, for Solexa-based platforms, in some embodiments, candidate clonotypes are generated from collections of paired reads from multiple clusters, e.g. 10 or more, as mentioned above

[00101] The cloud of sequence reads surrounding each candidate clonotype can be modeled using the binomial distribution and a simple model for the probability of a single base error. This latter error model can be inferred from mapping V and J segments or from the clonotype finding algorithm itself, via self-consistency and convergence. A model is constructed for the probability of a given ‘cloud’ sequence Y with read count C2 and E errors (with respect to sequence X) being part of a true clonotype sequence X with perfect read count C1 under the null model that X is the only true clonotype in this region of sequence space. A decision is made whether or not to coalesce sequence Y into the clonotype X according the parameters C1, C2, and E. For any given C1 and E a max value C2 is pre-calculated for deciding to coalesce the sequence Y. The max values for C2 are chosen so that the probability of failing to coalesce Y under the null hypothesis that Y is part of clonotype X is less than some value P after integrating over all possible sequences Y with error E in the neighborhood of sequence X. The value P is controls the behavior of the algorithm and makes the coalescing more or less permissive.

[00102] If a sequence Y is not coalesced into clonotype X because its read count is above the threshold C2 for coalescing into clonotype X then it becomes a candidate for seeding separate clonotypes. An algorithm implementing such principles makes sure that any other sequences Y2, Y3, etc. which are ‘nearer’ to this sequence Y (that had been deemed independent of X) are not aggregated into X. This concept of ‘nearness’ includes both error counts with respect to Y and X and the absolute read count of X and Y, i.e. it is modeled in the same fashion

as the above model for the cloud of error sequences around clonotype X. In this way ‘cloud’ sequences can be properly attributed to their correct clonotype if they happen to be ‘near’ more than one clonotype.

[00103] In one embodiment, an algorithm proceeds in a top down fashion by starting with the sequence X with the highest read count. This sequence seeds the first clonotype. Neighboring sequences are either coalesced into this clonotype if their counts are below the precalculated thresholds (see above), or left alone if they are above the threshold or ‘closer’ to another sequence that was not coalesced. After searching all neighboring sequences within a maximum error count, the process of coalescing reads into clonotype X is finished. Its reads and all reads that have been coalesced into it are accounted for and removed from the list of reads available for making other clonotypes. The next sequence is then moved on to with the highest read count. Neighboring reads are coalesced into this clonotype as above and this process is continued until there are no more sequences with read counts above a given threshold, e.g. until all sequences with more than 1 count have been used as seeds for clonotypes.

[00104] As mentioned above, in another embodiment of the above algorithm, a further test may be added for determining whether to coalesce a candidate sequence Y into an existing clonotype X, which takes into account quality score of the relevant sequence reads. The average quality score(s) are determined for sequence(s) Y (averaged across all reads with sequence Y) where sequences Y and X differ. If the average score is above a predetermined value then it is more likely that the difference indicates a truly different clonotype that should not be coalesced and if the average score is below such predetermined value then it is more likely that sequence Y is caused by sequencing errors and therefore should be coalesced into X.

Related Clonotypes

[00105] Frequently lymphocytes produce related clonotypes. That is, multiple lymphocytes may exist or develop that produce clonotypes whose sequences are similar. This may be due to a variety of mechanism, such as hypermutation in the case of IgH molecules. As another example, in cancers, such as lymphoid neoplasms, a single lymphocyte progenitor may give rise to many related lymphocyte progeny, each possessing and/or expressing a slightly different TCR or BCR, and therefore a different clonotype, due to cancer-related somatic mutation(s), such as base substitutions, aberrant rearrangements, or the like. A set of such related clonotypes is referred to herein as a “clan.” In some case, clonotypes of a clan may arise from the mutation of another clan member. Such an “offspring” clonotype may be referred to as a phylogenic clonotype. Clonotypes within a clan may be identified by one or more measures of

relatedness to a parent clonotype, or to each other. In one embodiment, clonotypes may be grouped into the same clan by percent homology, as described more fully below. In another embodiment, clonotypes may be assigned to a clan by common usage of V regions, J regions, and/or NDN regions. For example, a clan may be defined by clonotypes having common J and ND regions but different V regions; or it may be defined by clonotypes having the same V and J regions (including identical base substitutions mutations) but with different NDN regions; or it may be defined by a clonotype that has undergone one or more insertions and/or deletions of from 1-10 bases, or from 1-5 bases, or from 1-3 bases, to generate clan members. In another embodiment, members of a clan are determined as follows.

[00106] Clonotypes are assigned to the same clan if they satisfy the following criteria: i) they are mapped to the same V and J reference segments, with the mappings occurring at the same relative positions in the clonotype sequence, and ii) their NDN regions are substantially identical. “Substantial” in reference to clan membership means that some small differences in the NDN region are allowed because somatic mutations may have occurred in this region. Preferably, in one embodiment, to avoid falsely calling a mutation in the NDN region, whether a base substitution is accepted as a cancer-related mutation depends directly on the size of the NDN region of the clan. For example, a method may accept a clonotype as a clan member if it has a one-base difference from clan NDN sequence(s) as a cancer-related mutation if the length of the clan NDN sequence(s) is m nucleotides or greater, e.g. 9 nucleotides or greater, otherwise it is not accepted, or if it has a two-base difference from clan NDN sequence(s) as cancer-related mutations if the length of the clan NDN sequence(s) is n nucleotides or greater, e.g. 20 nucleotides or greater, otherwise it is not accepted. In another embodiment, members of a clan are determined using the following criteria: (a) V read maps to the same V region, (b) C read maps to the same J region, (c) NDN region substantially identical (as described above), and (d) position of NDN region between V-NDN boundary and J-NDN boundary is the same (or equivalently, the number of downstream base additions to D and the number of upstream base additions to D are the same). Clonotypes of a single sample may be grouped into clans and clans from successive samples acquired at different times may be compared with one another. In particular, in one aspect of the invention, clans containing clonotypes correlated with a disease, such as a lymphoid neoplasm, are identified from clonotypes of each sample and compared with that of the immediately previous sample to determine disease status, such as, continued remission, incipient relapse, evidence of further clonal evolution, or the like. As used herein, “size” in reference to a clan means the number of clonotypes in the clan.

[00107] As mentioned above, in one aspect, methods of the invention monitor a level of a clan of clonotypes rather than an individual clonotype. This is because of the phenomena of clonal evolution, e.g. Campbell et al, Proc. Natl. Acad. Sci., 105: 13081-13086 (2008); Gerlinger et al, Br. J. Cancer, 103: 1139-1143 (2010). The sequence of a clone that is present in the diagnostic sample may not remain exactly the same as the one in a later sample, such as one taken upon a relapse of disease. Therefore if one is following the exact clonotype sequence that matches the diagnostic sample sequence, the detection of a relapse might fail. Such evolved clones are readily detected and identified by sequencing. For example many of the evolved clones emerge by V region replacement (called VH replacement). These types of evolved clones are missed by real time PCR techniques since the primers target the wrong V segment. However given that the D-J junction stays intact in the evolved clone, it can be detected and identified in this invention using the sequencing of individual spatially isolated molecules. Furthermore, the presence of these related clonotypes at appreciable frequency in the diagnostic sample increases the likelihood of the relevance of the clonotype. Similarly the development of somatic hypermutations in the immune receptor sequence may interfere with the real time PCR probe detection, but appropriate algorithms applied to the sequencing readout (as disclosed above) can still recognize a clonotype as an evolving clonotype. For example, somatic hypermutations in the V or J segments can be recognized. This is done by mapping the clonotypes to the closest germ line V and J sequences. Differences from the germ line sequences can be attributed to somatic hypermutations. Therefore clonotypes that evolve through somatic hypermutations in the V or J segments can be readily detected and identified. Somatic hypermutations in the NDN region can be predicted. When the remaining D segment is long enough to be recognized and mapped, any somatic mutation in it can be readily recognized. Somatic hypermutations in the N+P bases (or in D segment that is not mappable) cannot be recognized for certain as these sequences can be modified in newly recombined cells which may not be progeny of the cancerous clonotype. However algorithms are readily constructed to identify base changes that have a high likelihood of being due to somatic mutation. For example a clonotype with the same V and J segments and 1 base difference in the NDN region from the original clone(s) has a high likelihood of being the result of somatic recombination. This likelihood can be increased if there are other somatic hypermutations in the V and J segments because this identifies this specific clonotype as one that has been the subject of somatic hypermutation. Therefore the likelihood of a clonotype being the result of somatic hypermutation from an original clonotype can be computed using several parameters: the number of differences in the NDN region, the length of NDN region, as well as the presence of other somatic hypermutations in the V and/or J segments.

[00108] The clonal evolution data can be informative. For example if the major clone is an evolved clone (one that was absent previously, and therefore, previously unrecorded) then this is an indication of that tumor has acquired new genetic changes with potential selective advantages. This is not to say that the specific changes in the immune cell receptor are the cause of the selective advantage but rather that they may represent a marker for it. Tumors whose clonotypes have evolved can potentially be associated with differential prognosis. In one aspect of the invention, a clonotype or clonotypes being used as a patient-specific biomarker of a disease, such as a lymphoid neoplasm, for example, a leukemia, includes previously unrecorded clonotypes that are somatic mutants of the clonotype or clonotypes being monitored. In another aspect, whenever any previously unrecorded clonotype is at least ninety percent homologous to an existing clonotype or group of clonotypes serving as patient-specific biomarkers, then such homologous clonotype is included with or in the group of clonotypes being monitored going forward. That is, if one or more patient-specific clonotypes are identified in a lymphoid neoplasm and used to periodically monitor the disease (for example, by making measurement on less invasively acquired blood samples) and if in the course of one such measurement a new (previously unrecorded) clonotype is detected that is a somatic mutation of a clonotype of the current set, then it is added to the set of patient-specific clonotypes that are monitored for subsequent measurements. In one embodiment, if such previously unrecorded clonotype is at least ninety percent homologous with a member of the current set, then it is added to the patient-specific set of clonotype biomarkers for the next test carried out on the patient; that is, the such previously unrecorded clonotype is included in the clan of the member of the current set of clonotypes from which it was derived (based on the above analysis of the clonotype data). In another embodiment, such inclusion is carried out if the previously unrecorded clonotype is at least ninety-five percent homologous with a member of the current set. In another embodiment, such inclusion is carried out if the previously unrecorded clonotype is at least ninety-eight percent homologous with a member of the current set.

[00109] It is also possible that a cell evolves through a process that replaces the NDN region but preserves the V and J segment along with their accumulated mutations. Such cells can be identified as previously unrecorded cancer clonotypes by the identification of the common V and J segment provided they contain a sufficient number of mutations to render the chance of these mutations being independently derived small. A further constraint may be that the NDN region is of similar size to the previously sequenced clone.

[00110] While the present invention has been described with reference to several particular example embodiments, those skilled in the art will recognize that many changes may

be made thereto without departing from the spirit and scope of the present invention. The present invention is applicable to a variety of sensor implementations and other subject matter, in addition to those discussed above.

Definitions

[00111] Unless otherwise specifically defined herein, terms and symbols of nucleic acid chemistry, biochemistry, genetics, and molecular biology used herein follow those of standard treatises and texts in the field, e.g. Kornberg and Baker, DNA Replication, Second Edition (W.H. Freeman, New York, 1992); Lehninger, Biochemistry, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human Molecular Genetics, Second Edition (Wiley-Liss, New York, 1999); Abbas et al, Cellular and Molecular Immunology, 6th edition (Saunders, 2007).

[00112] “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.

[00113] “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. patents 4,683,195; 4,965,188; 4,683,202; 4,800,159 (PCR); Gelfand et al, U.S. patent 5,210,015 (real-time PCR with “taqman” probes); Wittwer et al, U.S. patent 6,174,670;

Kacian et al, U.S. patent 5,399,491 (“NASBA”); Lizardi, U.S. patent 5,854,033; Aono et al, Japanese patent publ. 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.

[00114] “Cancer vaccine” means a composition comprising one or more tumor antigens. A cancer vaccine may also comprise components found in vaccines for infectious agents, such as, solvents, stabilizers, adjuvants, buffers, surfactants, preservatives, salts, and the like. Tumor antigens may be incorporated into a cancer vaccine in a variety of formats, including but not limited to, whole tumor cells, lysates of tumor cells, gene-modified tumor cells, DNA encoding one or more tumor antigens, peptides, plasmids, viral gene transfer vectors, RNA encoding one or more tumor antigens, dendritic cells loaded with tumor antigen (e.g. tumor antigen peptides, tumor lysates, whole protein tumor antigen, transfection solutions containing RNA that encodes one or more tumor antigen, and so on), see Berzofsky et al, *J. Clin. Investigation*, 113: 1515-1525 (2004). In one embodiment, a cancer vaccine comprises one or more tumor antigens and an adjuvant. In another embodiment, one or more tumor antigens are included in a cancer vaccine as whole tumor cells, lysates of tumor cells, or one or more tumor proteins expressed from genes derived from tumor cells. In another embodiment, one or more tumor antigens comprise one or more tumor antigen peptides operationally associated with an antigen presenting cell. In another embodiment, an antigen presenting cell is a dendritic cell. In one aspect, cancer vaccines are designed to directly or indirectly stimulate a recipient’s cytotoxic T cells to react to and destroy tumor cells. An anti-idiotypic cancer vaccine means a cancer vaccine in which the one or more tumor antigens are peptides or protein fragments which are fully or partially encoded by one or more clonotype of the tumor cells.

[00115] “Clonality” as used herein means a measure of the degree to which the distribution of clonotype abundances among clonotypes of a repertoire is skewed to a single or a few clonotypes. Roughly, clonality is an inverse measure of clonotype diversity. Many measures or statistics are available from ecology describing species-abundance relationships that may be used for clonality measures in accordance with the invention, e.g. Chapters 17 & 18, in

Pielou, *An Introduction to Mathematical Ecology*, (Wiley-Interscience, 1969). In one aspect, a clonality measure used with the invention is a function of a clonotype profile (that is, the number of distinct clonotypes detected and their abundances), so that after a clonotype profile is measured, clonality may be computed from it to give a single number. One clonality measure is Simpson's measure, which is simply the probability that two randomly drawn clonotypes will be the same. Other clonality measures include information-based measures and McIntosh's diversity index, disclosed in Pielou (cited above).

[00116] "Clonotype" means a 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.

[00117] "Clonotype profile" means a listing of distinct clonotypes and their relative abundances that are derived from a population of lymphocytes. Typically, the population of lymphocytes are obtained from a tissue sample. The term "clonotype profile" is related to, but more general than, the immunology concept of 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 at least 10^3 distinct clonotypes; in other embodiments, clonotype profiles may comprise at least 10^4 distinct clonotypes; in other embodiments, clonotype profiles may comprise at least 10^5 distinct

clonotypes; in other embodiments, clonotype profiles may comprise at least 10^6 distinct clonotypes. In such embodiments, such clonotype profiles may further comprise abundances or relative frequencies of each of the distinct 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. 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 25-200 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 25-200 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. In some embodiments, clonotype profiles are derived from samples comprising from 10^5 to 10^7 lymphocytes. Such numbers of lymphocytes may be obtained from peripheral blood samples of from 1-10 mL.

[00118] “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.

[00119] “Immune response” means the tumor-antigen induced proliferation and differentiation of lymphocytes into effector cells. An aspect of an immune response of particular interest is a tumor-antigen induced proliferation of T cells. In one aspect, immune response means the proliferation of cytotoxic T cells capable of specifically recognizing tumor cells. As used herein, the term proliferation means an increase in absolute number or an increase in proportion within a population, e.g. as determined by clonotype profiles. “Immune responsiveness” means the magnitude or level of an immune response.

“Immune activation” 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). Of particular interest is the T cell component of immune activation, that is, T cell activation. There are several conventional measures of T cell activation, including T cell proliferation in response to tumor antigen stimulation, ELISPOT assays, enumeration T cells labeled by tetramer-antigenic peptide conjugates, or the like.

[00120] “Lymphoid or myeloid proliferative disorder” means any abnormal proliferative disorder in which one or more nucleotide sequences encoding one or more rearranged immune receptors can be used as a marker for monitoring such disorder. “Lymphoid or myeloid neoplasm” means an abnormal proliferation of lymphocytes or myeloid cells that may be malignant or non-malignant. A lymphoid cancer is a malignant lymphoid neoplasm. A myeloid cancer is a malignant myeloid neoplasm. Lymphoid and myeloid neoplasms are the result of, or are associated with, lymphoproliferative or myeloproliferative disorders, and include, but are not limited to, follicular lymphoma, chronic lymphocytic leukemia (CLL), acute lymphocytic leukemia (ALL), chronic myelogenous leukemia (CML), acute myelogenous leukemia (AML), Hodgkins’s and non-Hodgkin’s lymphomas, multiple myeloma (MM), monoclonal gammopathy of undetermined significance (MGUS), mantle cell lymphoma (MCL), diffuse large B cell lymphoma (DLBCL), myelodysplastic syndromes (MDS), T cell lymphoma, or the like, e.g. Jaffe et al, Blood, 112: 4384-4399 (2008); Swerdlow et al, WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (e. 4th) (IARC Press, 2008).

[00121] “Percent homologous,” “percent identical,” or like terms used in reference to the comparison of a reference sequence and another sequence (“comparison sequence”) mean that in an optimal alignment between the two sequences, the comparison sequence is identical to the reference sequence in a number of subunit positions equivalent to the indicated percentage, the subunits being nucleotides for polynucleotide comparisons or amino acids for polypeptide comparisons. As used herein, an “optimal alignment” of sequences being compared is one that maximizes matches between subunits and minimizes the number of gaps employed in constructing an alignment. Percent identities may be determined with commercially available implementations of algorithms, such as that described by Needleman and Wunsch, J. Mol. Biol., 48: 443-453 (1970) (“GAP” program of Wisconsin Sequence Analysis Package, Genetics Computer Group, Madison, WI), or the like. Other software packages in the art for constructing alignments and calculating percentage identity or other measures of similarity include the “BestFit” program, based on the algorithm of Smith and Waterman, Advances in Applied Mathematics, 2: 482-489 (1981) (Wisconsin Sequence Analysis Package, Genetics Computer

Group, Madison, WI). In other words, for example, to obtain a polynucleotide having a nucleotide sequence at least 95 percent identical to a reference nucleotide sequence, up to five percent of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to five percent of the total number of nucleotides in the reference sequence may be inserted into the reference sequence.

[00122] “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 $>90^{\circ}\text{C}$, primers annealed at a temperature in the range $50\text{-}75^{\circ}\text{C}$, and primers extended at a temperature in the range $72\text{-}78^{\circ}\text{C}$. 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 5,210,015 (“taqman”); Wittwer et al, U.S. patents 6,174,670 and 6,569,627 (intercalating dyes); Tyagi et al, U.S. patent 5,925,517 (molecular beacons); which patents 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” means 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 from 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, β_2 -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.

[00123] “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 acid 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).

[00124] “Sequence read” means a sequence of nucleotides determined from a sequence or stream of data generated by a sequencing technique, which determination is made, for example, by means of base-calling software associated with the technique, e.g. base-calling software from a commercial provider of a DNA sequencing platform. A sequence read usually includes quality scores for each nucleotide in the sequence. Typically, sequence reads are made by extending a primer along a template nucleic acid, e.g. with a DNA polymerase or a DNA ligase. Data is generated by recording signals, such as optical, chemical (e.g. pH change), or electrical signals, associated with such extension. Such initial data is converted into a sequence read.

[00125] “Pulse” means exposure of APCs to antigen for a time sufficient to promote presentation of that antigen on the surface of the APC. Peptide pulsing protocols are known in the art (see for example Redchenko and Rickinson (1999) J. Virol. 334-342; Nestle et al (1998) Nat. Med. 4 328-332; Tjandrawan et al (1998) J. Immunotherapy 21 149-157). For example, in a standard protocol for loading dendritic cells with peptides, cells are incubated with peptide at 50 µg/ml with 3 µg/ml β-2 microglobulin for two hours in serum free medium. The unbound peptide is then washed off.

[00126] “Antigen-presenting cell” or “APC” means a specialized cell that express class II MHC proteins on its cell surface. Short peptides associate non-covalently with the surface class II MHC proteins which are then detected by other T cells such as T helper cells (HTL or helper T lymphocytes). Types of antigen presenting cells include, macrophages, B cells, and dendritic cells.

[00127] “Cytotoxic T cell” means a cell which will kill another cell that has foreign macromolecules on its surface. Frequently these foreign macromolecules will be peptides non-covalently bound to cell surface class I MHC molecules. Most, but not all cytotoxic T cells express surface CD8 protein. A small percentage of cytotoxic T cells express CD4 on their cell surface and a small percentage of cytotoxic T cells do not express either CD4 or CD8 on their cell surface. Cytotoxic T cell, CTL and Tc cell will be used interchangeably herein.

[00128] “Epitope” means a set of amino acid residues which is involved in recognition by a particular immunoglobulin, or in the context of T cells, those residues necessary for recognition by T cell receptor proteins and/or Major Histocompatibility Complex (MHC) receptors. In an immune system setting, in vivo or in vitro, an epitope is the collective features of a molecule, such as primary, secondary and tertiary peptide structure, and charge, that together

form a site recognized by an immunoglobulin, T cell receptor or HLA molecule. Throughout this disclosure epitope and peptide are often used interchangeably.

[00129] “Peptide” means a series of residues, typically L-amino acids, connected one to the other, typically by peptide bonds between the α -amino and carboxyl groups of adjacent amino acids. The preferred CTL-inducing peptides of the invention are 13 residues or less in length and usually consist of between about 8 and about 11 residues, preferably 9 or 10 residues. The preferred HTL-inducing oligopeptides are less than about 50 residues in length and usually consist of between about 6 and about 30 residues, more usually between about 12 and 25, and often between about 15 and 20 residues.

[00130] “Pharmaceutically acceptable” refers to a generally non-toxic, inert, and/or physiologically compatible composition.

[00131] “Pharmaceutical excipient” comprises a material such as an adjuvant, a carrier, pH-adjusting and buffering agents, tonicity adjusting agents, wetting agents, preservative, and the like

[00132] “Vaccine” means a composition that contains one or more peptides of the invention. There are numerous embodiments of vaccines in accordance with the invention, such as by a cocktail of one or more peptides; one or more epitopes of the invention comprised by a polyepitopic peptide; or nucleic acids that encode such peptides or polypeptides, e.g., a minigene that encodes a monoepitopic or polyepitopic peptide. The peptides or polypeptides can optionally be modified, such as by lipidation, addition of targeting or other sequences. HLA class I-binding peptides of the invention can be admixed with, or linked to, HLA class II-binding peptides, to facilitate activation of both cytotoxic T lymphocytes and helper T lymphocytes. Vaccines can also comprise peptide-pulsed antigen presenting cells, e.g., dendritic cells.

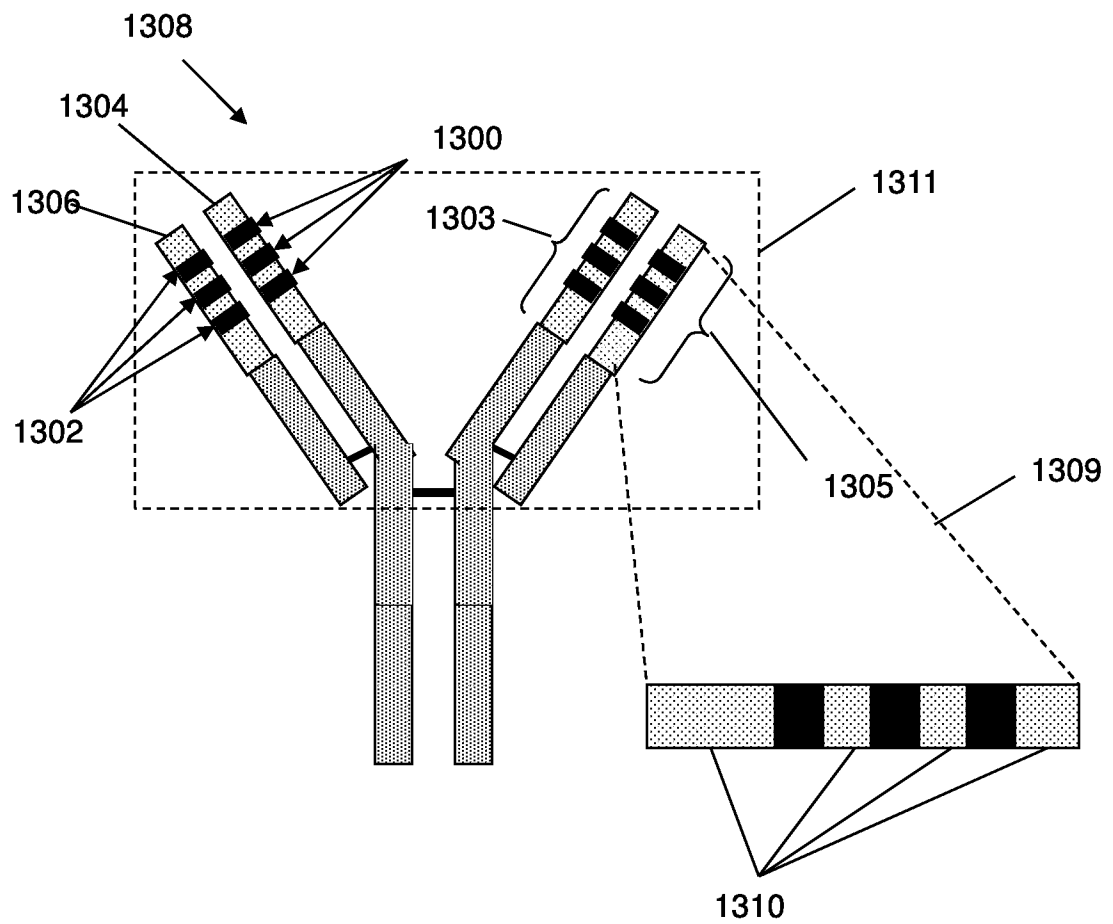
What is claimed is:

1. A method of controlling a myeloid or lymphoid proliferative disorder of a patient, the method comprising the steps of:
 - (a) obtaining from the patient a sample comprising T-cells and/or B-cells;
 - (b) generating a clonotype profile from one or more recombined nucleic acid sequences from T-cell receptor genes and/or immunoglobulin genes;
 - (c) determining from the clonotype profile a presence, absence and/or level of one or more patient-specific clonotypes correlated with the myeloid or lymphoid proliferative disorder and phylogenic clonotypes thereof;
 - (d) formulating a peptide vaccine composition whenever the levels of any of the one or more patient-specific clonotypes correlated with the myeloid or lymphoid proliferative disorder or phylogenic clonotypes thereof exceed a predetermined value, wherein the peptide vaccine composition comprises peptides encoded by nucleic acid sequences of the one or more patient-specific clonotypes correlated with the myeloid or lymphoid proliferative disorder or phylogenic clonotypes thereof which exceed such predetermined value; and
 - (e) administering an effective amount of the peptide vaccine composition to the patient to control the myeloid or lymphoid proliferative disorder.
2. The method of claim 1 further including the step of repeating said steps (a) through (e) after a predetermined monitoring interval.
3. The method of claim 2 wherein said predetermined monitoring interval is in the range of from 1 to 12 months.
4. The method of claim 1 further including a step of determining an immune response of said patient to said peptide vaccine composition and optionally repeating said steps (a) through (e) if the immune response of said patient is below a predetermined level.
5. The method of claim 4 wherein said immune response is a level of T cell activation by peptides of said peptide vaccine composition and wherein said predetermined level of said immune response to said peptide vaccine composition is at least a two-fold increase in the level of T cell activation by peptides of said peptide vaccine composition within one month of said step of administering.

6. The method of claim 5 wherein said level of T cell activation is measured by an ELISPOT assay.
7. The method of claim 1 wherein said peptide vaccine composition is an idiotypic vaccine composition comprising one or more peptides encoded by clonotypes correlated with said myeloid or lymphoid proliferative disorder.
8. The method of claim 7 wherein at least one of said one or more peptides is incorporated into a Fab fragment in said peptide vaccine composition.
9. The method of claim 1 wherein said predetermined value of said one or more patient-specific clonotypes correlated with said myeloid or lymphoid proliferative disorder or phylogenic clonotypes thereof is a value measured immediately after said patient received induction therapy for said myeloid or lymphoid proliferative disorder.
10. The method of claim 1 wherein said myeloid or lymphoid proliferative disorder is a B cell leukemia or lymphoma and wherein said clonotype profile includes segments encompassing at least a portion of a CDR1, CDR2 or CDR3 region of an IgH or IgL variable region.
11. The method of claim 1 wherein said myeloid or lymphoid proliferative disorder is a T cell leukemia or lymphoma and wherein said clonotype profile includes segments encompassing at least a portion of a CDR1, CDR2 or CDR3 region of a TCR α or a TCR β variable region.
12. A method of controlling a cancer of a patient, the method comprising the steps of:
 - (a) obtaining from the patient a sample comprising cancer cells;
 - (b) generating an exome profile and/or an expression profile from nucleic acid sequences from cells of the sample;
 - (c) determining from the exome profile and/or expression profile a presence, absence and/or level of one or more patient-specific exons and/or transcripts correlated with the cancer and phylogenic exons or transcripts thereof;
 - (d) formulating a peptide vaccine composition whenever the levels of any of the one or more patient-specific exons or transcripts correlated with the cancer or phylogenic clonotypes thereof exceed a predetermined value;

(e) administering the peptide vaccine composition to the patient to control the cancer.

13. The method of claim 12 further including the step of repeating said steps (a) through (e) after a predetermined monitoring interval.
14. The method of claim 13 wherein said predetermined monitoring interval is in the range of from 1 to 12 months.
15. The method of claim 12 further including a step of determining an immune response of said patient to said peptide vaccine composition and optionally repeating said steps (a) through (e) if the immune response of said patient is below a predetermined level.
16. The method of claim 15 wherein said immune response is a level of T cell activation by peptides of said peptide vaccine composition and wherein said predetermined level of said immune response to peptides of said peptide vaccine composition is at least a two-fold increase in the level of T cell activation by peptides of said peptide vaccine composition within one month of said step of administering.
17. The method of claim 16 wherein said level of T cell activation is measured by an ELISPOT assay.
18. The method of any of claims 12 through 17 wherein said cancer cells are B cells and/or T cells and wherein said exome profile and/or expression profile is a clonotype profile.
19. The method of claim 18 wherein said cancer is a myeloid or lymphoid cancer and wherein said peptide vaccine composition is an idiotype vaccine composition comprising one or more peptides encoded by clonotypes correlated with the myeloid or lymphoid cancer.

**Fig. 1A**

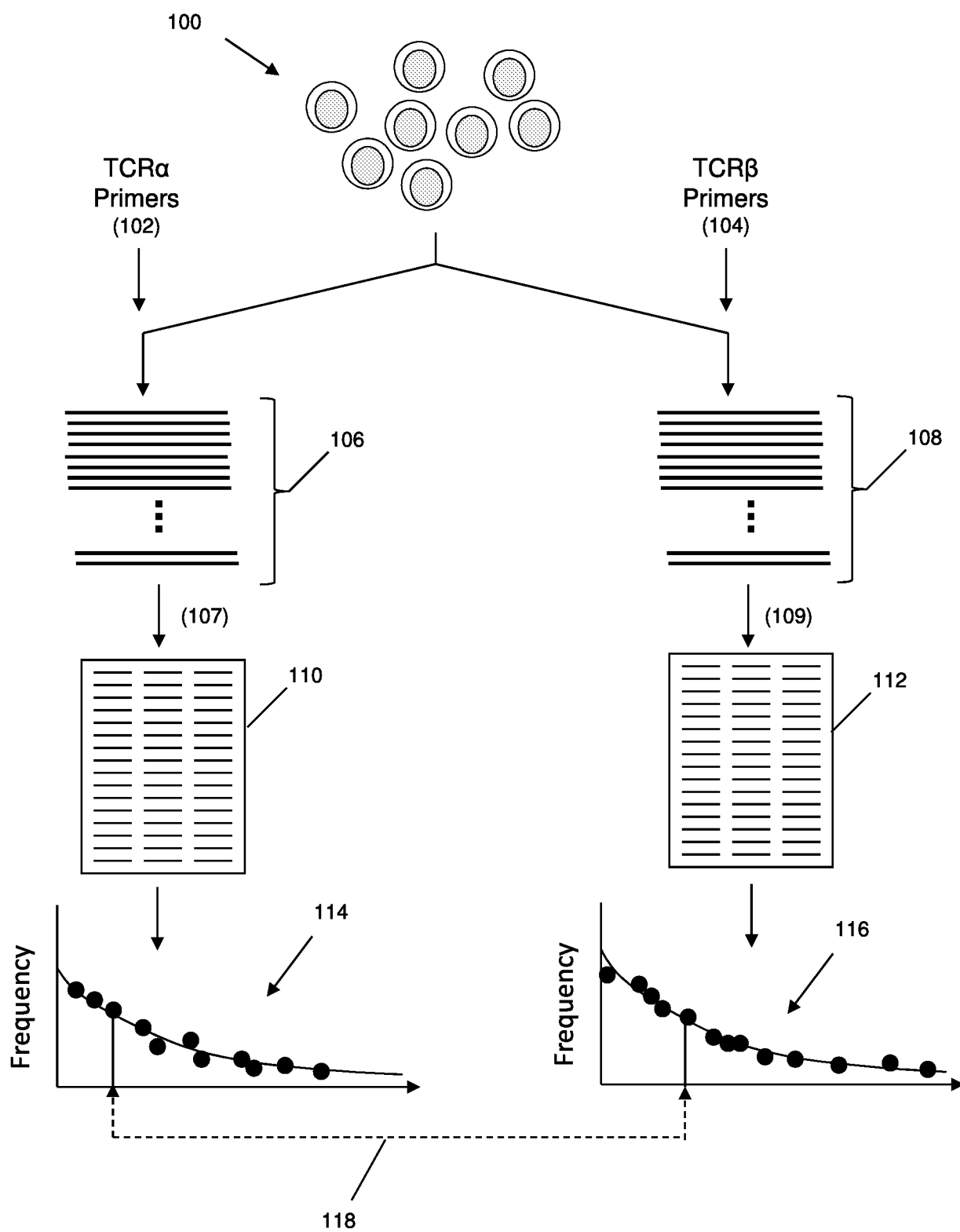


Fig. 1B

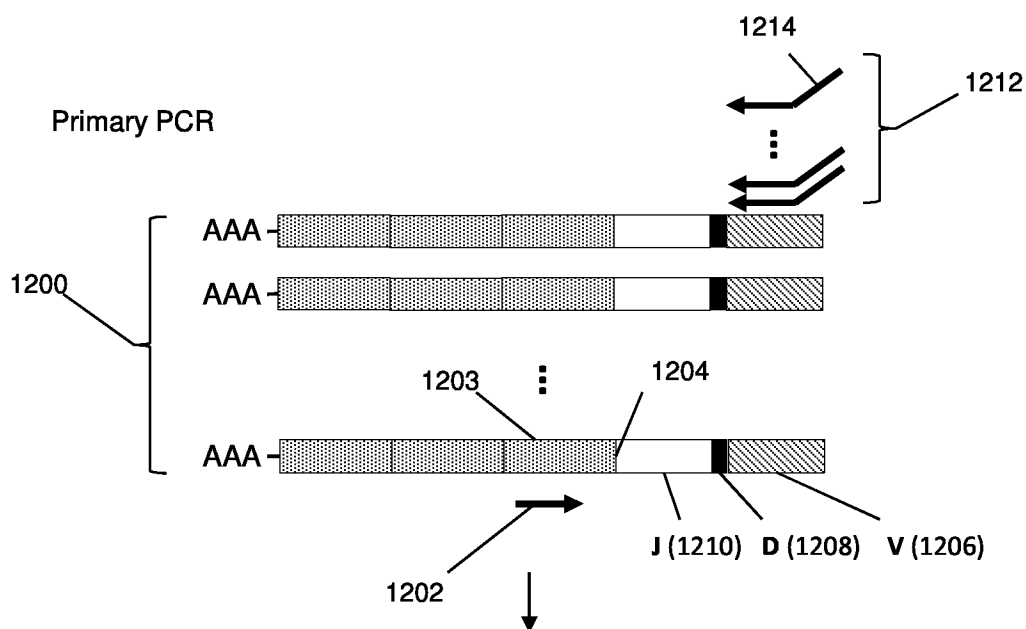


Fig. 1C

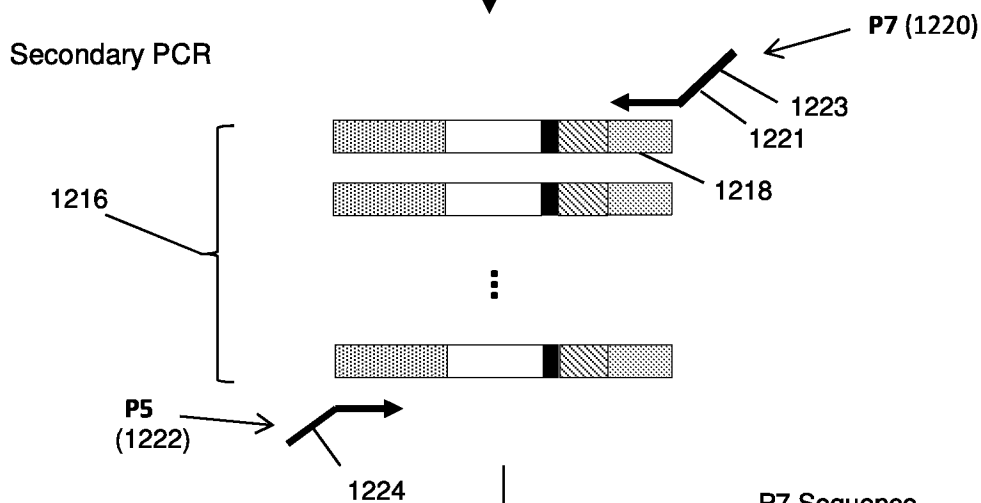


Fig. 1D

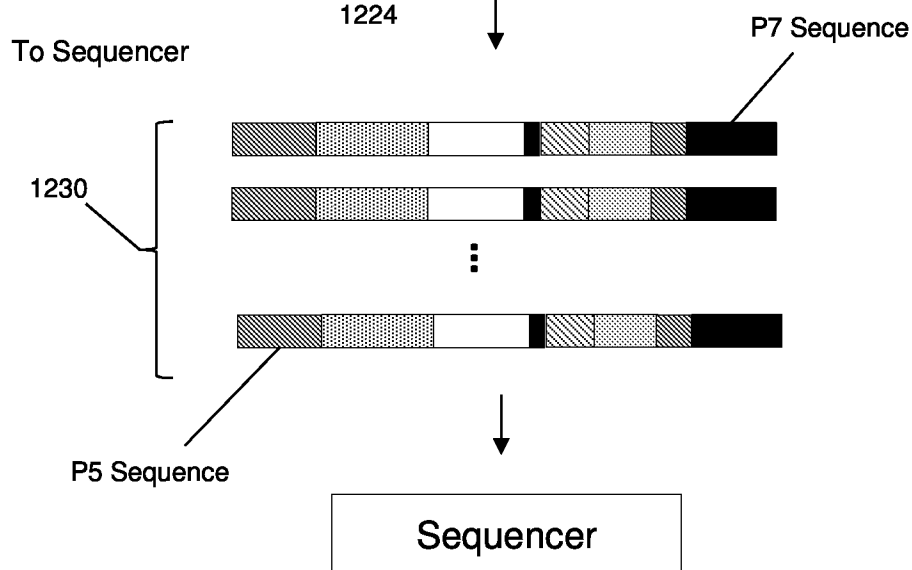


Fig. 1E

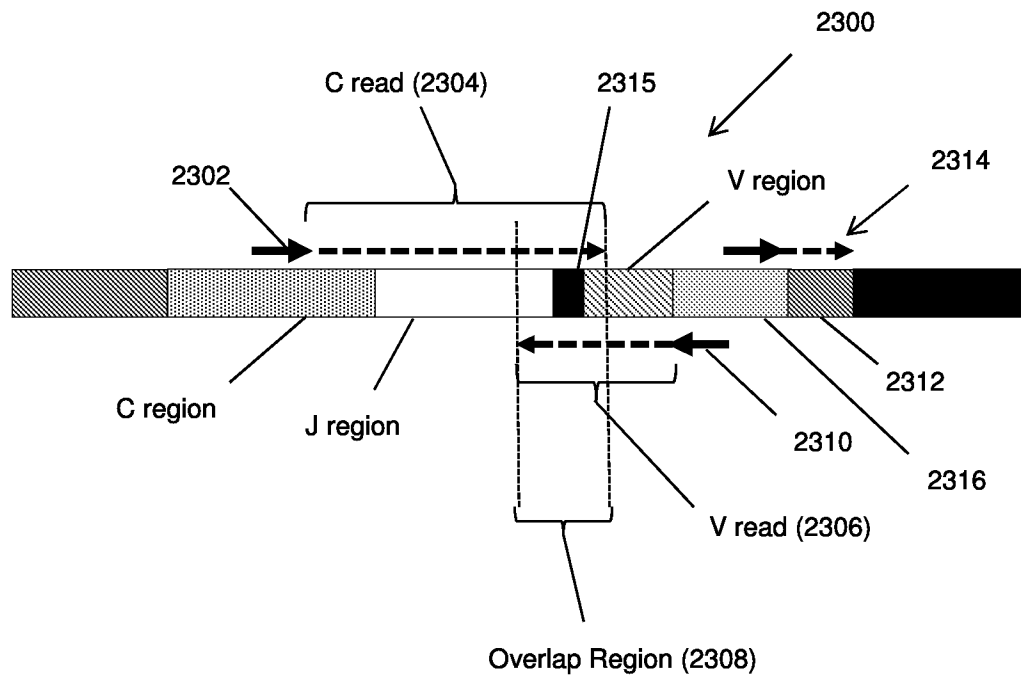


Fig. 2A

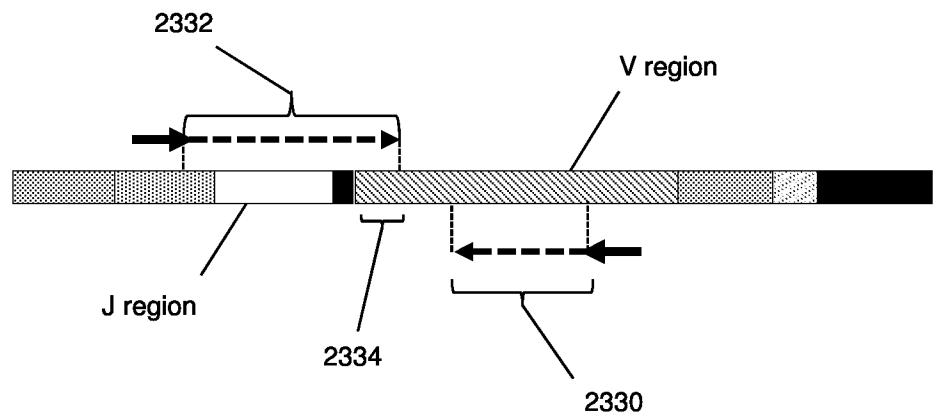


Fig. 2B

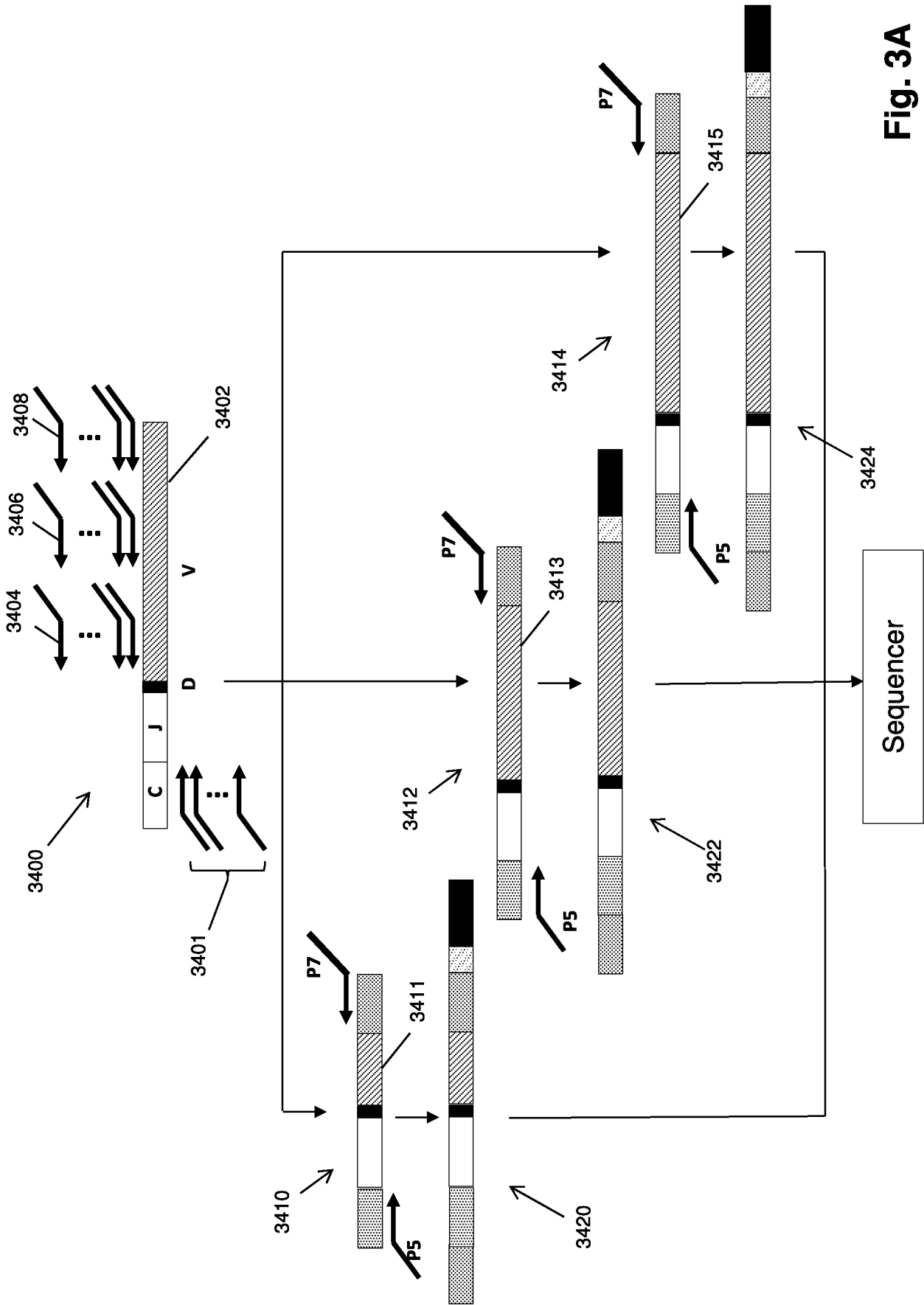


Fig. 3A

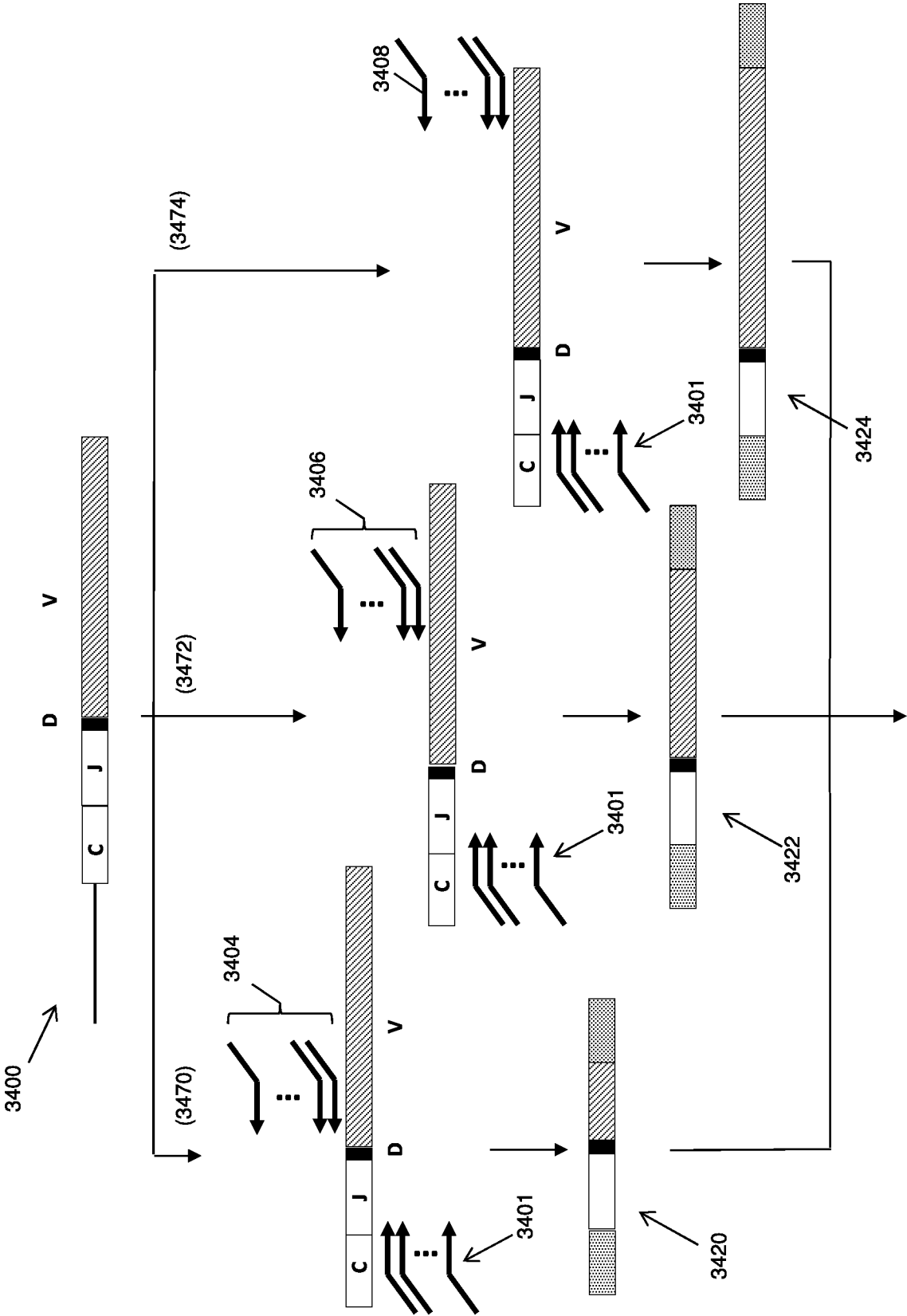


Fig. 3B

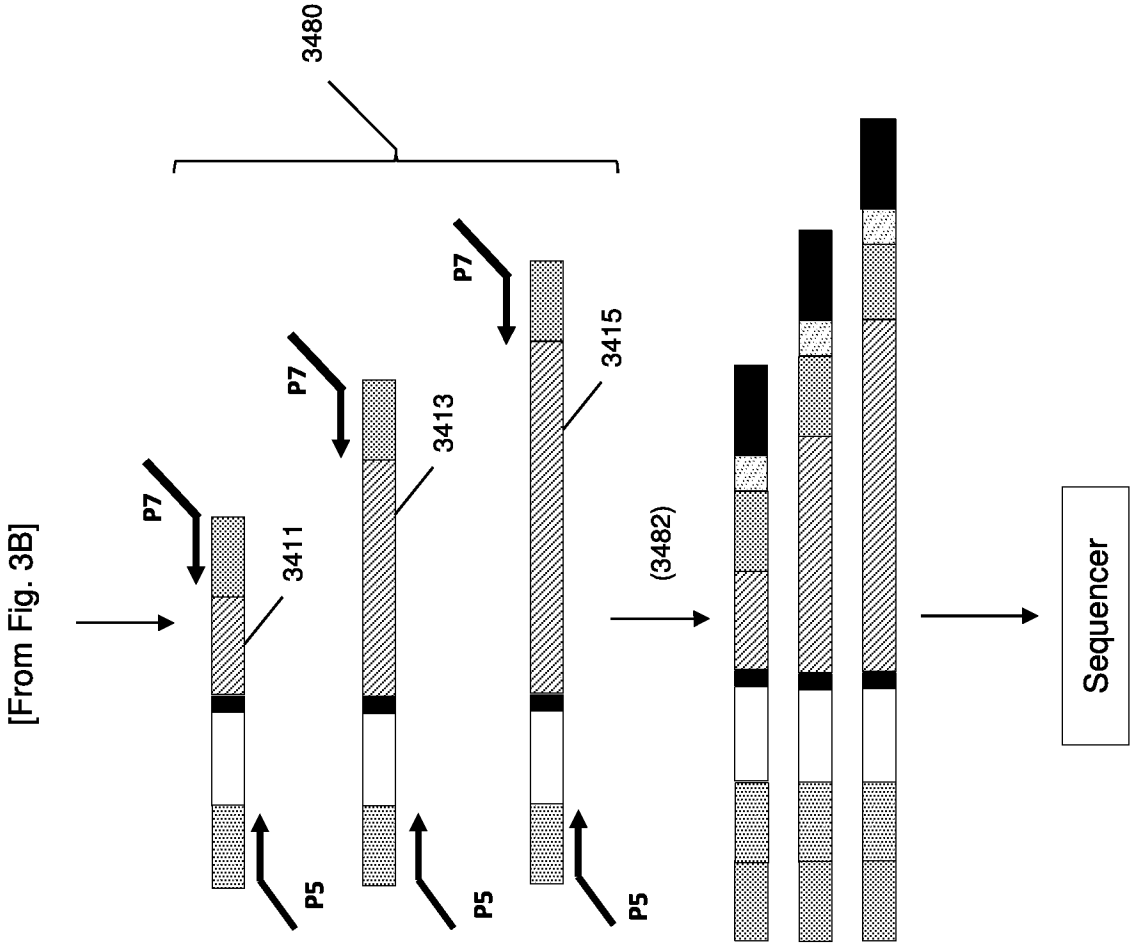


Fig. 3C

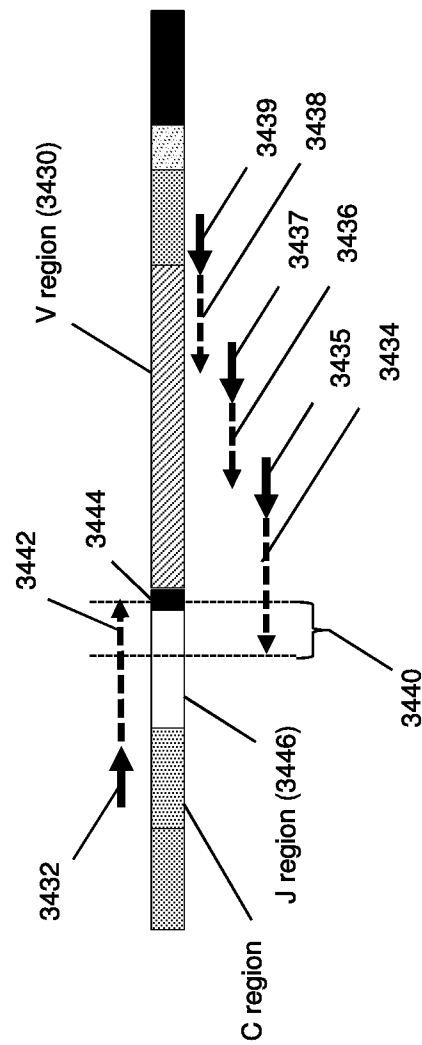


Fig. 3D