TITLE

Identification of Polynucleotides Associated With a Sample.

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application is related to U.S. provisional patent application No. 61/517,976, filed April 28, 2011, U.S. provisional patent application No. 61/575,652, filed August 24, 2011, U.S. provisional patent application No. 61/599,870, filed February 16, 2012, and U.S. provisional patent application No. 61/608,571, filed March 8, 2012; the disclosure of which is hereby incorporated by reference in its entirety for all purposes.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002] This invention was made with government support under N01-HV-28183 NHLBI Proteomics Center and N01-HV-00242 NHLBI Proteomics Center awarded by the National Heart Lung and Blood Institute of the U.S. National Institutes of Health. The government has certain rights in the invention.

SEQUENCE LISTING

[0003] This instant application contains a "lengthy" Sequence Listing which has been submitted via CD-R, in lieu of a printed paper copy, and is hereby incorporated by reference in its entirety. Said CD-R, recorded on April 25, 2012, are labeled "CRF," "Copy 1 – SEQUENCE LISTING PART, " "Copy 2 – SEQUENCE LISTING PART" and "Copy 3 - SEQUENCE LISTING PART," respectively, and each contains only one single self-extracting file named 20786PCT_CRF_Sequencelising.EXE (138,571,776 bytes), which subsequently contains one uncompressed ASCII text file named 20786PCT_CRF_Sequencelisting.TXT (795,566,080 bytes).

BACKGROUND

[0004] Producing therapeutic monoclonal antibodies from human sources is biologically and technically challenging. To date, several approaches have been described, including generating human hybridomas, using transgenic mice expressing human immunoglobulins, and using human immunoglobulin phage display libraries.

[0005] Human hybridomas can be difficult to generate because human myeloma fusion partners, unlike their mouse counterparts, are inefficient at generating hybridomas. Human hybridomas also have a tendency to spontaneously lose the expressed antibody genes after prolonged culture. Epstein-Barr virus (EBV) transformation immortalizes B cells, but only extremely small fractions of all of the EBV-transformed B cells are affinity matured or recognize the target

antigen. The generation of hybridomas typically includes large screens to obtain therapeutic monoclonal antibodies. None of the therapeutic monoclonal antibodies currently approved by the U.S. F.D.A. were created through the generation of human hybridomas or EBV transformation of B cells, attesting to the technical difficulties and challenges posed by these methods.

[0006] Phage display libraries of human antibody sequences represent another method for producing therapeutic human monoclonal antibodies. This method utilizes recombinant DNA technology to randomly express human antibody heavy- and light-chain sequences together to enable screening for combinations that bind to the target antigen. However, this strategy does not produce affinity-matured antibodies, and antibodies produced in this way usually bind to antigen with low affinity and avidity. Successive mutation and selection/screening steps are then needed to generate high-affinity antibodies.

[0007] Another way to produce therapeutic human monoclonal antibodies is by creating or using transgenic mice that possess a human antibody repertoire. When immunized, such mice produce antibodies that target the immunizing antigen, and hybridomas can then be generated for the production of therapeutic human monoclonal antibodies. Such transgenic mice are proprietary and not commonly available for use in generating human antibodies.

[0008] Thus the inventors have identified a need for compositions, kits, and methods that can, e.g., produce large numbers of affinity-matured human antibodies, avoiding the need for laborious and time-consuming humanization of an antibody, or the need to conduct extensive screens. The compositions, kits, and methods described herein address this need. In addition, the compositions, kits, and methods described herein are broadly applicable outside the human antibody space and can be used in a number of different applications including, e.g., matching together two or more polynucleotides of interest that are derived from a single sample and present in a library of polynucleotides.

SUMMARY

[0009] Disclosed herein is a composition comprising a polynucleotide, wherein the polynucleotide comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises at least one identification region, and wherein the first region is coupled to the second region.

[0010] In some aspects, the variable immunoglobulin region comprises a VDJ region of an IgG immunoglobulin nucleotide sequence isolated from an activated human B cell greater than or equal to 8 μm in diameter, and wherein the 5' end of the immunoglobulin region is coupled to

the 3' end of the identification region. In some aspects, the composition is comprised in a clonal family.

[0011] In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is an activated B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasmablast. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell located in the blood of a subject. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a human activated B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a memory B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasma cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is an antigen–specific B cell. In some aspects, the immunoglobulin region is isolated from a mammalian B cell. In some aspects, the immunoglobulin region is isolated from a human B cell. In some aspects, the immunoglobulin region is isolated from a mouse B cell. In some aspects, the immunoglobulin region is isolated from a B cell from a subject with a disease or condition of interest. In some aspects, the immunoglobulin region is isolated from a B cell from a subject recovering from or recovered from a disease or condition of interest. In some aspects, the immunoglobulin region is isolated from a B cell from a subject administered with at least one antigen of interest. In some aspects, the immunoglobulin region is isolated from a B cell from a subject administered with at least one antigen of interest and an adjuvant. In some aspects, the immunoglobulin region is isolated from a B cell located in the blood of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in the bone marrow of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in the spleen of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in at least one lymph node of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in lymphoid tissue of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in the gut of a subject. In some aspects, the immunoglobulin region is isolated from an activated B cell that is about 8-20 μm in diameter. In some aspects, the immunoglobulin region is isolated from an activated B cell that is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or greater than 20 μm in diameter. In some aspects, the immunoglobulin region is isolated from an activated B cell that is about 60, 70, 80, 90, 100, 120, 130, 140, 150, 200, 250, 300, 350, or greater than 350 μm^2 in area. In some

aspects, the immunoglobulin region is isolated from an activated B cell that is about 250, 268, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, or greater than 4000 μm^3 in volume. In some aspects, the immunoglobulin region is isolated from an activated B cell that has a diameter of 10% or greater in size than the median diameter of a control resting B cell. In some aspects, the immunoglobulin region is isolated from an activated B cell that has a diameter of 15% or greater in size than the median diameter of a control resting B cell. In some aspects, the immunoglobulin region is isolated from an activated B cell that has a diameter of 20% or greater in size than the median diameter of a control resting B cell. In some aspects, the immunoglobulin region is isolated from an activated B cell capable of secreting immunoglobulin. In some aspects, the immunoglobulin region is isolated from a B cell in the gap 1 (G1), synthesis (S), gap 2 (G2), or mitosis (M) phase of the cell cycle. In some aspects, the immunoglobulin region is isolated from a B cell is not in the gap 0 (G0) phase of the cell cycle. In some aspects, the immunoglobulin region is isolated from a B cell characterized as having an FSC greater than 1.2x of the FSC mean of resting B lymphocytes by flow cytometry. In some aspects, the immunoglobulin region is isolated from a B cell characterized as having an FSC mean between 0.7 - 1.15x of the FSC mean of human monocytes by flow cytometry. In some aspects, the immunoglobulin region is isolated from a single CD19 positive B cell. In some aspects, the immunoglobulin region is isolated from a single CD38 positive B cell. In some aspects, the immunoglobulin region is isolated from a single CD27 positive B cell. In some aspects, the immunoglobulin region is isolated from a single CD20 negative B cell. In some aspects, the immunoglobulin region is isolated from a single $CD19^{+}CD20^{-}CD27^{+}CD38^{hi}$ B cell.

[0012] In some aspects, the 5' end of the immunoglobulin region is coupled to the 3' end of the identification region.

[0013] In some aspects, the variable immunoglobulin region comprises a VDJ region of an immunoglobulin nucleotide sequence. In some aspects, the variable immunoglobulin region comprises a VJ region of an immunoglobulin nucleotide sequence. In some aspects, the variable immunoglobulin region comprises a V, D, and/or J region of an immunoglobulin nucleotide sequence. In some aspects, the variable immunoglobulin region comprises a heavy and/or light chain of an immunoglobulin nucleotide sequence. In some aspects, the variable immunoglobulin region comprises an IgG, IgM, IgD, IgE, or IgA immunoglobulin sequence. In some aspects, the variable immunoglobulin region comprises a human IgG1, IgG2, IgG3, or IgG4 immunoglobulin sequence. In some aspects, the variable immunoglobulin region comprises a mouse IgG1, IgG2a, IgG2b, or IgG3 immunoglobulin sequence. In some aspects,

the immunoglobulin region is about 200-2000 nucleotides in length. In some aspects, the immunoglobulin region is less than 200, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, or greater than 2000 nucleotides in length.

[0014] In some aspects, the identification region comprises a plurality of identification regions. In some aspects, the identification region comprises a plurality of identification regions, and wherein each identification region in the plurality has a distinct sequence. In some aspects, the identification region comprises at least one sample identification region and at least one plate identification region. In some aspects, the identification region comprises a sequence distinct from the sequence of the immunoglobulin region. In some aspects, the identification region is about 2 - 100 nucleotides in length. In some aspects, the identification region is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, or greater than 100 nucleotides in length. In some aspects, the identification region is 2 – 1,000 nucleotides in length. In some aspects, the identification region is equal to or greater than 100 nucleotides in length. In some aspects, the identification region comprises a contiguous non-coding nucleotide sequence. In some aspects, the identification region comprises a non-coding nucleotide sequence. In some aspects, the identification region comprises a non-contiguous, non-coding nucleotide sequence. In some aspects, the length of the sequence of the identification region is less than the length of the sequence of the immunoglobulin region.

[0015] In some aspects, a composition described herein can include a third region, wherein the third region comprises an adapter region. In some aspects, the third region comprises an adapter region, and wherein the third region is located between the first region and the second region. In some aspects, the third region comprises an adapter region, and wherein the adapter region comprises at least one G nucleotide located at its 3' end.

[0016] In some aspects, the identification region is 2-100 nucleotides long and has a sequence distinct from the immunoglobulin region sequence, and wherein the adaptor region comprises at least one G nucleotide at its 3' end and is located 3' of the sample identification region and 5' of the immunoglobulin region, and wherein the immunoglobulin variable region has undergone hypermutation and differs from the germline sequence of a naïve B cell.

[0017] In some aspects, the composition is present in a container. In some aspects, a plurality of the compositions are present in a container. In some aspects, a plurality of the compositions are present in a single well of a single plate comprising a plurality of wells.

[0018] In some aspects, the composition is in a library of compositions, wherein each composition is present in a separate container, wherein each composition comprises a

polynucleotide comprising a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, wherein the nucleotide sequence of each identification region o is distinct from the nucleotide sequence of the other identification regions present in the library, and wherein the last nucleotide sequences of a plurality of variable immunoglobulin regions in the library share at least 80-99% sequence identity.

[0019] In some aspects, the composition is comprised in a library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide, wherein the polynucleotide comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, and wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in each separate container in the library.

[0020] Also described herein is a polynucleotide composition library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide, wherein the polynucleotide comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, and wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in each separate container in the library.

[0021] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide of the plurality is present in a separate container, wherein each polynucleotide of the plurality comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in the library, and wherein at least two variable immunoglobulin regions in the plurality share at least 80-99% sequence identity.

[0022] Also described herein is a polynucleotide library comprising a clonal family of polynucleotides, wherein each polynucleotide in the family comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region

and the second region comprises an identification region, wherein the first region is coupled to the second region, wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in the family, and wherein each of the variable immunoglobulin regions in the family exhibit at least 80-99% sequence identity. In some aspects, the library comprises a plurality of clonal families.

[0023] Also described herein is a clonal family of immunoglobulin sequences wherein each sequence in the family is coupled to an identification region. In some aspects, each identification region is distinct from the other identification regions. In some aspects, the immunoglobulin sequences comprise heavy chain immunoglobulin sequences. In some aspects, the immunoglobulin sequences comprise light chain immunoglobulin sequences. In some aspects, the immunoglobulin sequences comprise heavy chain and light chain immunoglobulin sequences. In some aspects, one or more of the identification regions comprise a light chain immunoglobulin sequence. In some aspects, one or more of the identification regions comprise a heavy chain immunoglobulin sequence.

[0024] Also described herein is a set of two or more of the clonal families described herein.

[0025] Also described herein is a set of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more of the clonal families described herein.

[0026] Also described herein is a clonal family of immunoglobulin sequences wherein each sequence in the family is operatively coupled to at least one contiguous nucleotide sequence. In some aspects, the immunoglobulin sequences comprise heavy chain immunoglobulin sequences and the at least one contiguous nucleotide sequence comprises a light chain immunoglobulin sequence. In some aspects, the immunoglobulin sequences comprise light chain immunoglobulin sequences and the at least one contiguous nucleotide sequence comprises a heavy chain immunoglobulin sequence.

[0027] Also described herein is a method of producing a clonal family of immunoglobulin sequences comprising obtaining a plurality of immunoglobulin sequences each having V, D, and/or J regions and each coupled to an identification region; and grouping two or more sequences from the plurality to produce the clonal family wherein each sequence in the clonal family is a mutated version of the same germline immunoglobulin sequence having a V, D, and/or J region or the germline immunoglobulin sequence having the V, D, and/or J region.

[0028] In some aspects, each identification region is distinct from the other identification regions.

[0029] Also described herein is a method of producing a clonal family of immunoglobulin sequences comprising obtaining a plurality of immunoglobulin sequences each having V, D,

and/or J regions and each coupled to an identification region, and wherein each identification region is distinct from the other identification regions; removing one or more identification regions; and grouping two or more sequences from the plurality to produce the clonal family wherein each sequence in the clonal family is a mutated version of the same germline immunoglobulin sequence having a V, D, and/or J region or the germline immunoglobulin sequence having the V, D, and/or J region.

[0030] Also described herein is a method of identifying a second cDNA coupled to a first identification region comprising selecting a first cDNA coupled to the first identification region and identifying the second cDNA based on the shared identity of the identification region coupled to each cDNA.

[0031] Also described herein is a method of producing a 3' tail on a second nucleotide sequence comprising obtaining a first nucleotide sequence and contacting the first nucleotide sequence with a thermal stable RNase H^- reverse transcriptase having template switching activity at less than 50°C, wherein the contacting produces the 3' tail and the second nucleotide sequence. In some aspects, the first nucleotide sequence is contacted at about less than 50, 49, 48, 47, 46, 45 ,44, 43, 42, or less than 42°C. In some aspects, the first nucleotide sequence is contacted at 42°C. In some aspects, the first nucleotide sequence is contacted at 45.5°C. In some aspects, the transcriptase is a Moloney Murine Leukemia Virus (MMLV) RNase H^- reverse transcriptase. In some aspects, the transcriptase is SuperScript III.

[0032] Also described herein is a method for determining the naturally occurring sequence of a first sequence of interest comprising obtaining a plurality of sequences related to the first sequence and each coupled to a first identification region, wherein each first identification region is identical, and wherein one or more of the sequences in the plurality is distinct from the naturally occurring sequence; and comparing the sequences in the plurality to determine the naturally occurring sequence of the first sequence of interest. In some aspects, the plurality of sequences comprise immunoglobulin sequences. In some aspects, the plurality of sequences comprise immunoglobulin sequences. In some aspects, the plurality of sequences comprise immunoglobulin sequences. In some aspects, the plurality of sequences are each coupled to a second identification region and each second identification region is identical. In some aspects, the first sequence of interest is an immunoglobulin sequence. In some aspects, the plurality of sequences are immunoglobulin sequences.

[0033] Also described herein is a composition comprising a polynucleotide comprising a first region and a second region, wherein the first region comprises a B cell-derived variable

immunoglobulin region and the second region comprises an identification region, and wherein the first region is coupled to the second region.

[0034] Also described herein is a polynucleotide composition library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide comprising a B cell-derived variable immunoglobulin region and an identification region, wherein the variable immunoglobulin region is coupled to the identification region, wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in each separate container in the library.

[0035] Also described herein is a method for producing a polynucleotide composition, comprising: obtaining a polynucleotide comprising a first region, wherein the first region comprises an expressed B cell variable immunoglobulin region associated with a subject; and generating the polynucleotide composition comprising the first region and a second region by coupling the first region to the second region, wherein the second region comprises an identification region.

[0036] In some aspects, obtaining the polynucleotide comprises obtaining a B cell associated with the subject and processing the cell to prepare the polynucleotide. In some aspects, obtaining the polynucleotide comprises receiving the polynucleotide directly or indirectly from a third party that has processed a B cell associated with the subject to prepare the polynucleotide. In some aspects, obtaining the polynucleotide comprises receiving the polynucleotide directly or indirectly from a third party that has solubilized a B cell associated with the subject to prepare the polynucleotide. In some aspects, obtaining the polynucleotide comprises obtaining a B cell using a flow cytometer. In some aspects, obtaining the polynucleotide comprises obtaining a B cell using a microfluidic device.

[0037] In some aspects, the variable immunoglobulin region comprises a VDJ region of an IgG immunoglobulin nucleotide sequence isolated from an activated human B cell greater than or equal to 8 μm in diameter, and wherein the 5' end of the immunoglobulin region is coupled to the 3' end of the identification region. In some aspects, the composition is comprised in a clonal family.

[0038] In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is an activated B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasmablast. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell. In some

aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell located in the blood of a subject. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a human activated B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a memory B cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasma cell. In some aspects, the immunoglobulin region is isolated from a B cell, and wherein the B cell is an antigen–specific B cell. In some aspects, the immunoglobulin region is isolated from a mammalian B cell. In some aspects, the immunoglobulin region is isolated from a human B cell. In some aspects, the immunoglobulin region is isolated from a mouse B cell. In some aspects, the immunoglobulin region is isolated from a B cell from a subject with a disease or condition of interest. In some aspects, the immunoglobulin region is isolated from a B cell from a subject recovering from or recovered from a disease or condition of interest. In some aspects, the immunoglobulin region is isolated from a B cell from a subject administered with at least one antigen of interest. In some aspects, the immunoglobulin region is isolated from a B cell from a subject administered with at least one antigen of interest and an adjuvant. In some aspects, the immunoglobulin region is isolated from a B cell located in the blood of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in the bone marrow of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in the spleen of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in at least one lymph node of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in lymphoid tissue of a subject. In some aspects, the immunoglobulin region is isolated from a B cell located in the gut of a subject. In some aspects, the immunoglobulin region is isolated from an activated B cell that is about 8-20 μm in diameter. In some aspects, the immunoglobulin region is isolated from an activated B cell that is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or greater than 20 μm in diameter. In some aspects, the immunoglobulin region is isolated from an activated B cell that is about 60, 70, 80, 90, 100, 120, 130, 140, 150, 200, 250, 300, 350, or greater than 350 μm^2 in area. In some aspects, the immunoglobulin region is isolated from an activated B cell that is about 250, 268, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, or greater than 4000 μm^3 in volume. In some aspects, the immunoglobulin region is isolated from an activated B cell that has a diameter of 10% or greater in size than the median diameter of a control resting B cell. In some aspects, the immunoglobulin region is isolated from an activated B cell that has a diameter of 15% or greater in size than the median diameter of a control resting B cell. In some aspects, the immunoglobulin region is isolated from an activated B cell that has a diameter of 20% or greater

in size than the median diameter of a control resting B cell. In some aspects, the immunoglobulin region is isolated from an activated B cell capable of secreting immunoglobulin. In some aspects, the immunoglobulin region is isolated from a B cell in the gap 1 (G1), synthesis (S), gap 2 (G2), or mitosis (M) phase of the cell cycle. In some aspects, the immunoglobulin region is isolated from a B cell is not in the gap 0 (G0) phase of the cell cycle. In some aspects, the immunoglobulin region is isolated from a B cell characterized as having an FSC greater than 1.2x of the FSC mean of resting B lymphocytes by flow cytometry. In some aspects, the immunoglobulin region is isolated from a B cell characterized as having an FSC mean between 0.7 - 1.15x of the FSC mean of human monocytes by flow cytometry. In some aspects, the immunoglobulin region is isolated from a single CD19 positive B cell. In some aspects, the immunoglobulin region is isolated from a single CD38 positive B cell. In some aspects, the immunoglobulin region is isolated from a single CD27 positive B cell. In some aspects, the immunoglobulin region is isolated from a single CD20 negative B cell. In some aspects, the immunoglobulin region is isolated from a single $CD19^{+}CD20^{-}CD27^{+}CD38^{hi}$ B cell.

[0039] In some aspects, the 5' end of the immunoglobulin region is coupled to the 3' end of the identification region.

[0040] In some aspects, the variable immunoglobulin region comprises a VDJ region of an immunoglobulin nucleotide sequence. In some aspects, the variable immunoglobulin region comprises a VJ region of an immunoglobulin nucleotide sequence. In some aspects, the variable immunoglobulin region comprises a V, D, and/or J region of an immunoglobulin nucleotide sequence. In some aspects, the variable immunoglobulin region comprises a heavy and/or light chain of an immunoglobulin nucleotide sequence. In some aspects, the variable immunoglobulin region comprises an IgG, IgM, IgD, IgE, or IgA immunoglobulin sequence. In some aspects, the variable immunoglobulin region comprises a human IgG1, IgG2, IgG3, or IgG4 immunoglobulin sequence. In some aspects, the variable immunoglobulin region comprises a mouse IgG1, IgG2a, IgG2b, or IgG3 immunoglobulin sequence. In some aspects, the immunoglobulin region is about 200-2000 nucleotides in length. In some aspects, the immunoglobulin region is less than 200, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, or greater than 2000 nucleotides in length.

[0041] In some aspects, the identification region comprises a plurality of identification regions. In some aspects, the identification region comprises a plurality of identification regions, and wherein each identification region in the plurality has a distinct sequence. In some aspects, the

identification region comprises at least one sample identification region and at least one plate identification region. In some aspects, the identification region comprises a sequence distinct from the sequence of the immunoglobulin region. In some aspects, the identification region is about 2 - 100 nucleotides in length. In some aspects, the identification region is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, or greater than 100 nucleotides in length. In some aspects, the identification region is 2 – 1,000 nucleotides in length. In some aspects, the identification region is equal to or greater than 100 nucleotides in length. In some aspects, the identification region comprises a contiguous non-coding nucleotide sequence. In some aspects, the identification region comprises a non-coding nucleotide sequence. In some aspects, the identification region comprises a non-contiguous, non-coding nucleotide sequence. In some aspects, the length of the sequence of the identification region is less than the length of the sequence of the immunoglobulin region.

[0042] In some aspects, a composition described herein can include a third region, wherein the third region comprises an adapter region. In some aspects, the third region comprises an adapter region, and wherein the third region is located between the first region and the second region. In some aspects, the third region comprises an adapter region, and wherein the adapter region comprises at least one G nucleotide located at its 3' end.

[0043] In some aspects, the identification region is 2-100 nucleotides long and has a sequence distinct from the immunoglobulin region sequence, and wherein the adaptor region comprises at least one G nucleotide at its 3' end and is located 3' of the sample identification region and 5' of the immunoglobulin region, and wherein the immunoglobulin variable region has undergone hypermutation and differs from the germline sequence of a naïve B cell.

[0044] In some aspects, the composition is present in a container. In some aspects, a plurality of the compositions are present in a container. In some aspects, a plurality of the compositions are present in a single well of a single plate comprising a plurality of wells.

[0045] In some aspects, the composition is in a library of compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide comprising a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, wherein the nucleotide sequence of each identification region o is distinct from the nucleotide sequence of the other identification regions present in the library, and wherein the last nucleotide sequences of a plurality of variable immunoglobulin regions in the library share at least 80-99% sequence identity.

[0046] In some aspects, the composition is comprised in a library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide, wherein the polynucleotide comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, and wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in each separate container in the library.

[0047] Also described herein is a method for producing a polynucleotide composition, comprising: obtaining a B cell associated with a subject; isolating polynucleotides from the cell comprising an expressed B cell variable immunoglobulin region; and generating the polynucleotide composition comprising the variable immunoglobulin region and an identification region by coupling the variable immunoglobulin region to the identification region.

[0048] Also described herein is a method for producing a polynucleotide composition, comprising: obtaining a polynucleotide comprising a B cell-derived variable immunoglobulin region associated with a subject; and generating the polynucleotide composition comprising the variable immunoglobulin region and an identification region by coupling the variable immunoglobulin region to the identification region.

[0049] In some aspects, obtaining the polynucleotide comprises obtaining a B cell and processing the cell to prepare the polynucleotide. In some aspects, obtaining the polynucleotide comprises receiving the polynucleotide directly or indirectly from a third party that has processed a B cell to prepare the polynucleotide.

[0050] Also described herein is a method for producing two or more polynucleotide compositions, comprising: obtaining a polynucleotide library comprising a plurality of polynucleotides associated with a plurality of samples obtained from one or more subjects, wherein one or more polynucleotides comprises an expressed B cell variable immunoglobulin region, wherein each sample is associated with a B cell, and wherein each polynucleotide associated with each sample is present in a separate container; and generating two or more polynucleotide compositions each comprising a polynucleotide from the plurality of polynucleotides and an identification region by coupling the polynucleotide to the identification region, wherein the sequence of each identification region is distinct from the sequence of the identification regions coupled to the other polynucleotides in the library.

[0051] In some aspects, obtaining the polynucleotide library comprises obtaining a plurality of B cells and processing the cells to prepare the polynucleotide library. In some aspects, obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has processed a plurality of B cells to prepare the polynucleotide library.

[0052] Also described herein is a method for producing two or more polynucleotide compositions, comprising:obtaining a polynucleotide library comprising a plurality of polynucleotides associated with a plurality of samples obtained from one or more subjects, wherein one or more polynucleotides comprises a B cell-derived variable immunoglobulin region, and wherein each polynucleotide associated with each sample is present in a separate container; and generating two or more polynucleotide compositions each comprising a polynucleotide from the plurality of polynucleotides and an identification region by coupling the polynucleotide to the identification region, wherein the sequence of each identification region is distinct from the sequence of the identification regions coupled to the other polynucleotides in the library.

[0053] In some aspects, obtaining the polynucleotide library comprises obtaining a plurality of B cells and processing the cells to prepare the polynucleotide library. In some aspects, obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has processed a plurality of B cells to prepare the polynucleotide library.

[0054] Also described herein is a polynucleotide composition library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a single sample-derived cDNA region comprising the nucleotide C at the 3' end of the cDNA region and a sample identification-adapter region comprising a sample identification region coupled to an adapter region, wherein the nucleotide sequence of the sample identification region of each sample identification-adapter region is distinct from the nucleotide sequence of the sample identification region of the other sample identification-adapter regions present in each separate container in the library, wherein the adapter region comprises the nucleotide G at the 3' end of the adapter region, and wherein the sample identification-adapter region is attached to the cDNA region by binding between the C and G.

[0055] In some aspects, the cDNA region comprises an RNA polynucleotide hybridized to a DNA polynucleotide. In some aspects, the cDNA region comprises an mRNA polynucleotide hybridized to a cDNA polynucleotide. In some aspects, the cDNA region comprises at least one C at the 3' end and wherein the adapter region comprises at least one G at the 3' end.

[0056] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a sample identification region, an adapter region, and a single sample-derived cDNA region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the cDNA region is coupled to the 3' end of the adapter region, wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample, and wherein the sample identification region is double-stranded. In some aspects, each polynucleotide comprises a plurality of sample identifications regions.

[0057] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the amplicon region is operatively coupled to the adapter region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

[0058] In some aspects, the 5' end of the amplicon region is coupled to the 3' end of the adapter region, wherein the universal primer region comprises the sequence CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G. In some aspects, the sequence of the universal primer region is not fully complementary to any human gene exon, and wherein the universal primer region has minimal secondary structure that does not interfere with the adapter region. In some aspects, the universal primer region is the sequence CACGACCGGTGCTCGATTTAG. In some aspects, the amplicon region comprises a cDNA region comprising a cDNA nucleotide sequence. In some aspects, the sequence of the sample identification region of each polynucleotide from a first single sample differs by at least 1 nucleotide from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample. In some aspects, the sequence of the sample identification region of each polynucleotide from a first single sample differs by at least 2 nucleotides from the sequence of the sample identification region of

the other polynucleotides in the library from one or more samples distinct from the first single sample. In some aspects, the sample identification region is at least 10 nucleotides in length. In some aspects, the sample identification region is at least 1 nucleotide in length. In some aspects, the sequence of each sample identification region is selected from Tables 2 and 7. In some aspects, the sequence of the adapter region comprises at least one G nucleotide at its 3' end. In some aspects, the amplicon region comprises an immunoglobulin heavy chain amplicon sequence, an immunoglobulin light chain amplicon sequence, a T cell receptor alpha amplicon sequence, or a T cell receptor beta amplicon sequence.

[0059] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-3', wherein A is a universal primer region, wherein B is a sample identification region, wherein C is an adapter region, wherein D is an amplicon region from a single sample, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

[0060] Also described herein is a polynucleotide comprising a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the amplicon region is operatively coupled to the adapter region.

[0061] In some aspects, the 5' end of the amplicon region is coupled to the 3' end of the adapter region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

[0062] Also described herein is a polynucleotide comprising the sequence 5'-A-B-C-D-3', wherein A is a universal primer region, wherein B is a sample identification region, wherein C is an adapter region, and wherein D is an amplicon region from a single sample.

[0063] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a first plate identification region, a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first plate identification region is

operatively coupled to the universal primer region, wherein the amplicon region is operatively coupled to the adapter region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

[0064] In some aspects, the sequence of the first plate identification region of each polynucleotide from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the first plate identification region of each polynucleotide from the first set of single samples differs by at least 1 nucleotide from the sequence of the first plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the first plate identification region of each polynucleotide from the first set of single samples differs by at least 2 nucleotides from the sequence of the first plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples. In some aspects, the first plate identification region is at least 10 nucleotides in length. In some aspects, the sequence of the first plate identification region is selected from Tables 3 and 6. In some aspects, the 3' end of the first plate identification region is coupled to the 5' end of the universal primer region, wherein the 5' end of the amplicon region is coupled to the 3' end of the adapter region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, wherein the adapter region comprises at least one G, wherein each polynucleotide further comprises a second plate identification region, a first sequencing region, and a second sequencing region, wherein the 5' end of the second plate identification region is coupled to the 3' end of the amplicon region, wherein the 3' end of the first sequencing region is coupled to the 5' end of the first plate identification region, and wherein the 5' end of the second sequencing region is coupled to the 3' end of the second plate identification region. In some aspects, the sequence of the second plate identification region is identical to the sequence of the first plate identification region on each polynucleotide. In some aspects, the sequence of the second plate identification region of each polynucleotide from a first set of single samples is distinct from the sequence of the second plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the second plate identification region of each polynucleotide from the first set of single samples

differs by at least 1 nucleotide from the sequence of the second plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the second plate identification region of each polynucleotide from the first set of single samples differs by at least 2 nucleotides from the sequence of the second plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples. In some aspects, the second plate identification region is at least 10 nucleotides in length. In some aspects, the sequence of the second plate identification region is selected from Tables 3 and 6. In some aspects, the first sequencing region comprises GAGAGACTGACAGCGTATCGCCTCCCTCGCGCCATCAG. In some aspects, the second sequencing region comprises CTATGCGCCTTGCCAGCCCGCTCAG.

[0065] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-E-3', wherein A is a plate identification region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, wherein E is an amplicon region from a single sample, and wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample

[0066] Also described herein is a polynucleotide comprising a first plate identification region, a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first plate identification region is operatively coupled to the universal primer region, and wherein the amplicon region is operatively coupled to the adapter region.

[0067] In some aspects, the 3' end of the first plate identification region is coupled to the 5' end of the universal primer region, wherein the 5' end of the amplicon region is coupled to the 3' end of the adapter region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, wherein the adapter region comprises at least one G, wherein each polynucleotide further comprises a second plate identification region, a first sequencing region, and a second sequencing region, wherein the 5' end of the second plate identification region is coupled to the 3' end of the amplicon region, wherein the 3' end of the

first sequencing region is coupled to the 5' end of the first plate identification region, and wherein the 5' end of the second sequecing region is coupled to the 3' end of the second plate identification region.

[0068] Also described herein is a polynucleotide comprising the sequence 5'-A-B-C-D-E-3', wherein A is a plate identification region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, and wherein E is an amplicon region from a single sample.

[0069] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a first restriction site region, a universal primer region, a sample identification region, an adapter region, an amplicon region from a single sample, and a second restriction site region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first restriction site region is operatively coupled to the universal primer region, wherein the amplicon region is operatively coupled to the adapter region, wherein the second restriction site region is operatively coupled to the amplicon region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

[0070] In some aspects, the first restriction site region comprises one or more restriction sites. In some aspects, the first restriction site region comprises one or more restriction sites selected from the group consisting of: NheI, XhoI, BstBI, EcoRI, SacII, BbvCI, PspXI, AgeI, ApaI, KpnI, Acc65I, XmaI, BstEII, DraIII, PacI, FseI, AsiSI and AscI. In some aspects, the second restriction site region comprises one or more restriction sites. In some aspects, the second restriction site region comprises one or more restriction sites selected from the group consisting of: NheI, XhoI, BstBI, EcoRI, SacII, BbvCI, PspXI, AgeI, ApaI, KpnI, Acc65I, XmaI, BstEII, DraIII, PacI, FseI, AsiSI and AscI. In some aspects, the 3' end of the first restriction site region is coupled to the 5' end of the universal primer region, wherein the 3' end of the adapter region is coupled to the 5' end of the amplicon region, wherein the 3' end of the amplicon region is coupled to the 5' end of the second restriction site region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

[0071] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-E-F-3', wherein A is a first restriction site region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, wherein E is an amplicon region from a single sample, wherein F is a second restriction site region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

[0072] Also described herein is a polynucleotide for insertion into a vector, comprising a first restriction site region, a universal primer region, a sample identification region, an adapter region, an amplicon region from a single sample, and a second restriction site region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first restriction site region is operatively coupled to the universal primer region, wherein the amplicon region is operatively coupled to the adapter region, and wherein the second restriction site region is operatively coupled to the amplicon region.

[0073] In some aspects, the 3' end of the first restriction site region is coupled to the 5' end of the universal primer region, wherein the 3' end of the adapter region is coupled to the 5' end of the amplicon region, wherein the 3' end of the amplicon region is coupled to the 5' end of the second restriction site region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

[0074] Also described herein is a polynucleotide for insertion in a vector, comprising the sequence 5'-A-B-C-D-E-F-3', wherein A is a first restriction site region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, wherein E is an amplicon region from a single sample, and wherein F is a second restriction site region.

[0075] Also described herein is a polynucleotide adapter molecule, comprising a universal primer region, a sample identification region, and an adapter region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, and wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region. In some aspects, the universal primer region comprises CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

[0076] Also described herein is a polynucleotide primer, comprising a universal primer region and a plate identification region, and wherein the 3' end of the plate identification region is coupled to the 5' end of the universal primer region. In some aspects, the universal primer region comprises CACGACCGGTGCTCGATTTAG, wherein the primer further comprises a sequencing region, and wherein the 3' end of the sequencing region is coupled to the 5' end of the plate identification region.

[0077] Also described herein is a vector comprising a polynucleotide described herein. In some aspects, the vector comprises a plurality of polynucleotides. In some aspects, the vector is selected from the group consisting of: pEE6.4 and pEE12.4

[0078] Also described herein is an isolated host cell comprising a vector described herein or a polynucleotide described herein. In some aspects, the host cell is selected from the group consisting of: CHO cells, CHO-K1 cells, CHO-S cells, NS0 cells, CHO cells that are dhfr-, CHO-dhfr-, DUKX-B11 CHO cells, and DG44 CHO cells.

[0079] Also described herein is a method for producing one or more polynucleotides of interest, comprising: obtaining a cDNA library comprising a plurality of cDNAs associated with a plurality of samples obtained from one or more subjects, wherein each cDNA is associated with a single sample in the plurality of samples, and wherein each cDNA associated with each sample is present in a separate container; and adding an adapter molecule to the cDNA associated with each sample to produce the one or more polynucleotides of interest, wherein the adapter molecule comprises a sample identification region and an adapter region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the sequence of the sample identification region of each adapter molecule is distinct from the sequence of the sample identification region of the other adapter molecules added to each cDNA in the library.

[0080] In some aspects, the method further includes allowing the 3' end of the adapter region to attach to the 3' end of each cDNA in the library to produce the one or more polynucleotides of interest. In some aspects, obtaining the cDNA library comprises obtaining the plurality of samples and processing the samples to prepare the cDNA library. In some aspects, the adapter molecule further comprises a universal primer region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region. In some aspects, each cDNA region comprises an mRNA polynucleotide hybridized to a cDNA polynucleotide. In some aspects, each sample comprises a cell. In some aspects, the cell is a B cell. In some aspects, the B cell is a plasmablast, memory B cell, or a plasma cell. In some aspects, each sample comprises a plurality of cells. In some aspects, obtaining the cDNA library comprises receiving

the cDNA library directly or indirectly from a third party that has processed the plurality of samples to prepare the cDNA library. In some aspects, the adaptor is added by annealing the adaptor to the '3 tail of a cDNA generated during a reverse transcription reaction. In some aspects, each cDNA comprises at least one C nucleotide, wherein C is located at the 3' end of each cDNA, wherein the adapter region comprises at least one G nucleotide, wherein G is located at the 3' end of the adapter region, and wherein the adapter region is attached to each cDNA via binding between the G and C. In some aspects, the adapter molecule is single-stranded, and further comprising incorporating the adapter molecule into each cDNA by allowing an enzyme to make the adapter molecule double-stranded. In some aspects, the adapter molecule is incorporated into each cDNA to produce the polynucleotide of interest by an MMLV H$^-$ reverse transcriptase.

[0081] Also described herein is a method of producing one or more polynucleotides of interest for sequencing, comprising: obtaining a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the amplicon region is operatively coupled to the adapter region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample; and amplifying the polynucleotide library with a set of primers to produce the one or more polynucleotides of interest for sequencing, wherein the one or more polynucleotides of interest for sequencing comprises a first sequencing region, a first plate identification region, a universal primer region, a sample identification region, an adapter region, an amplicon region from a single sample, and a second sequencing region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first plate identification region is operatively coupled to the universal primer region, wherein the amplicon region is operatively coupled to the adapter region, wherein the first sequencing region is located at the 5' end of the polynucleotide of interest, and wherein the second sequencing region is located at the 3' end of the polynucleotide of interest.

[0082] In some aspects, the method further includes sequencing the one or more polynucleotides of interest. In some aspects, the method further includes sequencing the one or more polynucleotides of interest with 454 sequencing. In some aspects, the method further includes sequencing the one or more polynucleotides of interest with SMRT sequencing. In some aspects, the method further includes sequencing the one or more polynucleotides of interest with SOLiD sequencing. In some aspects, the method further includes sequencing the one or more polynucleotides of interest with SOLEXA sequencing. In some aspects, the method further includes sequencing the one or more polynucleotides of interest with tSMS sequencing. In some aspects, the set of primers is selected from the primers shown in Tables 1 and 5. In some aspects, obtaining the polynucleotide library comprises preparing the polynucleotide library in a laboratory. In some aspects, obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has prepared the polynucleotide library.

[0083] Also described herein is a method for analyzing sequencing data, comprising: obtaining a dataset associated with a plurality of polynucleotides, wherein the dataset comprises sequencing data for the plurality of polynucleotides, wherein each polynucleotide in the plurality of polynucleotides comprises a sample identification region, and wherein each sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the plurality of polynucleotides from one or more samples distinct from the first single sample; and analyzing the dataset to match together polynucleotides with identical sample identification regions, wherein a match indicates that the polynucleotides originated from the same sample.

[0084] In some aspects, each polynucleotide in the plurality of polynucleotides further comprises a first plate identification region, wherein each combination of each first plate identification region and sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the first plate identification region of each polynucleotide from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the plurality of polynucleotides from one or more single sample sets distinct from the first set of single samples, and further comprising analyzing the dataset to match together polynucleotides with identical first plate identification regions and identical sample identification regions, wherein a match between both regions indicates that the polynucleotides originated from the same sample. In some aspects, obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of

polynucleotides to experimentally determine the dataset. In some aspects, obtaining the dataset comprises receiving the dataset directly or indirectly from a third party that has sequenced the plurality of polynucleotides to experimentally determine the dataset. In some aspects, the dataset is stored on an electronic storage medium. In some aspects, the single sample is a single cell. In some aspects, the single sample comprises a single cell. In some aspects, the single sample comprises a single B cell. In some aspects, the single sample comprises a plurality of B cells. In some aspects, further comprising selecting one or more polynucleotides for cloning.

[0085] Also described herein is a method for identifying a second polynucleotide of interest based on selection of a first polynucleotide of interest, comprising: obtaining a dataset associated with a plurality of polynucleotides, wherein the dataset comprises sequencing data for the plurality of polynucleotides, wherein each polynucleotide in the plurality of polynucleotides comprises a sample identification region, and wherein each sample identification region on each polynucleotide is unique to a single sample thereby associating each polynucleotide in the plurality of polynucleotides with a distinct single sample, wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the plurality of polynucleotides from one or more samples distinct from the first single sample; and selecting a first polynucleotide of interest associated with a first single sample from the dataset and identifying a second polynucleotide of interest in the first single sample based on the sample identification region of the first polynucleotide of interest.

[0086] In some aspects, each polynucleotide in the plurality of polynucleotides further comprises a first plate identification region, wherein each combination of each first plate identification region and sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the first plate identification region of each polynucleotide from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the plurality of polynucleotides from one or more single sample sets distinct from the first set of single samples, and further comprising identifying a second polynucleotide of interest in the first single sample based on the sample identification region and first plate identification region of the first polynucleotide of interest. In some aspects, the first single sample comprises a B cell. In some aspects, the first single sample comprises a plurality of B cells. In some aspects, the first single sample comprises a B cell, wherein the first polynucleotide of interest comprises an antibody heavy chain nucleotide sequence, and wherein the second polynucleotide of interest comprises an antibody light chain nucleotide sequence. In some aspects, the first single sample comprises a B cell, wherein the

first polynucleotide of interest comprises an antibody light chain nucleotide sequence, and wherein the second polynucleotide of interest comprises an antibody heavy chain nucleotide sequence. In some aspects, obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of polynucleotides to experimentally determine the dataset. In some aspects, obtaining the dataset comprises receiving the dataset directly or indirectly from a third party that has sequenced the plurality of polynucleotides to experimentally determine the dataset. In some aspects, the dataset is stored on an electronic storage medium.

[0087] Also described herein is a method of producing one or more polynucleotides of interest for cloning, comprising: obtaining a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the amplicon region is operatively coupled to the adapter region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample; and amplifying the polynucleotide library with a set of primers to produce the one or more polynucleotides of interest for cloning, wherein the one or more polynucleotides of interest for cloning comprises a first restriction site region, a universal primer region, a sample identification region, an adapter region, an amplicon region from a single sample, and a second restriction site region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the amplicon region is operatively coupled to the adapter region, wherein the first restriction site region is located at the 5' end of the polynucleotide of interest, and wherein the second restriction site region is located at the 3' end of the polynucleotide of interest.

[0088] In some aspects, obtaining the polynucleotide library comprises preparing the polynucleotide library in a laboratory. In some aspects, obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has prepared the polynucleotide library.

[0089] Also described herein is a method of producing a molecule of interest, comprising: obtaining a host cell comprising a polynucleotide comprising a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the amplicon region is operatively coupled to the adapter region; and culturing the host cell under conditions sufficient to produce the molecule of interest. In some aspects, obtaining the host cell comprises preparing the host cell comprising the polynucleotide in a laboratory. In some aspects, obtaining the host cell comprises receiving the host cell comprising the polynucleotide directly or indirectly from a third party that has prepared the host cell. In some aspects, the molecule of interest is a protein. In some aspects, the molecule of interest is an antibody. In some aspects, the molecule of interest is a human monoclonal antibody. In some aspects, further comprising collecting the molecule of interest.

[0090] Also described herein is a kit, comprising a polynucleotide, a polynucleotide library, a vector, or a host cell described herein and instructions for use.

[0091] Also described herein is a method of linking and barcoding a plurality of non-contiguous polynucleotide sequences of interest, said method comprising: (a) providing a plurality of cDNA molecules; (b) physically linking cDNA molecules of interest; and (c) adding a barcode sequence to the cDNAs of interest prior to, during, or after physical linkage.

[0092] In some aspects, the physical linking is by ligation. In some aspects, the physical linking is by recombination. In some aspects, the physical linking comprises using an overlap-extension sequence. In some aspects, the barcode sequence is located at one or both of the ends of the physically linked cDNAs. In some aspects, the barcode sequence is located in between the physically linked cDNAs. In some aspects, the ligation is performed by annealing and ligation of compatible ends. In some aspects, the compatible ends are a restriction site. In some aspects, the ligation is performed by blunt end ligation. In some aspects, the overlap-extension sequence is added during the course of amplification using a primer comprising the overlap-extension tail. In some aspects, the overlap-extension sequence is added during the course of reverse transcription using a primer comprising the overlap-extension tail. In some aspects, the overlap-extension sequence is added by annealing an adaptor to the 3' tail of a cDNA generated during a reverse transcription reaction. In some aspects, the barcode sequence is added by ligation. In some aspects, the ligation is performed by annealing and ligation of compatible ends. In some aspects, the compatible ends are a restriction site. In some aspects, the ligation is performed by blunt end ligation of an adaptor comprising the barcode sequence. In some aspects, the barcode sequence is added during the course of an amplification reaction using a primer comprising the

barcode sequence. In some aspects, the barcode sequence is added during the course of a reverse transcription reaction using a primer comprising the barcode sequence. In some aspects, the barcode sequence is added by annealing an adaptor to the 3' tail of a cDNA generated during a reverse transcription reaction. In some aspects, the '3 end of the cDNA comprises at least one C nucleotide, and wherein the 3' end of the adaptor comprises at least one G nucleotide, and wherein the adaptor is attached to each cDNA via binding between the C and G. In some aspects, the adaptor is single-stranded, and further comprising incorporating the adaptor into each cDNA by allowing an enzyme to make the adaptor double-stranded. In some aspects, the adaptor is incorporated into each cDNA by an MMLV H- reverse transcriptase. In some aspects, the overlap-extension sequence comprises a barcode sequence. In some aspects, the polynucleotide sequences of interest comprise antibody heavy and light chains. In some aspects, further comprising (d) adding a sequencing region to the cDNAs of interest prior to, during, or after physical linkage. In some aspects, the sequencing region is added with an adaptor. In some aspects, further comprising (e) sequencing of the physically linked cDNA molecules of interest using a NextGen sequencing platform. In some aspects, the NextGen sequencing platform is 454 sequencing. In some aspects, the NextGen sequencing platform is SMRT sequencing. In some aspects, the NextGen sequencing platform is SOLiD sequencing. In some aspects, the NextGen sequencing platform is SOLEXA sequencing. In some aspects, the NextGen sequencing platform is tSMS sequencing. In some aspects, the plurality of cDNA molecules is from single samples contained in a plate with at least 6 wells, at least 12 wells, at least 24 wells, at least 48 wells, at least 96 wells, at least 384 wells, at least 1536 wells, or more wells. In some aspects, the plurality of cDNA molecules is from single samples contained in at least one, two, three, four, five, six, seven, eight, nine, ten, twenty, thirty, forty, fifty, seventy five, one hundred, or more plates with at least 96 wells each.

[0093] Also described herein is a method of linking and barcoding a plurality of samples containing polynucleotide sequences of interest, said method comprising: (a) distributing the samples into a plurality of containers; (b) synthesizing polynucleotide sequences of interest using templates from the sample, wherein said synthesis results in the addition of a barcode sequence; and (c) effecting linkage of the polynucleotide sequences of interest synthesized in step (b).

[0094] In some aspects, each sample comprises a cell. In some aspects, the cell is a B cell. In some aspects, the B cell is a plasmablast, memory B cell, or a plasma cell. In some aspects, each sample comprises a plurality of cells. In some aspects, the polynucleotide sequences of interest comprise antibody heavy and light chains. In some aspects, said synthesis comprises an

RT-PCR amplification. In some aspects, said RT-PCR amplification is performed in a single step. In some aspects, said linkage of the polynucleotide of interest is performed during the course of an RT-PCR amplification using an overlap-extension primer. In some aspects, further comprising (d) adding a sequencing region to the polynucleotide sequences of interest prior to, during, or after barcode sequence addition or linkage. In some aspects, the sequencing region is added with an adaptor. In some aspects, further comprising (e) sequencing of the linked polynucleotide sequences of interest using a NextGen sequencing platform. In some aspects, the NextGen sequencing platform is 454 sequencing. In some aspects, the NextGen sequencing platform is SMRT sequencing. In some aspects, the NextGen sequencing platform is SOLiD sequencing. In some aspects, the NextGen sequencing platform is SOLEXA sequencing. In some aspects, the NextGen sequencing platform is tSMS sequencing. In some aspects, the plurality of samples are single samples contained in a plate with at least 6 wells, at least 12 wells, at least 24 wells, at least 48 wells, at least 96 wells, at least 384 wells, at least 1536 wells, or more wells. In some aspects, the plurality of samples are single samples contained in at least one, two, three, four, five, six, seven, eight, nine, ten, twenty, thirty, forty, fifty, seventy five, one hundred, two hundred, five hundred or more plates with at least 96 wells each.

[0095] Also described herein is a method of linking and barcoding a plurality of non-contiguous polynucleotide sequences of interest, said method comprising: (a) distributing cells into a plurality of containers to obtain isolated one or more cells; (b) amplifying polynucleotide sequences of interest using templates from said isolated one or more cells, wherein said amplification results in the addition of a barcode sequence; and (c) effecting linkage of the polynucleotide sequences of interest amplified in step (b).

[0096] In some aspects, the nucleotide sequences of interest comprise antibody heavy and light chains. In some aspects, said amplification comprises an RT-PCR amplification. In some aspects, said RT-PCR amplification is performed in a single step. In some aspects, said linkage of the nucleotide of interest is performed during the course of an RT-PCR amplification using an overlap-extension primer. In some aspects, further comprising (d) adding a sequencing region to the polynucleotide sequences of interest prior to, during, or after barcode sequence addition or linkage. In some aspects, the sequencing region is added with an adaptor. In some aspects, further comprising (e) sequencing of the linked polynucleotide sequences of interest using a NextGen sequencing platform. In some aspects, the NextGen sequencing platform is 454 sequencing. In some aspects, the NextGen sequencing platform is SMRT sequencing. In some aspects, the NextGen sequencing platform is SOLiD sequencing. In some aspects, the NextGen sequencing platform is SOLEXA sequencing. In some aspects, the NextGen sequencing

platform is tSMS sequencing. In some aspects, the one or more cells are contained in a plate with at least 6 wells, at least 12 wells, at least 24 wells, at least 48 wells, at least 96 wells, at least 384 wells, at least 1536 wells, or more wells. In some aspects, the one or more cells are contained in at least one, two, three, four, five, six, seven, eight, nine, ten, twenty, thirty, forty, fifty, seventy five, one hundred, or more plates with at least 96 wells each.

[0097] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-3', wherein A is a sample identification region (barcode sequence), wherein B is a first cDNA region from a single sample, wherein C is a linker region, wherein D is a second cDNA region from the same single sample, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

[0098] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-3', wherein A is a first cDNA region from a single sample, wherein B is a linker region, wherein C is a second cDNA region from the same single sample, wherein D is a sample identification region (barcode sequence), and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

[0099] Also described herein is a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-3', wherein A is a first cDNA region from a single sample, wherein B is a linker region comprising a sample identification region (barcode sequence), wherein C is a second cDNA region from the same single sample, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

[00100] In some aspects, said first cDNA region comprises an antibody heavy chain and said second cDNA region comprises an antibody light chain. In some aspects, the library comprises at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 3000, at least 10,000, at least 30,000, at least 100,000, at least 300, 000, at least 1,000,000, at least 3,000,000, at least 10,000,000, at least 30,000,000, or more members. In some aspects, the

library comprises at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 3000, at least 10,000, at least 30,000, or more genes of a cell sample's whole transcriptome. In some aspects, the library comprises at least 1, at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000 or more of the different antibody species present in the blood of an individual. In some aspects, the antibodies are expressed by plasmablasts, plasma cells, memory B cells, long-lived plasma cells, other B lineage cells or combinations thereof.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[00101] These and other features, aspects, and advantages will become better understood with regard to the following description, and accompanying drawings, where:

[00102] **Figure 1. B cell differentiation.** Mature naive B cells are $CD19^+$ and can be activated to proliferate and differentiate upon antigenic challenge in secondary lymphoid tissues such as lymph nodes and spleen. They proliferate and differentiate in either extra-follicular foci or in germinal centers. Differentiating B cells in extrafollicular foci typically differentiate to become short-lived plasma cells and usually reside in the secondary lymphoid tissue they originated from. B cells differentiating in germinal centers can either become memory B cells, which can be further stimulated to differentiate via subsequent antigenic challenge, or become plasmablasts that have the potential to become long-lived plasma cells. These plasmablasts can enter the circulation and are trafficked to the various tissues in which long-lived plasma cells reside, such as the bone marrow, mucosal tissues (for IgA+ plasma cells) and inflamed tissues. Some transiting plasma cells are also present in blood. All of the above mentioned cell types can also be found in circulation in the blood.

[00103] **Figure 2. Schematic of high-throughput sequencing, cloning, and expression of paired genes from single sorted cells.** Desired cell populations are single cell sorted based on their expression of cell surface markers into 96-well PCR plates. During reverse transcription, barcoded (sample-IDs) DNA adaptor molecules are added onto synthesized 1^{st} strand cDNA utilizing the template-switching property of MMLV H^- reverse transcriptases. RT products from each plate are then pooled separately and 2 rounds of PCR performed to amplify specific amplicon regions (amplicons). PCR is done with primers with 5' flanking barcodes to add plate identification regions (plate-IDs) to amplicons. Amplicons are then sent for 454 sequencing (a). Obtained sequences are subset by plate-IDs and sample-IDs before sequence assembly. Amplicons from the same cell are paired together using plate-IDs and sample-IDs and desired clones selected for cloning and expression (b). Specific amplicons can be amplified from each

pooled plate of amplicons by using cloning primers that are complementary to the sample-ID of that particular amplicon. Cloning primers also add restriction site regions (RS) which are then used to insert the clone into mammalian expression vectors for downstream expression and screening. In this example, the amplicons are immunoglobulin (Ig) heavy and light chain genes which code for antibody. Expressed antibodies can then be used for downstream screening (c). 5' RS and 3' RS: 5' and 3' restriction sites respectively. HC and LC: heavy chain and light chain respectively.

[00104] Figure 3. Schematic of reverse transcription and PCR to add sample-IDs and plate-IDs to Ig amplicons. Reverse transcription (RT) was performed with Superscript II or Superscript III, which are MMLV H$^-$ reverse transcriptases. These transcriptases have a 3' tailing activity and add a couple of cytosines to the 3' end of newly synthesized 1st strand cDNAs. An oligonucleotide ending with –GGG (an adapter region) can complement base-pair to this and the reverse transcriptase then switches template to the oligonucleotide and carry on transcription, resulting in the incorporation of the sample-ID adaptor to the 3' end of the cDNA (a). As the sample-ID adaptor contains a 5' invariable sequence (universal primer region), forward primers complementary to this sequence can be used for subsequent PCRs. The 1st PCR was done with Fw Long Primer1, Fw Short Primer1 and GSP1. The Fw Long Primer1 has a 5' flanking region containing a plate-ID barcode and the 454 Titanium Primer A for 454 sequencing, which were incorporated into the amplicon. The Fw Short Primer1 has a Tm similar to the GSP1 primer and was included to slightly increase the efficiency of the PCR. Each GSP1 (gene-specific primer 1) has a complementary gene specific sequence and was used to amplify a specific gene. Here, the gene-specific primers are for the kappa and lambda light chains and the gamma heavy chain to amplify these genes. Sequences for primers are shown (b). The second PCR is a nested PCR and was done with Fw Primer2, Long GSP2 Primer and RV Primer2. Sequences for primers are as shown. Long GSP2 Primer has a 5' flanking region containing a plate-ID barcode and the 454 Titanium Primer B for 454 sequencing, which were incorporated into the amplicon. Long GSP2 primers again amplify the kappa and lambda light chains and the gamma heavy chain. The RV Primer2 has a Tm similar to Fw Primer2 and was included to slightly increase PCR efficiency. Sequences for primers are shown. After RT and 2 PCRs, each amplicon will have 454 Titanium Primers A and B for 454 sequencing, two identical plate-IDs, each identifying the amplicon as coming from a particular single cell-sorted 96-well plate and a sample-ID determining its position on the 96-well plate (c).

[00105] Figure 4. Successful amplification of single-cell sorted B cells using sequencing and cloning primers. A 96-well plate of single-cell sorted B cells were reverse transcribed,

pooled and amplified as shown in the schematics in Fig. 1a and Fig. 2. Bands for the kappa light chain, lambda light chain, and gamma heavy chain were visualized on an agarose gel at the expected sizes: ~600bp for kappa and lambda and ~700bp for gamma (a). DNA chromatogram of Sanger sequencing of the gamma heavy chain from the 5' end showed a 'variable' sequence corresponding to multiple sample-IDs for the pooled plate (b). Sanger sequencing of the kappa light chain from the 3' end showed a 'variable' sequence after the constant region and beginning at the VJ junction corresponding to multiple light chains (c). A cloning primer pair specific for well A1 was used to amplify the kappa light chain. Sanger sequencing showed that in contrast to (c), only one clean sequence was amplified (d). All results are representative of the other amplified immunoglobulin genes. M: 100bp DNA ladder, K: kappa light chain, L: lambda light chain, G: gamma heavy chain.

[00106] Figure 5. CCP+ RA patients have peripheral blood plasmablasts percentages that correlated with disease activity and secreted anti-citrulline autoantibodies. Plasmablasts are $CD19^{+}CD20^{-}CD27^{+}CD38^{++}$ and were first gated on CD3- cells then gated as shown (a). Peripheral blood was obtained from consented RA patients and plasmablasts plotted as a percentage of total PBMCs. A Mann-Whitney test showed that CCP+ RA patients possess significantly higher ($p<0.05$) plasmablast percentages than CCP- RA patients (b). Plasmablast percentages were significantly correlated with clinical disease activity index (CDAI) in CCP+ patients by linear regression. Linear regression was performed on log transformed plasmablasts percentages to achieve normality of dataset (c). CCP+ plasmablasts were also either mock-sorted or underwent a plasmablast-depletion sort with >95% elimination of plasmablasts (d) and cultured for 7 days in RPMI supplemented with 10% FBS. Supernatant were collected and analyzed for anti-citrullinated peptide reactivity with Luminex. Mean fluorescence intensities for antibody reactivity against each peptide were plotted as peptide reactivities (e).

[00107] Figure 6. Strategy for selection and screening of clones for neutralizing antibodies. Paired LC and HC antibody sequences are obtained from bioinformatic analysis of 454-sequenced amplicons and grouped into clonal families based on their LC and HC V(D)J usage. The clone(s) occurring at the highest frequency in each clonal family will be selectively cloned, expressed, and screened for binding to the target antigen of interest using ELISA. Representative clone(s) from the entire clonal family which secrete binding antibodies will then be cloned and expressed for screening of neutralizing antibodies. Each antibody diagram represents a clone.

[00108] Figure 7. Characterization of immunoglobulin heavy chain V(D)J sequences and clonal families derived from individual human subjects. Blood was obtained from

humans with the following conditions: (i) chronic *Staphylococcus aureus* osteomyelitis in a human who was not taking antibiotics (due to non-compliance) but whose immune system effectively suppressed the infection and prevented fulminant infection for several months off antibiotics; (ii) a human with acute and fulminant *Staphylococcus aureus* bacteremia requiring transfer to the intensive care unit and aggressive intravenous antibiotic treatment; (iii) 3 humans with chronic active rheumatoid arthritis (RA) (with disease activity scores (DAS) > 5); (iv) a human 7 days following receipt of the trivalent influenza vaccine (Fluzone, Sanofi); and (v) a human with metastatic lung adenocarcinoma who was expected to expire but following chemotherapy went into a state of long-term non-progression. In all cases, the human patients exhibited elevations in their peripheral blood plasmablast levels (ranging from 1.5 – 6% of peripheral blood B cells being plasmablasts [$CD20^{-}CD19^{+}CD38^{++}CD27^{+}$], with levels in normal humans being 0.1-0.2% of peripheral blood B cells), indicating an activated immune response. Plasmablasts were single-cell sorted into 96-well plates, and barcoding and 454 sequencing of the expressed immunoglobulin cDNA was performed as described in Figures 2 and 3. Bioinformatic analysis was used to pair the heavy and light chain immunoglobulins expressed by individual plasmablasts. Pie chart diagrams of the percent of heavy chain V(D)J usage for the individual patients are presented – each wedge represents the percent of plasmablasts expressing a distinct heavy chain V(D)J sequence rearrangement.

[00109] Figure 8. Clustering of immunoglobulin heavy chain V(D)J sequences from human subjects demonstrates clonal families and clonal subfamilies. The immunoglobulin heavy chain sequence datasets generated in the studies described in Figure 7 were subject to hierarchical clustering using the program Clustal. Hierarchical clustering yielded evolutionary trees representing the antibody response in each individual human.

[00110] Figure 9. Schematic of RT and PCR to add sample-IDs and plate-IDs to any amplicon and downstream utility. Individual samples comprising either single cells or multiple cells are separately reverse transcribed in wells. Reverse transcription adds a sample-ID and a 5' universal primer region to all 1st strand cDNA as previously described (a). cDNA from all wells of a plate are pooled and undergo 2 rounds of PCR. The 1st PCR uses Fw Short Primer1, Fw Long Primer 1 as forward primers and adds a 454 Titanium Primer A for 454 sequencing and a plate-ID to the 5' end of the sequence. The Fw Short Primer1 has a Tm similar to the GSP1 primer and was included to slightly increase the efficiency of the PCR. Each GSP1 primer has a gene specific sequence and can specifically amplify that gene. Sequences for primers are shown. Note that regardless of which gene is amplified, the forward primers remain constant (b). The second PCR is a nested PCR. Fw Primer 2 is the forward

primer, and the reverse primers are Long GSP2 Primer and Rv Primer 2. Long GSP2 is gene-specific and only amplifies a specific gene. It also adds the 454 Titanium Primer B for 454 sequencing and a plate-ID to the 3' end of the amplicon. RV Primer2 has a Tm similar to Fw Primer2 and was included to slightly increase PCR efficiency. Sequences for primers are shown. After RT and 2 PCRs, amplicons from all plates are pooled and 454-sequenced. The combination of plate-IDs and sample-IDs allows for identification of sequences that originate from the same sample. This allows for comparison of sequences between multiple samples. Sequences from the same origin may also be expressed in pairs to obtain the exact protein from the original cell, such as the T-cell receptor and other Ig isotypes such as IgM, IgE and IgA (c).

[00111] Figure 10. Gene-specific primers for reverse transcription (RT-GSPs) of immunoglobulin heavy and light chains. RT-GSPs were used instead of oligo(dT)s as primers in reverse transcription of heavy and light chain genes. cDNA were then amplified by PCR and visualized on an agarose gel. RT-GSP primers IgKC_v3(a), IgLC_v5, IgLC_v6, IgLC_v7 and IgLC_v8 in lanes 1-4 respectively (b), IgHGC_v10, IgHGC_v11, IgHGC_v13 and IgGC_v15 in lanes 1-4 respectively (c) and IgHGC_v16 (d). KC, LC and GC in the primer names indicate that the primer is specific for the kappa chain, lambda chain and gamma heavy chain respectively. White bands in gel photos indicate where non-relevant lanes had been cropped out.

[00112] Figure 11. Adaptor region sequences. RNA was reversed transcribed with oligonucleotides comprising a universal primer region and an adaptor region at the 3' terminal end. cDNA was then amplified using the universal primer region sequence as a forward primer and gene-specific sequences as reverse primers. Amplified products were visualized on an agarose gel. Adaptor region consists of G (a), GGGGG and rGrGrG in lanes 1 and 2 respectively (b). rG indicates RNA nucleotides instead of DNA nucleotides.

[00113] Figure 12. Universal primer sequences. RNA was reverse transcribed with oligonucleotides comprising a universal primer sequence and an adaptor region at the 3' terminal. cDNA were then amplified by PCR using a forward primer complementary to the universal primer region and a reverse primer complementary to the gene specific sequence. Univ_seq_4 (a), univ_seq_5 (b) and univ_seq_f (c). Vertical white bands in gel photos indicate where non-relevant lanes have been cropped out. Otherwise lanes belong to the same gel photo.

[00114] Figure 13. Gene-specific primer sequences for 1st PCR reaction. Gene-specific reverse primers used in amplification of sequences in the first PCR reaction. Either the 1st PCR reaction or the subsequent 2nd nested PCR products were run and visualized on an agarose gel. Reverse primers used are IgKC_v4, IgLC_v5, IgHGC_v13 on lanes 1-3 respectively (a), K_GSP1, L_GSP1, G_GSP1 on lanes 1-3 respectively (b), K_GSP1c, L_GSP1c on lanes 1-2

respectively (c), G_GSP1 (d), L_GSP1d, G_GSP1g on lanes 1-2 respectively (e), G_GSP1h, G_GSP1k, L_GSP1f, L_GSP1g on lanes 1-4 respectively (f), G_GSP1d (g) L_GSP1h-o on lanes 1-8 respectively (h), G_GSP1m-q and G_GSP1t on lanes 1-6 respectively (i). K, L and G in the primer names indicate that the primers are specific for the kappa, lambda and gamma immunoglobulin constant regions respectively. Each gel starts with a lane marker on the left followed by sample lanes. White bars between lanes on the same gel photo indicate where non-relevant lanes in-between have been cropped out.

[00115] **Figure 14. Gene-specific sequences for the 2nd PCR reaction.** Gene-specific reverse primers used in amplification of sequences in the 2nd PCR reaction. PCR products were run and visualized on an agarose gel. Reverse primers used are K_GSP2, L_GSP2, G_GSP2 in lanes 1-3 respectively (a), K_GSP2v2a, K_GSP2v2b, L_GSP2v2 in lanes 1-3 respectively (b), K_GSP2v2c, K_GSP2v2c, G_GSP2v2c1, G_GSP2v2c2 in lanes 1-4 respectively (c), K_GSP2v2d-f in lanes 1-3 respectively (d),K_GSP2v2g, L_GSP2v2d and G_GSP2b in lanes 1-3 respectively (e). K, L, G in the primer names indicates that they are specific for the kappa, lambda and gamma immunoglobulin constant regions respectively. Each gel starts with a lane marker on the left followed by sample lanes. White bars between lanes on the same gel photo indicate that non-relevant lanes in-between have been cropped out.

[00116] **Figure 15. Potential locations of barcode sequences to identify a linked pair of polynucleotide sequences.** The schematic illustrates the physical linkage of two nucleic acid segments, A and B (e.g., two cDNAs). A barcode (BC) is appended to any one of the ends, or both ends, or anywhere in the sequence linking A and B. In one embodiment, A and B represent immunoglobulin heavy and light chain sequences.

[00117] **Figure 16. Different types of overlap-extension tails.** The bold line corresponds to a gene specific sequence and the thin line corresponds to the overlapping tail. As indicated, the overlap can be entirely due to the overlap of the primer sequence or else due to partial or total overlap with a gene specific sequence. As indicated, the overlap can also contain a barcode sequence. Structures I, II, and III indicate potential locations of the overlaps.

[00118] **Figure 17. Schematic overview of external barcode addition to a linked pair of antibody light and heavy chains.** Shown are the products of a reverse transcription reaction. The LC gene specific PCR primer contains a bar code, sequencing primer site, and restriction site (RE1) to allow these elements to be added to the 3' end of the resulting PCR product. Primers specific for LC and HC with overlap-extensions and encoding a restriction site (RE3) are indicated. A reverse primer specific for HC containing a sequencing primer site and a

restriction site (RE2) is also indicated. Amplification results in a nucleic acid with the linked structure shown with a bar code at one end.

[00119] Figure 18. Schematic overview of internal barcode addition to a linked pair of antibody light and heavy chains. Shown is a method of using adaptors containing extension overlap and barcode sequences to join cDNAs resulting from reverse transcription of mRNAs using oligo (dT) primers. The method shown takes advantage of the 3' tailing and template switching activities of reverse transcriptase to add overlap-extension sequences to the cDNAs to be joined. In this example, one of the adaptors adds both a barcode and overlap-extension sequence to one of the cDNAs to be joined, while only the overlap-extension sequence is added to the other cDNA to be joined. After amplification, a linked structure carrying one barcode sequence in between the linked cDNAs is generated.

[00120] Figure 19. Schematic overview of addition of two internal barcodes to a linked pair of antibody light and heavy chains using universal sequence overlap-extension primers. Shown is a method of using adaptors containing both a universal sequence and a barcode to join cDNAs resulting from reverse transcription of mRNAs using oligo (dT) primers. In this example, PCR primers to the universal sequence add an overlap-extension sequence to each of the cDNAs to be joined. After the amplification scheme shown, a linked structure carrying two barcodes in between the linked cDNAs is generated.

[00121] Figure 20. Schematic overview of addition of two internal barcodes to a linked pair of antibody light and heavy chains using overlap-extension adaptors. Shown is a method of using adaptors containing both a barcode and overlap-extension sequence to join cDNAs resulting from reverse transcription of mRNAs using gene specific primers (GSP). In this example, the overlap extension sequences on the adaptors added to each of the cDNAs allow for joining of the cDNAs by annealing. After the amplification scheme shown, a linked structure carrying two barcodes in between the linked cDNAs is generated

[00122] Figure 21. Use of barcoded GSPs during reverse transcription in combination with template-switch added adaptors. RT was performed with total PBMC RNA and univ_seq_2 template-switching oligo and IgKC_v3 GSP (lanes 1-2) and IgLC_v5 GSP (lanes 3-4) with an additional 5' flanking sequence, of which the first part is the Fixed_PCR3 sequence, and the last 8 bp is a barcode. Aliquots of the RT reaction were used in subsequent PCR reactions, with either a 5' V_K (lane 1) or V_L (lane 3) primer or the Univ_seq_2 (lanes 2 and 4) as the 5' primer, and Fixed_PCR3 as the 3' primer. The PCR products in lanes 2 and 4 ran as a smear, showing that the barcoded GSPs are non-specific in the RT reaction, and are not suitable for use with template-switch added adaptors.

[00123] Figure 22. Gating scheme for flow cytometry sorting of single cells into 96-well plates. Plasmablasts are defined as $CD19^{+}CD20^{-}CD27^{+}CD38^{++}$. Single PBMCs were first gated on based on their FSC and SSC profile (not shown). Live $CD19^{+}$ B cells were then gated on (left panel), and further narrowed down to CD20- B cells (2^{nd} panel from left), and refined to $CD27^{+}CD38^{++}$ cells. From this, IgG^{+} plasmablasts were determined as IgA^{-} and IgM^{-}, as IgG^{+} plasmablasts do not express cell surface IgG. This population was single cell sorted into 96-well plates.

[00124] Figure 23. Plasmablasts are present in people undergoing immunological challenge. Plasmablasts constituted 0.15% of peripheral blood B cells in a representative healthy donor, and range from 0.5%-16.4% in people undergoing a variety of immunological challenges including infections (*Staphylococcus aureus* and *Clostridium difficile* infections), cancer (a patient with metastatic melanoma who was a non-progressor for >4 years due to treatment with ipilimumab and a patient with metastatic adenocarcinoma of the lung who was a long-term non-progressor for >3 years after receipt of chemotherapy), and vaccination (receipt of influenza virus vaccine). This shows that plasmablasts are elevated in and obtainable from a range of subjects mounting immune responses of interest for isolation of individual plasmablasts for high-throughput sequencing of the antibody repertoire to characterize the active humoral response.

[00125] Figure 24. Expressed recombinant antibodies were secreted for 2-3 weeks in transient transfections. As outlined in Figure 2, the paired heavy and light chain immunoglobulin cDNA were cloned by PCR and co-transfected into 293T cells at the 48-well scale. Supernatants were collected every other day for 18 days. Anti-human IgG ELISA was performed to determine the amount of secreted antibodies in the collected supernatants, and the concentration of the antibodies in the supernatants of a panel of individual co-transfectants are graphed. Secretion tended to peak by day 9 and was substantially diminished by day 18.

[00126] Figure 25. Paired antibody heavy chain (HC) and light chain (LC) from an influenza vaccinated human exhibit variation across the complement determining regions (CDRs). Figure 25A: Partial dendrogram of flu antibodies. After pairing of heavy and light chains, a multiple sequence alignment was generated for heavy chains, and another multiple sequence alignment was generated for light chains. Both multiple sequence alignments were generated using Clustalw2 2.1 with default parameters. The two alignments were concatenated together and used to build a tree in CLC Sequence Viewer v. 6.5.2 using the neighbor joining method with 100 bootstrap replicates. Figure 25B: Heavy chain CDRs for a clonal family from flu-vaccinated patient. Identifiers in figure correspond to sequence names in Sequence Listing

as follows: 51.A11.1 = NA.51.11.A11.1.454.heavy.3.nb-aa, 49.A08.1 = NA.49.8.A08.1.454.heavy.3.nb-aa, 51.D07.1 is the amino acid sequence obtained by translating NA.51.40.D07.1.454.heavy.3.nb in frame 1. Figure 25C: Light chain CDRs for a clonal family from flu-vaccinated patient. Identifiers in figure correspond to sequence names in Sequence Listing as follows: 51.A11.1 = NA.51.11.A11.1.454.light.4.nb-aa, 49.A08.1 = NA.49.8.A08.1.454.light.4.nb-aa, 51.D07.1 = NA.51.40.D07.1.454.light.4.zerom50-aa.

[00127] Figure 26. Recombinant anti-influenza antibodies bound to Fluzone influenza virus vaccine. Analysis of the evolutionary tree (Figure 8) of the heavy and light chain antibody repertoire dataset generated for the influenza vaccinated human described in Figure 7 was performed to select antibodies representative of the clonal families identified. The heavy and light chains for the selected antibodies representing both clonal families as well as several singlet branches were cloned by PCR and co-transfected into 293T cells (as outlined in Figure 2), and supernatants collected from transfectants as described in Figure 24. The recombinant antibodies were then tested for reactivity against the Fluzone influenza virus vaccine (Sanofi) by ELISA, with the Fluzone vaccine coated on the ELISA plate. The recombinant influenza virus antibodies were incubated in the ELISA plate at 100ng/ml, and a horse radish peroxidase (HRP)-conjugated anti-human IgG antibody used to detect antibody binding. The TMB substrate reaction was allowed to go for 30 minutes before quenching with acid stop. Readout is displayed as 450nm absorbance as no standards were available. Multiple recombinant antibodies representing the identified clonal families bound to the influenza virus vaccine, while recombinant antibodies representative of other clonal families and the "dead ends" did not bind influenza vaccine.

[00128] Figure 27. Recombinant anti-influenza antibodies representative of clonal families bind influenza virus hemaglutinins with picomolar affinities. The recombinant anti-influenza virus antibodies representative of clonal families from the Fluzone-vaccinated human (Figure 7) that bound influenza vaccine in an ELISA assay (Figure 26) were tested using a surface plasmon resonance (SPR) instrument (ProteOn System, Bio-Rad Laboratories) to determine their binding affinities for influenza hemagglutinin (both the H3N2 A/Perth/16/2009 and H1N1 A/California/07/2009 strains present in the vaccine). The recombinant anti-influenza virus antibodies were bound to the surface using EDAC-NHS chemistry, and the H3N2 Perth and H1N1 California hemaglutinins were independently tested as the ligands, with hemagglutinin as the analyte. Ka column denotes the on-rates, Kd column the off-rates and K_D the dissociation constant. Multiple recombinant antibodies bound either the H3N2 Perth or the H1N1 California hemagglutinin with picomolar affinities.

[00129] **Figure 28. Recombinant anti-influenza antibodies neutralize influenza virus infectivity in microneutralization assays.** Six antibodies exhibiting reactivity on the Fluzone ELISA (Figure 26) were sent to the contract research organization (CRO) Virapur, LLC, (San Diego, CA) for testing in a microneutralization assay using the H1N1 California/07/2009 influenza virus strain and the H3N2 A/Perth/16/2009 influenza virus strain, 2 of the three stains of influenza virus in the Fluzone vaccine. 5 out of the 6 recombinant antibodies neutralized influenza virus in the microneutralization assay, preventing infectivity at microgram per milliliter levels and possibly sub-microgram per milliliter concentrations. The recombinant antibody F21 neutralized H3N2 Perth, and although it bound Fluzone in the ELISA assay it did not show binding in the SPR analysis (Figure 27) likely because the concentrations of the hemagglutinin analyte that were used were too low for binding to be detectable.

[00130] **Figure 29. Recombinant anti-*Staph. aureus* antibodies bound to fixed *S. aureus* by flow cytometry.** Analysis of the evolutionary tree (Figure 8) of the heavy and light chain antibody repertoire dataset generated from the human who controlled (without antibiotics) a chronic *Staph. aureus* osteomyelitis (as described in Figure 7) enabled selection of antibodies representative of the clonal families identified. The heavy and light chains for the selected antibodies representing both clonal families as well as several singlet branches were cloned by PCR and co-transfected into 293T cells (as outlined in Figure 2), and supernatants collected from transfectants as described in Figure 24. The recombinant anti-*Staph. aureus* antibodies were then tested for reactivity against fixed *S. aureus*. The secondary antibody used was a FITC-conjugated mouse anti-human IgG, and samples were analyzed on a BD LSR II or LSR Fortessa. The percentage of positive staining is shown, with 2 anti-influenza antibodies used as negative controls. The 2° antibody alone did not result in binding over background, as protein A binds weakly to mouse IgG1, which is the isotype of the 2° antibody. The staining observed above background is due to the binding of the recombinant anti-*Staph. aureus* antibodies to the small percentage of the Wood strain of *S. aureus* that express protein A. (a). Flow cytometry plots are shown for the 2 positive binding anti-*Staph. aureus* antibodies, S6 and S11, along with the isotype-matched negative control anti-influenza antibodies (b). The level of binding of the anti-*Staph. aureus* antibodies to *S. aureus* was proportional to the amount of antibody used. The dark solid line represents an antibody concentration of 10ug/ml, the dark dotted line 5ug/ml, and grey dotted line 1ug/ml (c).

[00131] **Figure 30. Anti-*Staph. aureus* antibodies reduced the number of *S. aureus* colony forming units.** Recombinant anti-*Staph.* aureus antibodies were incubated with *S. aureus* in combination with baby rabbit serum as a complement source, before being serially

diluted and grown overnight on 5% trypticase soy agar (TSA) blood agar plates. Colony forming units (CFUs) were then counted and graphed. Two recombinant anti-*Staph. aureus antibodies* (Ab-a and Ab-b) resulted in killing of the *Staph. aureus* and thus a reduced number of CFU/ml.

[00132] Figure 31. Identification of *S. aureus* antigen targets of recombinant anti-*Staph. aureus* antibodies generated from a human mounting an effective immune response against a chronic *Staph. aureus* infection. A protein lysate was generated from a clinical *Staph. aureus* isolate. Recombinant anti-*Staph. aureus* antibodies representative of clonal families identified in the antibody repertoire from a human mounting an immune response that was preventing progression of a chronic *Staph. aureus* osteomyelitis infection were used to immunoprecipitate proteins from a *Staph. aureus* Spa⁻ clinical isolate. The immunoprecipitates were separated by SDS-PAGE, identified bands excised, and mass spectrometry (an Agilent XCT-Plus ion trap mass spectrometer) used to identify the immunoprecipiated proteins which are presented in the Figure.

[00133] Figure 32. Generation of anti-lung adenocarcinoma antibodies from a human with metastatic lung adenocarcinoma who was a long-term non-progressor. A human with metastatic lung adenocarcinoma who became a long-term non-progressor following chemotherapy exhibited persistently elevated blood plasmablasts, indicating a persistently activated immune response (Figure 7). The patient's peripheral blood plasmablasts were sorted, the antibody repertoire sequenced, and antibodies representative of clonal families (Figure 8) were cloned and expressed recombinantly. One of the expressed recombinant antibodies, which is representative of one of the identified clonal families, bound to an independent lung adenocarcinoma in immunohistochemical stains. Tissue arrays were then used to further characterize the reactivity of this antibody. Tissue arrays containing multiple independent lung adenocarcinomas, squamous cell carcinoma and healthy lung tissue were blocked overnight with 100ug/ml of F(ab) goat anti-human antibody. Slides were stained with anti-lung adenocarcinoma antibodies or an anti-influenza antibody as a negative control, and visualized with Vector Red. Slides were counterstained with hematoxylin. Hematoxylin blue color was removed using Photoshop so that only nuclei and Vector Red (red) staining shows up as darker grey in the image. This recombinant antibody bound to 4 out of 5 independent lung adenocarcinoma tissue samples tested (contained in tissue arrays), but did not bind to lung squamous cell carcinoma or to healthy lung tissue.

[00134] Figure 33. The identified anti-lung adenocarcinoma antibody (Figure 32) binds to the surface of a lung adenocarcinoma cell line. The recombinant anti-lung

adenocarcinoma antibody (Figure 32) strongly stained the surface of the lung adenocarcinoma cell line H1650, and exhibited only low-level staining of a kidney epithelial and a lung squamous tumor cell lines. This anti-lung adenocarcinoma antibody did not bind a second lung adenocarcinoma cell line (H2009), consistent with our observation that this antibody bound 4 out of the 5 independent lung adenocarcinoma tissue samples tested by immunohistochemistry (Figure 32).

[00135] Figure 34. Generation of rheumatoid factor antibodies from rheumatoid arthritis (RA) patients. Recombinant antibodies representative of clonal families identified in the evolutionary trees of antibody repertoires generated from humans with RA (Figure 8) were selected, cloned and recombinantly expressed. Recombinant antibodies derived from RA patients were used as the primary antibody in a direct ELISA and anti-human IgG-HRP was used as the secondary antibody, and binding visualized with TMB substrate. Recombinant antibodies RA2 and RA3 exhibited reactivity, and thus represent rheumatoid factor antibodies.

[00136] Figure 35. Generation of anti-CCP and anti-histone 2A antibodies from RA patients. Additional recombinant antibodies generated from RA patients with active disease were characterized using a histone 2A ELISA and a cyclic-citrullinated peptide (CCP) ELISA (using the CCP2 ELISA kit [Axis Shield]). Recombinant antibodies were used at 125ug/ml. Panel (a) presents the results from a histone 2A ELISA, and multiple recombinant antibodies bound to histone 2A. Panel (b) present the results of the CCP2 ELISA, and several recombinant antibodies exhibited positive reactivity. The anti-CCP2 ELISA included a seronegative and 2 seropositive controls. For both assays, absorbance was recorded as the readout. Absorbance values above the background (dotted line) were considered to be positive.

[00137] Figure 36. Confirmatory independent experiment demonstrating generation of anti-histone 2A antibodies from active RA patients. Recombinant antibodies derived from RA patient evolutionary trees (Figure 8 and Figure 35) were further tested in a histone 2A ELISA assay. Antibodies were used at 30ug/ml, a 4-fold lower concentration that that used in Figure 35. Absorbance was recorded as the readout. Absorbance values above the background were considered to be positive.

[00138] Figure 37. Identification of anti-histone and anti-citrullinated protein antibodies using RA antigen arrays. Antibodies derived from RA patients were used to probe an RA antigen array containing a spectrum of native and citrullinated proteins and peptides. Following incubation with a Cy-3-labeled anti-human IgG secondary antibody, recombinant antibody binding was quantitated by scanning with an Axon Instruments GenePix microarray scanner.

Reactivities are displayed as a heatmap. Recombinant antibodies derived from RA bound to several distinct citrullinated or native antigens.

[00139] Figure 38. Pacific Biosciences sequencing provides full-length sequencing reads of IgG heavy chain amplicon. IgG heavy chain amplicons from plate 44 were provided to Pacific Biosciences for SMRT sequencing. The number of circular consensus sequence (CCS) reads with barcodes corresponding to selected wells are shown.

[00140] Figure 39. Use of alternative cell surface markers and other cellular features to identify blood plasmablasts. Plasmablasts can be identified and sorted through use of a variety of cell surface markers and/or cellular features. Panel (a) demonstrates that plasmablasts exhibit higher forward scatter (FSC) than resting B cells. Plasmablasts were identified based on $CD19^{+}CD20^{-}CD27^{+}CD38^{hi}$ staining, and these results demonstrate that B cells (grey) are smaller than plasmablasts (black). Panel (b) demonstrates that use of anti-CD19 staining combined with FSC identifies a population of B cells that contains 72% plasmablasts. Panel (c) demonstrates that, for a population of $CD19^{+}$ B cells (cells were pre-gated as being CD19 positive), side scatter (SSC) and FSC can be used to identify a population of B cells that contains 37% plasmablasts. Panels (d-f) present several approaches to identify plasmablasts within the $CD19^{+}$ B cell population. Gating on FSC^{hi} cells give 37% purity of plasmablasts (c). Gating on $FSC^{hi}CD20^{-}$ cells gave 71% purity in plasmablasts (d). Gating on $FSC^{hi}CD38^{+}$ cells gave 80% purity in plasmablasts (e). Gating on $FSC^{hi}CD27^{+}$ cells gave 44% purity in plasmablasts (f).

[00141] Figure 40. Human blasting B cells (plasmablasts) are larger than resting B cells but smaller than monocytes on average. Singlet monocytes, B cells and plasmblasts were gated and compared for side- and forward-scatter parameters. Monocytes were defined by their characteristic FSC and SSC profile, and as $CD19^{-}CD3^{-}$. B cells were defined as $CD19^{+}CD20^{+}$. Plasmablasts were defined as $CD19^{+}CD20^{-}CD27^{+}CD38^{++}$. Cells shown on the FSC-A (forward scatter area) and SSC-A (side scatter area) axes (a). Cells shown on the FSC-W (forward scatter width) and SSC-W (side scatter width) axes (b). The median of the FSC-A, SSC-A, FSC-W, SSC-W of plasmablasts were divided by that of resting B cells or monocytes to obtain a ratio which represent the size relationship between the cell types (c). The median of the FSC-A, SSC-A, FSC-W, SSC-W of the 20^{th} percentile of plasmablasts were divided by that of the median of resting B cells or monocytes to obtain a ratio which represent the size relationship between the cell types wherein at least 80% of plasmablasts are larger than the ratio (d). Error bars indicate 95% confidence interval.

[00142] Figure 41. Size of human plasmablasts compared to resting B cells by microscopy. Plasmablasts and resting B cells were sorted as $CD19^{+}CD20^{-}CD27^{+}CD38^{++}$ and $CD19^{+}CD20^{+}$ respectively. Cells were then imaged using an Olympus microscope at 200x. Cell area was measured using ImageJ and diameter determined using area=π x r^2 where diameter = 2 x radius, and volume determined by 4/3 x πr^3 . Error bars denote the inter-quartile range. A cut-off of ≥8uM or ≥50uM^2 or ≥268uM^3 will include 96% of plasmablasts and exclude 92% of resting B cells.

[00143] Figure 42. Superscript III has template switching activity at temperatures at and below 50°C. Reverse transcription (RT) was performed for 90 minutes using the temperatures indicated above the lanes using an adaptor ending with rGrGrG, and 1 round of PCR was done using GAPDH as the 3' primer (sequence ATGGTTCACACCCATGACG). As can be seen, no template switching activity to add on the adaptor could be seen at 55°C, and template switching activity increases from minimal at 50°C to highest at 42°C, the lowest temperature tested, as indicated by the brightness of the band at ~450bp. Marker is a 100bp marker. Superscript III is an MMLV reverse transcriptase that has specific mutations that result in a loss of RNAse H activity, and also has mutations made to the polymerase domain to increase thermal stability and has a half-life of 220 minutes at an RT temperature of 50oC. Other MMLV H^- enzymes that have been engineered for higher thermal stability are expected to exhibit similar activity.

[00144] Figure 43. Additional primers for human kappa, lambda and gamma constant regions. These primers were used for the 1st PCR, and then the 2nd PCR was performed using the primers from Table 1 and PCR products separated on a 2% agarose gel and the image was taken. Primers used for 1st PCR are Kappa GSP1, kappa GSP1e, kappa GSP1f, lambda GSP1, lambda GSP1x and lambda GSP1y respectively. Sequences are in Table 10. White bars between lanes on the same gel photo indicate that non-relevant lanes in-between have been cropped out.

[00145] Figure 44. Additional primers for other human constant regions and genes. 1st and 2nd PCR were done and products ran on a 2% agarose-gel and imaged. Lanes are from left: marker, mu, alpha constant regions, TCR alpha (a) and marker, TCR beta (b). Primers used and sequences are in Table 10. White bars between lanes on the same gel photo indicate that non-relevant lanes in-between have been cropped out.

[00146] Figure 45. Additional primers for mouse genes. 1st and 2nd PCR were done and products ran on a 2% agarose gel and imaged. Lanes are from left: marker, kappa, lambda, lambda, lambda, lambda light chains and mu heavy chain. The 4 lambda lanes had this combination of primers used: mouse_lambda_GSP1a with mouse_lambda_GSP2a,

mouse_lambda_GSP1a with mouse_lambda_GSP2b, mouse_lambda_GSP1b with mouse_lambda_GSP2a, and mouse_lambda_GSP1b with mouse_lambda GSP2a (a). Marker and alpha heavy chain (b). Gamma1, 2a, 2c heavy chains with 2nd PCR using mo_g12_GSP2d and mo_g12_GSP2e respectively, marker (c). Marker, gamma 3 heavy chain with 2nd PCR using mo_g3_GSP2d, mo_g3_GSP2e respectively followed by gamma 2b heavy chain with 2nd PCR using mo_g2b_GSP2d, mo_g2b_GSP2e respectively (d). Marker, TCR alpha (e). Marker, TCR beta (f). White bars between lanes on the same gel photo indicate that non-relevant lanes in-between have been cropped out.

[00147] Figure 46. Anti-*S. aureus* antibody-mediated killing of *S. aureus* by the HL-60 neutrophil cell line. Various recombinant anti-*S. aureus* antibodies (staph 1, staph 4, staph 6, staph 7, staph 9, staph 12) were incubated at 4°C with *S. aureus* for 30 minutes, following which non-bound antibody was washed away, and the *S. aureus* incubated with activated HL-60 cells and baby rabbit complement for 45 minutes at 37°C. Cells were then washed twice and extracellular bacteria were serially plated on 5% TSA blood agar plates, incubated overnight, and colony forming units (CFUs) counted. Recombinant antibodies staph 6, stepha 9 and staph 12 induced greater than 20% killing of *S. aureus*.

DETAILED DESCRIPTION

Compositions

Polynucleotides

[00148] In some aspects, a composition can include a polynucleotide. The term "polynucleotide(s)" refers to nucleic acids such as DNA molecules and RNA molecules and analogs thereof (e.g., DNA or RNA generated using nucleotide analogs or using nucleic acid chemistry). As desired, the polynucleotides may be made synthetically, e.g., using art-recognized nucleic acid chemistry or enzymatically using, e.g., a polymerase, and, if desired, can be modified. Typical modifications include methylation, biotinylation, and other art-known modifications. In addition, a polynucleotide can be single-stranded or double-stranded and, where desired, linked to a detectable moiety. In some aspects, a polynucleotide can include hybrid molecules, e.g., comprising DNA and RNA.

[00149] "G," "C," "A," "T" and "U" each generally stand for a nucleotide that contains guanine, cytosine, adenine, thymidine and uracil as a base, respectively. However, it will be understood that the term "ribonucleotide" or "nucleotide" can also refer to a modified nucleotide or a surrogate replacement moiety. The skilled person is well aware that guanine, cytosine, adenine, and uracil may be replaced by other moieties without substantially altering the base

pairing properties of an oligonucleotide comprising a nucleotide bearing such replacement moiety. For example, without limitation, a nucleotide comprising inosine as its base may base pair with nucleotides containing adenine, cytosine, or uracil. Hence, nucleotides containing uracil, guanine, or adenine may be replaced in nucleotide sequences by a nucleotide containing, for example, inosine. In another example, adenine and cytosine anywhere in the oligonucleotide can be replaced with guanine and uracil, respectively to form G-U Wobble base pairing with the target mRNA. Sequences containing such replacement moieties are suitable for the compositions and methods described herein.

[00150] As used herein, and unless otherwise indicated, the term "complementary," when used to describe a first nucleotide sequence in relation to a second nucleotide sequence, refers to the ability of a polynucleotide comprising the first nucleotide sequence to hybridize and form a duplex structure under certain conditions with a polynucleotide comprising the second nucleotide sequence, as will be understood by the skilled person. Such conditions can, for example, be stringent conditions, where stringent conditions may include: 400 mM NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA, 50°C or 70°C for 12-16 hours followed by washing. Other conditions, such as physiologically relevant conditions as may be encountered inside an organism, can apply. The skilled person will be able to determine the set of conditions most appropriate for a test of complementarity of two sequences in accordance with the ultimate application of the hybridized nucleotides.

[00151] Complementary sequences include base-pairing of a region of a polynucleotide comprising a first nucleotide sequence to a region of a polynucleotide comprising a second nucleotide sequence over the length or a portion of the length of one or both nucleotide sequences. Such sequences can be referred to as "complementary" with respect to each other herein. However, where a first sequence is referred to as "substantially complementary" with respect to a second sequence herein, the two sequences can be complementary, or they may include one or more, but generally not more than about 5, 4, 3, or 2 mismatched base pairs within regions that are base-paired. For two sequences with mismatched base pairs, the sequences will be considered "substantially complementary" as long as the two nucleotide sequences bind to each other via base-pairing.

[00152] "Complementary" sequences, as used herein, may also include, or be formed entirely from, non-Watson-Crick base pairs and/or base pairs formed from non-natural and modified nucleotides, in as far as the above embodiments with respect to their ability to hybridize are fulfilled. Such non-Watson-Crick base pairs includes, but are not limited to, G:U Wobble or Hoogstein base pairing.

[00153] The term percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (e.g., BLASTP and BLASTN or other algorithms available to persons of skill) or by visual inspection. Depending on the application, the percent "identity" can exist over a region of the sequence being compared, e.g., over a functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

[00154] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[00155] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., infra).

[00156] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information web-site.

[00157] Identical sequences include 100% identity of a polynucleotide comprising a first nucleotide sequence to a polynucleotide comprising a second nucleotide sequence over the entire length of one or both nucleotide sequences. Such sequences can be referred to as "fully identical" with respect to each other herein. However, in some aspects, where a first sequence is referred to as "substantially identical" with respect to a second sequence herein, the two sequences can be fully complementary, or they may have one or more, but generally not more than about 5, 4, 3, or 2 mismatched nucleotides upon alignment. In some aspects, where a first sequence is referred to as "substantially identical" with respect to a second sequence herein, the

two sequences can be fully complementary, or they may be about 50, 60, 70, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identical to each other.

[00158] Where a first sequence is referred to as "distinct" with respect to the identity of a second sequence herein, the two sequences have at least one or more mismatched nucleotides upon alignment. In some aspects, distinct sequences can have 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more mismatched nucleotides upon alignment. In some aspects, distinct sequences can be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or less than 100% identical to each other. In some aspects, where a first sequence is referred to as "distinct" with respect to a second sequence herein, the two sequences can have substantially or fully identical sequences, but instead differ from one another based upon differing patterns of modification within the sequences. Such modifications are generally known in the art, e.g., methylation.

[00159] In some aspects, a polynucleotide can be present in a library of polynucleotides. In some aspects, a polynucleotide library can include a plurality of polynucleotides. In some aspects, each polynucleotide in the plurality of polynucleotides can be derived from a single sample. In some aspects, a single sample can include a single cell such as a B cell.

[00160] Conventional notation is used herein to describe nucleotide sequences: the left-hand end of a single-stranded nucleotide sequence is the 5'-end; the left-hand direction of a double-stranded nucleotide sequence is referred to as the 5'-direction. The direction of 5' to 3' addition of nucleotides to nascent RNA transcripts is referred to as the transcription direction. The DNA strand having the same sequence as an mRNA is referred to as the "coding strand;" sequences on the DNA strand having the same sequence as an mRNA transcribed from that DNA and which are located 5' to the 5'-end of the RNA transcript are referred to as "upstream sequences;" sequences on the DNA strand having the same sequence as the RNA and which are 3' to the 3' end of the coding RNA transcript are referred to as "downstream sequences."

[00161] The term "messenger RNA" or "mRNA" refers to an RNA that is without introns and that can be translated into a polypeptide.

[00162] The term "cDNA" refers to a DNA that is complementary or identical to an mRNA, in either single stranded or double stranded form.

[00163] The term "amplicon" refers to the amplified product of a nucleic acid amplification reaction, e.g., RT-PCR.

[00164] The term "hybridize" refers to a sequence specific non-covalent binding interaction with a complementary nucleic acid. Hybridization may occur to all or a portion of a nucleic acid sequence. Those skilled in the art will recognize that the stability of a nucleic acid duplex, or hybrids, can be determined by the Tm. Additional guidance regarding hybridization conditions may be found in: Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 1989, 6.3.1-6.3.6 and in: Sambrook et al., Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, Vol. 3.

[00165] As used herein, "region" refers to a contiguous portion of the nucleotide sequence of a polynucleotide. Examples of regions are described herein an include identification regions, sample identification regions, plate identification regions, adapter regions, and the like. In some aspects, a polynucleotide can include one or more regions. In some aspects, a polynucleotide can include less than 2, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more regions. In some aspects, regions can be coupled. In some aspects, regions can be operatively coupled. In some aspects, regions can be physically coupled.

[00166] As used herein, "variable region" refers to a variable nucleotide sequence that arises from a recombination event, for example, it can include a V, J, and/or D region of an immunoglobulin or T cell receptor sequence isolated from a T cell or B cell of interest, such as an activated T cell or an activated B cell.

[00167] As used herein "B cell variable immunoglobulin region" refers to a variable immunoglobulin nucleotide sequence isolated from a B cell. For example, a variable immunoglobulin sequence can include a V, J, and/or D region of an immunoglobulin sequence isolated from a B cell of interest such as a memory B cell, an activated B cell, or plasmablast.

[00168] As used herein "identification region" refers to a nucleotide sequence label (e.g., a unique barcode sequence) that can be coupled to at least one nucleotide sequence for, e.g., later identification of the at least one nucleotide sequence.

[00169] As used herein "immunoglobulin region" refers to a contiguous portion of nucleotide sequence from one or both chains (heavy and light) of an antibody.

[00170] As used herein "adapter region" refers to a linker that couples a first nucleotide sequence to a second nucleotide sequence. In some aspects, an adapter region can include a contiguous portion of nucleotide sequence that acts as a linker. For example, an adapter region can have the sequence GGG and couples a first sequence to a second sequence via binding between GGG and CCC.

[00171] In some aspects, a polynucleotide can include a cDNA region. In some aspects, a polynucleotide can include a sample identification-adapter region. In some aspects, a polynucleotide can include a sample identification region. In some aspects, a polynucleotide can include an adapter region. In some aspects, a polynucleotide can include a universal primer region. In some aspects, a polynucleotide can include an amplicon region. In some aspects, a polynucleotide can include a plate identification region. In some aspects, a polynucleotide can include a first plate identification region. In some aspects, a polynucleotide can include a second plate identification region. In some aspects, a polynucleotide can include a restriction site region. In some aspects, a polynucleotide can include a first restriction site region. In some aspects, a polynucleotide can include a second restriction site region. In some aspects, a polynucleotide can include a sequencing region. In some aspects, a polynucleotide can include a first sequencing region. In some aspects, a polynucleotide can include a second sequencing region.

[00172] In some aspects, a polynucleotide can include a plurality of any region described herein. For example, a polynucleotide can include a first sample identification region and a second sample identification region. In some aspects, the first sample identification region and the second sample identification region are identical or substantially identical. In some aspects, the first sample identification region and the second sample identification region are distinct. In some aspects, an identification region is coupled to a variable immunoglobulin region.

[00173] In some aspects the sequence of a region will be at least long enough to serve as a target sequence for a primer or a probe in a PCR reaction. In some aspects, a region can be 1 to greater than 5000 base pairs in length. For example, a region can be from 1-10,000 nucleotides in length, e.g., 2-30 nucleotides in length, including all sub-ranges therebetween. As non-limiting examples, a region can be from 1-30 nucleotides, 1-26 nucleotides, 1-23 nucleotides, 1-22 nucleotides, 1-21 nucleotides, 1-20 nucleotides, 1-19 nucleotides, 1-18 nucleotides, 1-17 nucleotides, 18-30 nucleotides, 18-26 nucleotides, 18-23 nucleotides, 18-22 nucleotides, 18-21 nucleotides, 18-20 nucleotides, 19-30 nucleotides, 19-26 nucleotides, 19-23 nucleotides, 19-22 nucleotides, 19-21 nucleotides, 19-20 nucleotides, 20-30 nucleotides, 20-26 nucleotides, 20-25 nucleotides, 20-24 nucleotides, 20-23 nucleotides, 20-22 nucleotides, 20-21 nucleotides, 21-30 nucleotides, 21-26 nucleotides, 21-25 nucleotides, 21-24 nucleotides, 21-23 nucleotides, or 21-22 nucleotides. In some aspects, a region can be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more nucleotides in length. In some aspects, a region can be less than 50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800,

800-900, 900-1000, or greater than 1000 nucleotides in length. In some aspects, a region can be less than 1000, 1000-2000, 2000-3000, 3000-4000, 4000-5000, 5000-6000, 6000-7000, 7000-8000, 8000-9000, 9000-10000, or greater than 10000 nucleotides in length. In some aspects, a region can include at least two nucleotides, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20 or more nucleotides of a polynucleotide disclosed herein.

[00174] The term "sample" can include RNA, DNA, a single cell or multiple cells or fragments of cells or an aliquot of body fluid, taken from a subject (e.g., a mammalian subject, an animal subject, a human subject, or a non-human animal subject). Samples can be selected by one of skill in the art using any means now known or later discovered including centrifugation, venipuncture, blood draw, excretion, swabbing, ejaculation, massage, biopsy, needle aspirate, lavage sample, scraping, surgical incision, laser capture microdissection, gradient separation, or intervention or other means known in the art. Samples can also be selected by one of skill in the art using one or more markers known to be associated with a sample of interest. Samples can also be selected using methods known in the art such as cell sorting and FACS. Further examples of sample selection methods are described in the Examples section below.

[00175] In some aspects a polynucleotide can be derived from or associated with a single sample. In some aspects a region can be derived from or associated with a single sample. In some aspects, a cDNA region can be derived from or associated with a single sample. In some aspects, an amplicon region can be derived from or associated with a single sample. A "single sample" includes a sample comprising polynucleotides that is taken from a single source. In some aspects, a single source includes a sample taken at a particular time point or at a particular location, e.g., in a subject or flask of cells or plate of cells. In some aspects, a first single sample is taken from a first subject at a first time point and a second single sample is taken from the first subject at a second time point that is distinct from the first time point. In some aspects, a first single sample is taken from a first subject at a first location and a second sample is taken from the first subject at a second location that is distinct from the first location. In some aspects, a first single sample is taken from a first subject at a time point and a second single sample is taken from a second subject at a time point. In some aspects, a first single sample is taken from a first subject at a location and a second sample is taken from a second subject at a location. In one embodiment, a sample comprises polynucleotides that include mRNA derived from one or more B cells. In another embodiment, a sample comprises polynucleotides including cDNA derived from one or more B cells. In another embodiment, a single sample comprises mRNA

derived from one or more B cells sorted into a single well of a 96-well or 384-well plate. Samples are generally derived from a prokaryotic cell(s) (e.g., a bacterial cell(s)), a eukaryotic cell(s) (e.g., a mammalian and yeast cell(s)), or other sources of genetic material such as a virus or phage. The term "mammal" or "mammalian" as used herein includes both humans and non-humans and include but is not limited to humans, non-human primates, canines, felines, murines, bovines, equines, and porcines. In some aspects, the methods of the invention are applied to single samples in a plate with at least 96 wells, at least 384 wells, at least 1536 wells, or more wells. In further aspects, the methods of the invention are applied to single samples in at least one, two, three, four, five, six, seven, eight, ten, fifteen, twenty, thirty or more plates with at least 96 wells each.

[00176] In some aspects a 5' adaptor region sequence and/or a sample identification region are added to all cDNAs from a single sample, e.g., during RT and not just to Ig genes. In some aspects, 3' gene specific primers (GSPs) can be used to amplify any expressed gene in the single sample. In some aspects, genes are amplified that have a 5' variable region, e.g., T cell receptors and B cell receptors without needing multiple degenerate 5' primers to amplify the gene(s) of interest. GSPs can include primers specific for IgG, IgM, IgD, IgA, IgE, TCR chains, and other genes of interest.

[00177] In some aspects, multiple rounds of PCR can also be performed, e.g., using nested GSPs. For such nested GSPs, the GSP for the second round of PCR hybridizes to its target gene sequence at a position 5' along that sequence relative to the position hybridized to by the GSP used in the first round of PCR.

[00178] In some aspects, cDNA region or an amplicon region can include a DNA polynucleotide. In some aspects, cDNA region or an amplicon region can include a cDNA polynucleotide. In some aspects, cDNA region or an amplicon region can include an RNA polynucleotide hybridized to a DNA polynucleotide. In some aspects, cDNA region or an amplicon region can include an mRNA polynucleotide hybridized to a cDNA polynucleotide.

[00179] In some aspects, a universal primer region is not fully complementary to any human exon. In some aspects, a universal primer region is not fully complementary to any expressed human gene. In some aspects, a universal primer region has minimal secondary structure.

[00180] In some aspects, an amplicon region comprises an immunoglobulin heavy chain amplicon sequence. In some aspects, an amplicon region comprises an immunoglobulin light chain amplicon sequence. In some aspects, an amplicon region comprises a T cell receptor alpha amplicon sequence. In some aspects, an amplicon region comprises a T cell receptor beta amplicon sequence.

[00181] In some aspects, a polynucleotide is present in a library of polynucleotides and can be differentiated from other polynucleotides present in the library based on a region of the polynucleotide.

[00182] In some aspects, the sequence of the sample identification region of each polynucleotide in a library derived from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library derived from one or more samples distinct from the first single sample. In some aspects, the sequence of the sample identification region of each polynucleotide in a library derived from a first single sample differs by at least 1 nucleotide from the sequence of the sample identification region of the other polynucleotides in the library derived from one or more samples distinct from the first single sample. In some aspects, the sequence of the sample identification region of each polynucleotide in a library derived from a first single sample differs by at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more nucleotides from the sequence of the sample identification region of the other polynucleotides in the library derived from one or more samples distinct from the first single sample. In some aspects, the sequence of the sample identification region of each polynucleotide in a library derived from a first single sample can be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or less than 100% identical to the sequence of the sample identification region of the other polynucleotides in the library derived from one or more samples distinct from the first single sample. In some aspects, the sequence of the sample identification region of each polynucleotide in a library derived from a first single sample is less than 100% identical to the sequence of the sample identification region of the other polynucleotides in the library derived from one or more samples distinct from the first single sample. In some aspects, a sample-identification region acts as a digital barcode on all 1st strand cDNA reverse transcribed from a single sample. In some aspects, the sample identification region is at least 1 nucleotide in length. In some aspects, a sample-identification region can comprise at least 3 nucleotides, and sample-identification regions can differ from each other by at least 1 nucleotide. In one embodiment, sample-identification regions are 3-15 nucleotides in length and differ from each other by at least 1 nucleotide. In some aspects, sample-identification regions can comprise at least 64 variants (using sample-identification regions 3 nucleotides in length with each sample-ID differing from each other by at least 1 nucleotide), or in some aspects larger numbers of

variants. In some aspects, the sequence attached 3' to the sample-identification region can be an adapter region comprising at least 1 G. In a preferred embodiment, the sequence attached 3' to the sample-identification region can be an adapter region comprising at least 2 G's. In one embodiment, a sequence attached to the 5' end of a sample-identification region is a universal primer sequence that can be used during PCR amplification to avoid the need for the subsequent addition of a 5' universal primer sequence (by ligation or another method) or the use of multiple degenerate 5' primers to amplify genes with variable 5' regions. Examples of sample identification regions are shown in Tables 2 and 8.

[00183] In some aspects, the sequence of the first plate identification region of each polynucleotide in a library derived from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the first plate identification region of each polynucleotide in a library derived from the first set of single samples differs by at least 1 nucleotide from the sequence of the first plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the first plate identification region of each polynucleotide in a library derived from the first set of single samples differs by at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more nucleotides from the sequence of the first plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the first plate identification region of each polynucleotide in a library derived from the first set of single samples can be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or less than 100% identical to sequence of the first plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the first plate identification region of each polynucleotide in a library derived from the first set of single samples is less than 100% identical to sequence of the first plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. Examples of first plate identification regions are shown in Tables 3 and 7.

[00184] In some aspects, the sequence of the second plate identification region of each polynucleotide in a library derived from a first set of single samples is distinct from the sequence of the second plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the second plate identification region of each polynucleotide in a library derived from the first set of single samples differs by at least 1 nucleotide from the sequence of the second plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the second plate identification region of each polynucleotide in a library derived from the first set of single samples differs by at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more nucleotides from the sequence of the second plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the second plate identification region is identical to the sequence of the first plate identification region on a polynucleotide. In some aspects, the sequence of the second plate identification region of each polynucleotide in a library derived from the first set of single samples can be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or less than 100% identical to sequence of the second plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. In some aspects, the sequence of the second plate identification region of each polynucleotide in a library derived from the first set of single samples is less than 100% identical to sequence of the second plate identification region of the other polynucleotides in the library derived from one or more single sample sets distinct from the first set of single samples. Examples of second plate identification regions are shown in Tables 3 and 7.

[00185] In some aspects, a plate-identification region (e.g., a first plate identification region or a second plate identification region) can comprise at least 2 nucleotides, and plate-identification regions differ from each other by at least 1 nucleotide. In one embodiment, plate-identification regions are 2-10 nucleotides in length and differ from each other by at least 1 nucleotide. In some aspects, use of plate-identification regions is found in only some embodiments, as the use of a larger number of different sample-identification regions (one per single sample to be analyzed) can eliminate the need for plate-identification regions. In some

aspects, plate-identification regions are used to reduce the number of unique oligonucleotides containing a sample-identification region that need to be synthesized.

[00186] In some aspects, a polynucleotide includes one or more adapter regions. In some aspects, an adapter region includes one or more G's. In some aspects, an adapter region includes 2, 3, 4, 5, 6, 7, 8, 9, 10 or more G's. In some aspects, adapter regions are attached to the 3' ends of cDNAs using the template switching property of MMLV H$^-$ reverse transcriptases. Different methods to attach adaptor regions exist, including but not limited to, doing PCR with primers with 5' flanking adaptor region sequences, sticky and blunt end ligations, template-switching-mediated addition of nucleotides, or other methods to covalently attach nucleotides to the 5' end, to the 3' end, or to the 5' and 3' ends of the polynucleotides. These methods can employ properties of enzymes commonly used in molecular biology. PCR can use, e.g., thermophilic DNA polymerase. Sticky ends that are complementary or substantially complementary are created through either cutting dsDNA with restriction enzymes that leave overhanging ends or through 3' tailing activities of enzymes such as TdT (terminal transferase). Sticky and blunt ends can then be ligated with a complementary adaptor region using ligases such as T4 ligase. Template-switching utilizes the 3' tailing activity of MMLV H$^-$ reverse transcriptase to add one or more cytosines (C's) to the 3' end of cDNAs and its ability to switch template from mRNA to an adaptor region with complementary G's. In some aspects, a cDNA includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more C's on its 3' end.

[00187] In some aspects, a polynucleotide includes one or more restriction site regions. Restriction site regions include one or more restriction sites. Restrictions sites can include: NheI, XhoI, BstBI, EcoRI, SacII, BbvCI, PspXI, AgeI, ApaI, KpnI, Acc65I, XmaI, BstEII, DraIII, PacI, FseI, AsiSI, and AscI. In some aspects, any rare 8-cutter enzyme restriction site can be used.

[00188] In some aspects, one or more regions of a polynucleotide described herein can be operatively coupled to one or more other regions of the polynucleotide. In some aspects, two or more distinct regions of a single polynucleotide can be operatively coupled. For example, a universal primer region can be operatively coupled to an adapter region. In some aspects two or more regions can be operatively coupled together that are substantially identical to each other in sequence or identical in description. For example, a first sample identification region can be operatively coupled to a second sample identification region. In some aspects, the sequences of the first sample identification region and the second sample identification region are identical or substantially identical. In some aspects, the sequences of the first sample identification region and the second sample identification region are different or distinct.

[00189] In some aspects, one or more regions of a polynucleotide described herein can be coupled to one or more other regions of the polynucleotide. In some aspects, two or more distinct regions of a single polynucleotide can be coupled. For example, a universal primer region can be coupled to an adapter region. In some aspects two or more regions can be coupled together that are substantially identical to each other in sequence or identical in description. For example, a first sample identification region can be coupled to a second sample identification region. In some aspects, the sequences of the first sample identification region and the second sample identification region are identical or substantially identical. In some aspects, the sequences of the first sample identification region and the second sample identification region are different or distinct.

[00190] In some aspects, a polynucleotide includes the sequence 5'-A-B-3', wherein A is a sample identification region, and wherein B is an adapter region. In some aspects, a polynucleotide includes the sequence 5'-A-B-C-3', wherein A is a universal primer region, wherein B is a sample identification region, and wherein C is an adapter region. In some aspects, a polynucleotide includes the sequence 5'-A-B-C-3', wherein A is a sample identification region, wherein B is an adapter region, and wherein C is an amplicon region derived from a single sample. In some aspects, a polynucleotide includes the sequence 5'-A-B-C-D-3', wherein A is a universal primer region, wherein B is a sample identification region, wherein C is an adapter region, and wherein D is an amplicon region derived from a single sample. In some aspects, a polynucleotide includes the sequence 5'-A-B-C-D-E-3', wherein A is a plate identification region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, and wherein E is an amplicon region derived from a single sample. In some aspects, a polynucleotide includes the sequence 5'-A-B-C-D-E-F-3', wherein A is a first restriction site region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, wherein E is an amplicon region derived from a single sample, and wherein F is a second restriction site region.

[00191] In some aspects, the regions of each of the above sequences can be rearranged in a different order, e.g., 5'-C-A-D-B-3' or 5'-E-A-C-B-D-F-3' or 5'-B-A-3'. In some aspects, one or more regions of the above sequences can be deleted, e.g., 5'-A-D-3' or 5'-B-C-3'. In some aspects, one or more additional regions can be added to the above sequences, e.g., 5'-A-A_2-B-3' or 5'-A-B-C-D-E-F-G-3'. In such examples the one or more additional regions can be any region disclosed herein or equivalents thereof. In some aspects, one or more regions of the sequences above can be modified, e.g., methylated.

[00192] In some aspects, a polynucleotide can include an adapter molecule. In some aspects, a polynucleotide adapter molecule can include a universal primer region, a sample identification region, and an adapter region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, and wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region. In some aspects, an adapter molecule includes a polynucleotide comprising at least 2 nucleotides that bind to C's added by a reverse transcriptase at the 3' end of a 1st strand cDNA. In some aspects, an adapter molecule includes a deoxyribose polynucleotide comprising 3-6 G's (DNA G's). In another embodiment, an adapter molecule includes a ribose polynucleotide consisting of 3-6 G's (RNA G's). In other embodiments, the adapter molecule can utilize nucleotide analogues, such locked nucleic acids (LNAs), e.g., LNA G's. In other embodiments, the nucleotide base may also be a universal or degenerate base such as 5-nitroindole and 3-nitropyrrole that can base-pair to C's as well as other nucleotides, in any combination.

[00193] In some aspects, a polynucleotide can include a primer or a probe. In some aspects, a primer can include a universal primer region and a plate identification region, and wherein the 3' end of the plate identification region is coupled to the 5' end of the universal primer region.

[00194] In some aspects, a composition can include a polynucleotide composition library. In some aspects, a polynucleotide composition library includes a plurality of polynucleotide compositions. In some aspects each composition is present in a separate container. In some aspects, a container can be a test tube. In some aspects, a container can be a well in a plate. In some aspects, a container can be a well in a 96-well plate. In some aspects, a container can be a well in a 384-well plate. In some aspects, each composition comprises a cDNA region derived from a single sample. In some aspects, each composition comprises a sample identification-adapter region comprising a sample identification region coupled to an adapter region. In some aspects the sequence of the sample identification region of each sample identification-adapter region in a library is distinct from the nucleotide sequence of the sample identification region of the other sample identification-adapter regions present in each separate container in the library. In some aspects the sample identification-adapter region is attached to the cDNA region. In some aspects the sample identification-adapter region is attached to the cDNA region by binding between their 3' regions. In some aspects the sample identification-adapter region is attached to the cDNA region by G:C binding. In some aspects, the cDNA region comprises an RNA polynucleotide hybridized to a DNA polynucleotide. In some aspects, the cDNA region comprises an mRNA polynucleotide hybridized to a cDNA polynucleotide.

[00195] In some aspects, the plurality of polynucleotide compositions in a polynucleotide library can comprise at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 3000, at least 10,000, at least 30,000, at least 100,000, at least 300, 000, at least 1,000,000, at least 3,000,000, at least 10,000,000, at least 30,000,000, or more members. In other aspects, the plurality of polynucleotide compositions in a polynucleotide library can comprise at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 3000, at least 10,000, at least 30,000, or more genes of a cell sample's whole transcriptome. In other aspects, the plurality of polynucleotide compositions in a polynucleotide library comprises at least 1, at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 1,000,000,000 or more of the different antibody species present in the blood of an individual. These the antibody species can be expressed by plasmablasts, plasma cells, memory B cells, long-lived plasma cells, naïve B cells, other B lineage cells, or combinations thereof.

Vectors

[00196] In some aspects, a composition can include a vector. Vectors can be used in the transformation of a host cell with a nucleic acid sequence. In some aspects, a vector can include one or more polynucleotides described herein. In one embodiment, a library of nucleic acid sequences encoding target polypeptides may be introduced into a population of cells, thereby allowing screening of a library. The term "vector" is used to refer to a carrier nucleic acid molecule into which a nucleic acid sequence can be inserted for introduction into a cell where it can be replicated. A nucleic acid sequence can be "exogenous" or "heterologous" which means that it is foreign to the cell into which the vector is being introduced or that the sequence is homologous to a sequence in the cell but in a position within the host cell nucleic acid in which the sequence is ordinarily not found. Vectors include plasmids, cosmids, and viruses (e.g., bacteriophage). One of skill in the art may construct a vector through standard recombinant techniques, which are described in Maniatis et al., 1988 and Ausubel et al., 1994, both of which references are incorporated herein by reference. In some aspects, a vector can be a vector with the constant regions of an antibody pre-engineered in. In this way, one of skill can clone just the VDJ regions of an antibody of interest and clone those regions into the pre-engineered vector.

[00197] The term "expression vector" refers to a vector containing a nucleic acid sequence coding for at least part of a gene product capable of being transcribed. In some cases, RNA molecules are then translated into a protein, polypeptide, or peptide. Expression vectors can contain a variety of "control sequences," which refer to nucleic acid sequences for the

transcription and possibly translation of an operably linked coding sequence in a particular host organism. In addition to control sequences that govern transcription and translation, vectors and expression vectors may contain nucleic acid sequences that serve other functions as well.

[00198] In some aspects, a vector can include a promoter. In some aspects, a vector can include an enhancer. A "promoter" is a control sequence that is a region of a nucleic acid sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind such as RNA polymerase and other transcription factors. The phrases "operatively positioned," "operatively linked," "under control," and "under transcriptional control" mean that a promoter is in a correct functional location and/or orientation in relation to a nucleic acid sequence to control transcriptional initiation and/or expression of that sequence. A promoter may or may not be used in conjunction with an "enhancer," which refers to a cis-acting regulatory sequence involved in the transcriptional activation of a nucleic acid sequence.

[00199] A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment and/or exon. Such a promoter can be referred to as "endogenous." Similarly, an enhancer may be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence. Alternatively, certain advantages will be gained by positioning the coding nucleic acid segment under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with a nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers also to an enhancer not normally associated with a nucleic acid sequence in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes, and promoters or enhancers isolated from any other prokaryotic cell, and promoters or enhancers not "naturally occurring," i.e., containing different elements of different transcriptional regulatory regions, and/or mutations that alter expression. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR, in connection with the compositions disclosed herein (see U.S. Pat. No. 4,683,202, U.S. Pat. No. 5,928,906, each incorporated herein by reference).

[00200] In some aspects, a promoter and/or enhancer that effectively directs the expression of the DNA segment in the cell type chosen for expression. One example of such promoter that may be used is the E. coli arabinose or T7 promoter. Those of skill in the art of molecular biology generally are familiar with the use of promoters, enhancers, and cell type combinations for protein expression, for example, see Sambrook et al. (1989), incorporated herein by

reference. The promoters employed may be constitutive, tissue-specific, inducible, and/or useful under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins and/or peptides. The promoter may be heterologous or endogenous.

[00201] In some aspects, vectors can include initiation signals and/or internal ribosome binding sites. A specific initiation signal also may be included for efficient translation of coding sequences. These signals include the ATG initiation codon or adjacent sequences. Exogenous translational control signals, including the ATG initiation codon, may need to be provided. One of ordinary skill in the art would readily be capable of determining this and providing the necessary signals. It is well known that the initiation codon must be "in-frame" with the reading frame of the desired coding sequence to ensure translation of the entire insert. The exogenous translational control signals and initiation codons can be either natural or synthetic. The efficiency of expression may be enhanced by the inclusion of appropriate transcription enhancer elements.

[00202] In some aspects, a vector can include sequences that increase or optimize the expression level of the DNA segment encoding the gene of interest. An example of such sequences includes addition of introns in the expressed mRNA (Brinster, R.L. et al. (1988) Introns increase transcriptional efficiency in transgenic mice. Proc. Natl. Acad. Sci. USA 85, 836–40; Choi, T. et al. (1991) A generic intron increases gene expression in transgenic mice. Mol. Cell. Biol. 11, 3070–4). Another example of a method for optimizing expression of the DNA segment is "codon optimization". Codon optimization involves insertion of silent mutations in the DNA segment to reduce the use of rare codons to optimize protein translation (Codon engineering for improved antibody expression in mammalian cells. Carton JM, Sauerwald T, Hawley-Nelson P, Morse B, Peffer N, Beck H, Lu J, Cotty A, Amegadzie B, Sweet R. Protein Expr Purif. 2007 Oct;55(2):279-86. Epub 2007 Jun 16.).

[00203] In some aspects, a vector can include multiple cloning sites. Vectors can include a multiple cloning site (MCS), which is a nucleic acid region that contains multiple restriction enzyme sites, any of which can be used in conjunction with standard recombinant technology to digest the vector (see Carbonelli et al., 1999, Levenson et al., 1998, and Cocea, 1997, incorporated herein by reference.) "Restriction enzyme digestion" refers to catalytic cleavage of a nucleic acid molecule with an enzyme that functions only at specific locations in a nucleic acid molecule. Many of these restriction enzymes are commercially available. Use of such enzymes is understood by those of skill in the art. Frequently, a vector is linearized or fragmented using a restriction enzyme that cuts within the MCS to enable exogenous sequences to be ligated to the

vector. "Ligation" refers to the process of forming phosphodiester bonds between two nucleic acid fragments, which may or may not be contiguous with each other. Techniques involving restriction enzymes and ligation reactions are well known to those of skill in the art of recombinant technology.

[00204] In some aspects, a vector can include a termination signal. The vectors or constructs will generally comprise at least one termination signal. A "termination signal" or "terminator" is comprised of the DNA sequences involved in specific termination of an RNA transcript by an RNA polymerase. Thus, in certain embodiments, a termination signal that ends the production of an RNA transcript is contemplated. A terminator may be necessary in vivo to achieve desirable message levels.

[00205] Terminators contemplated for use include any known terminator of transcription described herein or known to one of ordinary skill in the art, including but not limited to, for example, rho dependent or rho independent terminators. In certain embodiments, the termination signal may be a lack of transcribable or translatable sequence, such as due to a sequence truncation.

[00206] In some aspects, a vector can include an origin of replication.

[00207] In order to propagate a vector in a host cell, it may contain one or more origins of replication sites (often termed "ori"), which is a specific nucleic acid sequence at which replication is initiated.

[00208] In some aspects, a vector can include one or more selectable and/or screenable markers. In certain embodiments, cells containing a nucleic acid construct may be identified in vitro or in vivo by including a marker in the expression vector. Such markers would confer an identifiable change to the cell permitting easy identification of cells containing the expression vector. Generally, a selectable marker is one that confers a property that allows for selection. A positive selectable marker is one in which the presence of the marker allows for its selection, while a negative selectable marker is one in which its presence prevents its selection. An example of a positive selectable marker is a drug resistance marker.

[00209] Usually the inclusion of a drug selection marker aids in the cloning and identification of transformants, for example, genes that confer resistance to neomycin, puromycin, hygromycin, DHFR, GPT, zeocin and histidinol are useful selectable markers. In addition to markers conferring a phenotype that allows for the discrimination of transformants based on the implementation of conditions, other types of markers including screenable markers such as GFP, whose basis is colorimetric analysis, are also contemplated. Alternatively, screenable enzymes such as chloramphenicol acetyltransferase (CAT) may be utilized. One of skill in the art would

also know how to employ immunologic markers, possibly in conjunction with FACS analysis. The marker used is not believed to be important, so long as it is capable of being expressed simultaneously with the nucleic acid encoding a gene product. Further examples of selectable and screenable markers are well known to one of skill in the art.

[00210] In one aspect, the vector can express DNA segments encoding multiple polypeptides of interest. For example, DNA segments encoding both the immunoglobulin heavy chain and light chain can be encoded and expressed by a single vector. In one aspect, both DNA segments can be included on the same expressed RNA and internal ribosome binding site (IRES) sequences used to enable expression of the DNA segments as separate polypeptides (Pinkstaff JK, Chappell SA, Mauro VP, Edelman GM, Krushel LA., Internal initiation of translation of five dendritically localized neuronal mRNAs., Proc Natl Acad Sci U S A. 2001 Feb 27;98(5):2770-5. Epub 2001 Feb 20.). In another aspect, each DNA segment has its own promoter region resulting in expression of separate mRNAs (Andersen CR, Nielsen LS, Baer A, Tolstrup AB, Weilguny D. Efficient Expression from One CMV Enhancer Controlling Two Core Promoters. Mol Biotechnol. 2010 Nov 27. [Epub ahead of print]).

Host cells and Expression Systems

[00211] In some aspects, a composition can include a host cell. In some aspects, a host cell can include a polynucleotide or vector described herein. In some aspects, a host cell can include a eukaryotic cell (e.g., insect, yeast, or mammalian) or a prokaryotic cell (e.g., bacteria). In the context of expressing a heterologous nucleic acid sequence, "host cell" can refer to a prokaryotic cell, and it includes any transformable organism that is capable of replicating a vector and/or expressing a heterologous gene encoded by a vector. A host cell can, and has been, used as a recipient for vectors. A host cell may be "transfected" or "transformed," which refers to a process by which exogenous nucleic acid is transferred or introduced into the host cell. A transformed cell includes the primary subject cell and its progeny.

[00212] In particular embodiments, a host cell is a Gram negative bacterial cell. These bacteria are suited for use in that they possess a periplasmic space between the inner and outer membrane and, particularly, the aforementioned inner membrane between the periplasm and cytoplasm, which is also known as the cytoplasmic membrane. As such, any other cell with such a periplasmic space could be used. Examples of Gram negative bacteria include, but are not limited to, E. coli, Pseudomonas aeruginosa, Vibrio cholera, Salmonella typhimurium, Shigella flexneri, Haemophilus influenza, Bordotella pertussi, Erwinia amylovora, Rhizobium sp. The Gram negative bacterial cell may be still further defined as bacterial cell which has been

transformed with the coding sequence of a fusion polypeptide comprising a candidate binding polypeptide capable of binding a selected ligand. The polypeptide is anchored to the outer face of the cytoplasmic membrane, facing the periplasmic space, and may comprise an antibody coding sequence or another sequence. One means for expression of the polypeptide is by attaching a leader sequence to the polypeptide capable of causing such directing.

[00213] Numerous prokaryotic cell lines and cultures are available for use as a host cell, and they can be obtained through the American Type Culture Collection (ATCC), which is an organization that serves as an archive for living cultures and genetic materials. An appropriate host can be determined by one of skill in the art based on the vector backbone and the desired result. A plasmid or cosmid, for example, can be introduced into a prokaryote host cell for replication of many vectors. Bacterial cells used as host cells for vector replication and/or expression include DH5-alpha, JM109, and KC8, as well as a number of commercially available bacterial hosts such as SURE™ Competent Cells and SOLOPACK™ Gold Cells (STRATAGENE™, La Jolla). In some aspects, other bacterial cells such as E. coli LE392 are contemplated for use as host cells.

[00214] Many host cells from various cell types and organisms are available and would be known to one of skill in the art. Similarly, a viral vector may be used in conjunction with a prokaryotic host cell, particularly one that is permissive for replication or expression of the vector. Some vectors may employ control sequences that allow it to be replicated and/or expressed in both prokaryotic and eukaryotic cells. One of skill in the art would further understand the conditions under which to incubate all of the above described host cells to maintain them and to permit replication of a vector. Also understood and known are techniques and conditions that would allow large-scale production of vectors, as well as production of the nucleic acids encoded by vectors and their cognate polypeptides, proteins, or peptides.

[00215] In some aspects, a host cell is mammalian. Examples include CHO cells, CHO-K1 cells, or CHO-S cells. Other mammalian host cells include NS0 cells and CHO cells that are dhfr-, e.g., CHO-dhfr-, DUKX-B11 CHO cells, and DG44 CHO cells.

[00216] Numerous expression systems exist can that comprise at least a part or all of the compositions disclosed herein. Expression systems can include eukaryotic expression systems and prokaryotic expression systems. Such systems could be used, for example, for the production of a polypeptide product identified as capable of binding a particular ligand. Prokaryote-based systems can be employed to produce nucleic acid sequences, or their cognate polypeptides, proteins and peptides. Many such systems are commercially and widely available. Other examples of expression systems comprise of vectors containing a strong prokaryotic

promoter such as T7, Tac, Trc, BAD, lambda pL, Tetracycline or Lac promoters, the pET Expression System and an E. coli expression system.

Polypeptides

[00217] In some aspects, a composition can include a polypeptide. In some aspects, a polypeptide encoded by a polynucleotide described herein can be expressed, e.g., from a host cell. The terms "polypeptide" or "protein" include a macromolecule having the amino acid sequence of a native protein, that is, a protein produced by a naturally-occurring and non-recombinant cell; or it is produced by a genetically-engineered or recombinant cell, and comprise molecules having the amino acid sequence of the native protein, or molecules having deletions from, additions to, and/or substitutions of one or more amino acids of the native sequence. The term also includes amino acid polymers in which one or more amino acids are chemical analogs of a corresponding naturally-occurring amino acid and polymers. The terms "polypeptide" and "protein" encompass antigen binding proteins, antibodies, or sequences that have deletions from, additions to, and/or substitutions of one or more amino acids of antigen-binding protein. The term "polypeptide fragment" refers to a polypeptide that has an amino-terminal deletion, a carboxyl-terminal deletion, and/or an internal deletion as compared with the full-length native protein. Such fragments can also contain modified amino acids as compared with the native protein. In certain embodiments, fragments are about five to 500 amino acids long. For example, fragments can be at least 5, 6, 8, 10, 14, 20, 50, 70, 100, 110, 150, 200, 250, 300, 350, 400, or 450 amino acids long. Useful polypeptide fragments include immunologically functional fragments of antibodies, including binding domains. In the case of a binding antibody, useful fragments include but are not limited to a CDR region, a variable domain of a heavy and/or light chain, a portion of an antibody chain or just its variable region including two CDRs, and the like.

[00218] The term "isolated protein" means that a subject protein (1) is free of at least some other proteins with which it would normally be found, (2) is essentially free of other proteins from the same source, e.g., from the same species, (3) is expressed by a cell from a different species, (4) has been separated from at least about 50 percent of polynucleotides, lipids, carbohydrates, or other materials with which it is associated in nature, (5) is operably associated (by covalent or noncovalent interaction) with a polypeptide with which it is not associated in nature, or (6) does not occur in nature. Typically, an "isolated protein" constitutes at least about 5%, at least about 10%, at least about 25%, or at least about 50% of a given sample. Genomic DNA, cDNA, mRNA or other RNA, nucleic acids of synthetic origin, or any combination

thereof can encode such an isolated protein. Preferably, the isolated protein is substantially free from proteins or polypeptides or other contaminants that are found in its natural environment that would interfere with its therapeutic, diagnostic, prophylactic, research or other use.

[00219] In some aspects, a polypeptide can include an antigen binding protein (ABP). An "antigen binding protein" ("ABP") as used herein means any protein that binds a specified target antigen. "Antigen binding protein" includes but is not limited to antibodies and binding parts thereof, such as immunologically functional fragments. Peptibodies are another example of antigen binding proteins. The term "immunologically functional fragment" (or simply "fragment") of an antibody or immunoglobulin chain (heavy or light chain) antigen binding protein, as used herein, is a species of antigen binding protein comprising a portion (regardless of how that portion is obtained or synthesized) of an antibody that lacks at least some of the amino acids present in a full-length chain but which is still capable of specifically binding to an antigen. Such fragments are biologically active in that they bind to the target antigen and can compete with other antigen binding proteins, including intact antibodies, for binding to a given epitope. In some embodiments, the fragments are neutralizing fragments. These biologically active fragments can be produced by recombinant DNA techniques, or can be produced by enzymatic or chemical cleavage of antigen binding proteins, including intact antibodies. Immunologically functional immunoglobulin fragments include, but are not limited to, Fab, a diabody (heavy chain variable domain on the same polypeptide as a light chain variable domain, connected via a short peptide linker that is too short to permit pairing between the two domains on the same chain), Fab', F(ab')2, Fv, domain antibodies and single-chain antibodies, and can be derived from any mammalian source, including but not limited to human, mouse, rat, camelid or rabbit. It is further contemplated that a functional portion of the antigen binding proteins disclosed herein, for example, one or more CDRs, could be covalently bound to a second protein or to a small molecule to create a therapeutic agent directed to a particular target in the body, possessing bifunctional therapeutic properties, or having a prolonged serum half-life. As will be appreciated by one of skill in the art, an antigen binding protein can include nonprotein components. Additional details about antigen binding proteins and antibodies such as modifications, variants, methods of making, and methods of screening can be found in U.S. Pat. Pub. 20110027287, herein incorporated by reference in its entirety for all purposes.

[00220] In some aspects, a polypeptide can include an antibody. The term "antibody" refers to an intact immunoglobulin of any isotype, or a fragment thereof that can compete with the intact antibody for specific binding to the target antigen, and includes, for instance, chimeric, humanized, fully human, and bispecific antibodies. An "antibody" is a species of an antigen

binding protein. An intact antibody will generally comprise at least two full-length heavy chains and two full-length light chains, but in some instances can include fewer chains such as antibodies naturally occurring in camelids which can comprise only heavy chains. Antibodies can be derived solely from a single source, or can be "chimeric," that is, different portions of the antibody can be derived from two different antibodies. The antigen binding proteins, antibodies, or binding fragments can be produced in hybridomas, by recombinant DNA techniques, or by enzymatic or chemical cleavage of intact antibodies. Unless otherwise indicated, the term "antibody" includes, in addition to antibodies comprising two full-length heavy chains and two full-length light chains, derivatives, variants, fragments, and muteins thereof. Furthermore, unless explicitly excluded, antibodies include monoclonal antibodies, bispecific antibodies, minibodies, domain antibodies, synthetic antibodies (sometimes referred to herein as "antibody mimetics"), chimeric antibodies, humanized antibodies, human antibodies, antibody fusions (sometimes referred to herein as "antibody conjugates"), and fragments thereof, respectively. In some embodiments, the term also encompasses peptibodies.

[00221] A therapeutically effective amount of an ABP can be administered to a subject in need thereof. ABPs can be formulated in pharmaceutical compositions. These compositions can comprise, in addition to one or more of the ABPs, a pharmaceutically acceptable excipient, carrier, buffer, stabilizer or other materials well known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material can depend on the route of administration, e.g. oral, intravenous, cutaneous or subcutaneous, nasal, intramuscular, intraperitoneal routes.

[00222] Pharmaceutical compositions for oral administration can be in tablet, capsule, powder or liquid form. A tablet can include a solid carrier such as gelatin or an adjuvant. Liquid pharmaceutical compositions generally include a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene glycol can be included.

[00223] For intravenous, cutaneous or subcutaneous injection, or injection at the site of affliction, the active ingredient will be in the form of a parenterally acceptable aqueous solution which is pyrogen-free and has suitable pH, isotonicity and stability. Those of relevant skill in the art are well able to prepare suitable solutions using, for example, isotonic vehicles such as Sodium Chloride Injection, Ringer's Injection, Lactated Ringer's Injection. Preservatives, stabilizers, buffers, antioxidants and/or other additives can be included, as required.

[00224] ABP administration is preferably in a "therapeutically effective amount" or "prophylactically effective amount" (as the case can be, although prophylaxis can be considered therapy), this being sufficient to show benefit to the individual. The actual amount administered, and rate and time-course of administration, will depend on the nature and severity of disease being treated. Prescription of treatment, e.g. decisions on dosage etc., is within the responsibility of general practitioners and other medical doctors, and typically takes account of the disorder to be treated, the condition of the individual patient, the site of delivery, the method of administration and other factors known to practitioners. Examples of the techniques and protocols mentioned above can be found in Remington's Pharmaceutical Sciences, 16th edition, Osol, A. (ed), 1980.

[00225] A composition can be administered alone or in combination with other treatments, either simultaneously or sequentially dependent upon the condition to be treated.

Immune cells

[00226] A sample can include immune cells. The immune cells can include T cells and B cells. T-cells (T lymphocytes) include, for example, cells that express T cell receptors. B-cells include, for example, activated B cells, blasting B cells, plasma cells, plasmablasts, memory B cells, B1 cells, B2 cells, marginal-zone B cells, and follicular B cells. T cells include activated T cells, blasting T cells, Helper T cells (effector T cells or Th cells), cytotoxic T cells (CTLs), memory T cells, central memory T cells, effector memory T cells and regulatory T cells. A sample can include a single cell in some applications (e.g., a calibration test to define relevant T or B cells) or more generally at least 1,000, at least 10,000, at least 100,000, at least 250,000, at least 500,000, at least 750,000, or at least 1,000,000 cells.

B cells

[00227] As used herein a "B cell" refers to any cell that has at least one rearranged immunoglobulin gene locus. A B cell can include at least one rearranged immunoglobulin heavy chain locus or at least one rearranged immunoglobulin light chain locus. A B cell can include at least one rearranged immunoglobulin heavy chain locus and at least one rearranged immunoglobulin light chain locus. B cells are lymphocytes that are part of the adaptive immune system. B cells can include any cells that express antibodies either in the membrane-bound form as the B-cell receptor (BCR) on the cell surface or as secreted antibodies. B cells can express immunoglobulins (antibodies, B cell receptor). Antibodies can include heterodimers formed from the heavy and light immunoglobulin chains. The heavy chain is formed from gene rearrangements of the variable, diversity, and junctional (VDJ) genes to form the variable

region, which is joined to the constant region. The light chain is formed from gene rearrangements of the variable and junctional (VJ) genes to form the variable region, which is then joined to the constant region. Owing to a large possible number of junctional combinations, the variable regions of the antibody gene (which is also the BCR) have huge diversity, enabling B cells to recognize any foreign antigen and mount a response against it.

B-cell activation and differentiation

[00228] B cells are activated and differentiate when they recognize an antigen in the context of an inflammatory immune response. They usually include 2 signals to become activated, one signal delivered through BCR (a membrane-bound form of the rearranged immunoglobulin), and another delivered through CD40 or another co-stimulatory molecule. This second signal can be provided through interaction with helper T cells, which express the ligand for CD40 (CD40L) on their surface. B cells then proliferate and may undergo somatic hypermutation, where random changes in the nucleotide sequences of the antibody genes are made, and B cells whose antibodies have a higher affinity B cells are selected. They may also undergo "class-switching", in which the constant region of the heavy chain encoding the IgM isotype is switched to the constant region encoding the IgG, IgA, or IgE isotype. Differentiating B cells may end up as memory B cells, which are usually of higher affinity and classed switched, though some memory B cells are still of the IgM isotype. Memory B cells can also become activated and differentiate into plasmablasts and ultimately, into plasma cells. Differentiating B cells may also first become plasmablasts, which then differentiate to become plasma cells.

Affinity maturation and clonal families

[00229] A clonal family is generally defined by the use of related immunoglobulin heavy chain and/or light chain V(D)J sequences by 2 or more samples. Related immunoglobulin heavy chain V(D)J sequences can be identified by their shared usage of V(D)J gene segments encoded in the genome. Within a clonal family there are generally subfamilies that vary based on shared mutations within their V(D)J segments, that can arise during B cell gene recombination and somatic hypermutation.

[00230] Activated B cells migrate and form germinal centers within lymphoid or other tissues, where they undergo affinity maturation. B cells may also undergo affinity maturation outside of germinal centers. During affinity maturation, B cells undergo random mutations in their antibody genes, concentrated in the complementary determining regions (CDRs) of the genes, which encode the parts of the antibody that directly bind to and recognize the target antigen against which the B cell was activated. This creates sub-clones from the original

proliferating B cell that express immunoglobulins that are slightly different from the original clone and from each other. Clones compete for antigen and the higher-affinity clones are selected, while the lower-affinity clones die by apoptosis. This process results in the "affinity maturation" of B cells and consequently in the generation of B cells expressing immunoglobulins that bind to the antigen with higher affinity. All the B cells that originate from the same 'parent' B cell form clonal families, and these clonal families include B cells that recognize the same or similar antigenic epitopes. In some aspects, we expect that clones present at higher frequencies represent clones that bind to antigen with higher affinity, because the highest-affinity clones are selected during affinity maturation. In some aspects, clones with different V(D)J segment usage exhibit different binding characteristics. In some aspects, clones with the same V(D)J segment usage but different mutations exhibit different binding characteristics.

Memory B cells

[00231] Memory B cells are usually affinity-matured B cells, and may be class-switched. These are cells that can respond more rapidly to a subsequent antigenic challenge, significantly reducing the time included for affinity-matured antibody secretion against the antigen from ~14 days in a naive organism to ~7 days.

Plasmablasts and plasma cells

[00232] Plasma cells can be either long-lived or short-lived. Long-lived plasma cells may survive for the lifetime of the organism, whereas short-lived plasma cells can last for 3-4 days. Long-lived plasma cells reside either in areas of inflammation, in the mucosal areas (in the case of IgA-secreting plasma cells), in secondary lymphoid tissues (such as the spleen or lymph nodes), or in the bone marrow. To reach these divergent areas, plasmablasts fated to become long-lived plasma cells may first travel through the bloodstream before utilizing various chemokine gradients to traffic to the appropriate areas. Plasmablasts are cells that are affinity matured, are typically classed-switched, and usually secrete antibodies, though generally in lower quantities than the quantity of antibody produced by plasma cells. Plasma cells are dedicated antibody secretors.

Characteristics of TCR and BCR genes

[00233] Since identifying recombinations are present in the DNA of each individual adaptive immune cell as well as their associated RNA transcripts, either RNA or DNA can be sequenced. A recombined sequence from a T-cell or B-cell can also be 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 alpha, beta, gamma, or delta chains of a TCR. In a majority of T-cells, the TCR is a heterodimer consisting of an alpha-chain and beta-chain. The TCR-alpha chain is generated by VJ recombination, and the beta chain receptor is generated by V(D)J recombination. For the TCR-beta 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 gamma and delta chains. The TCR gamma chain is generated by VJ recombination, and the TCR delta 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).

[00234] The DNA and RNA analyzed in the methods can correspond to sequences encoding heavy chain immunoglobulins (IgH) with constant regions (alpha, delta, gamma, epsilon, or mu) or light chain immunoglobulins (IgK or IgL) with constant regions lambda or kappa. Each antibody can have 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. In this process mutations occur in those B-cells that are able to recognize the specific epitope leading to greater diversity in antibodies that may be able to bind the specific epitope more strongly. All these factors contribute to great diversity of antibodies generated by the B-cells. Many billions and maybe more than a trillion distinct antibodies may be generated. The basic premise for generating T-cell diversity is similar to that for generating antibodies by B-cells. An element of T-cell and B-cell activation is their binding to epitopes. The activation of a specific cell leads to the production of more of the same type of cells leading to a clonal expansion.

[00235] Complementarity determining regions (CDR), or hypervariable regions, are sequences in the variable domains of antigen receptors (e.g., T cell receptor and immunoglobulin) that can bind an antigen. The chain of each antigen receptor contains three CDRs (CDR1, CDR2, and CDR3). The two polypeptides making T cells (alpha and beta) and immunoglobulin (IgH and IgK or IgL) contribute to the formation of the three CDRs.

[00236] The part of CDR1 and CDR2 that is coded for by TCR-beta lies within one of 47 functional V segments. Most of the diversity of CDRs is found in CDR3, with the diversity being generated by somatic recombination events during the development of T lymphocytes.

[00237] A great diversity of BCR is present inter and intra-individuals. The BCR is composed of two genes IgH and IgK (or IgL) coding for antibody heavy and light chains. Three Complementarity Determining Region (CDR) sequences that bind antigens and MHC molecules have the most diversity in IgH and IgK (or IgL). The part of CDR1 and CDR2 coded for by IgH lies within one of 44 functional V segments. Most of the diversity in naive B cells emerges in the generation of CDR3 through somatic recombination events during the development of B lymphocytes. The recombination can generate a molecule with one of each of the V, D, and J segments. In humans, there are 44 V, 27 D, and 6 J segments; thus, there is a theoretical possibility of more than 7,000 combinations. In a small fraction of BCRs (about 5%) two D segments are found. Furthermore, several bases may be deleted and others added (called N and P nucleotides) at each of the two junctions generating a great degree of diversity. After B cell activation a process of affinity maturation through somatic hypermutation occurs. In this process progeny cells of the activated B cells accumulate distinct somatic mutations throughout the gene with higher mutation concentration in the CDR regions leading to generating antibodies with higher affinity to the antigens. In addition to somatic hypermutation activated B cells undergo the process of isotype switching. Antibodies with the same variable segments can have different forms (isotypes) depending on the constant segment. Whereas all naive B cells express IgM (or IgD), activated B cells mostly express IgG but also IgM, IgA and IgE. This expression switching from IgM (and/or IgD) to IgG, IgA, or IgE occurs through a recombination event causing one cell to specialize in producing a specific isotype. There is one segment for each IgM, IgD, and IgE, two segments for IgA, and four segments for IgG.

Methods

Application to health care and biotechnology uses

[00238] Use of the compositions and methods described herein to identify antibodies and TCRs and to group antibody and TCR sequences into clonal families has many useful and novel

applications to health care and biotechnology research. Antibody clonal families can comprise affinity-matured non-identical clones and TCR clonal families can comprise identical clones. These applications include, but are not limited to: 1) the discovery and development of antibody or antibody-derived therapeutics; 2) the discovery and development of diagnostics; 3) the discovery and development of research tools useful in health and biotechnology research; and 4) the development and assessment of candidate vaccines and identification of antigens useful as vaccine components.

[00239] As the invention can be used with any type of B or T cell, the cell source and specific B or T cell subtype(s) are chosen based on the profile of the desired ultimate product. Examples of specific subclasses of B or T cells and their use are described in the subsection, "Isolation and enrichment of cells and cell subpopulations" in the General Materials and Methods section. In general, the cells can be from a particular human or animal subject having a particular clinical state or course of disease, or having received a particular treatment regimen, or having been exposed to a particular challenge, immunization, or set of conditions that induces an immune response.

Application to discovery and development of therapeutics, diagnostics, and research tools

[00240] To develop an antibody or molecule derived from an antibody for use as a therapeutic, diagnostic, or research tool, the antibody and/or derivatives of the antibody's antigen-binding regions can first be identified or discovered as binding a/the desired antigen(s) or epitope(s) and/or having a desired functional consequence in an in vivo or in vitro system. These candidate antibodies are then further screened for other desired properties specific to the intended product. These target product properties will be different for different types of therapeutic, diagnostic, and research tool antibodies, and the invention provides a useful means of identifying candidates for further development toward any of these product paths.

[00241] Based on the desired profile of the properties of the ultimate product, the source of relevant B cells can be, but is not limited to, a patient with disease, such as an infectious disease, cancer, or an autoimmune condition; a patient receiving a treatment, such as cancer therapy or a vaccine; or an animal with disease or treated in a manner to induce an immune response, such as immunization or induction/establishment of a disease model.

[00242] In general, candidate antibodies, or candidate macromolecules derived from the antigen-binding regions, that are intended for development as therapeutics, diagnostics, or research tools are discovered via multiple technologies that fall into one of two general approaches: 1) isolation of antibodies of interest from B cells of a human's or an animal's

immune response; and 2) isolation of antibodies derived from expression libraries of immunoglobulin molecules, or derivatives thereof, expressed heterologously and screened using one or more display technologies (reviewed in Hoogenboom HR, Trends Biotechnol., 1997,15:62-70; Hammond PW, MAbs, 2010, 2:157-64; Nissim A, Chernajovsky Y, Handb Exp Pharmacol., 2008, (181):3-18; Steinitz M, Hum Antibodies, 2009;18:1-10; Bradbury AR, Sidhu S, Dübel S, and McCafferty, Nat Biotechnol., 2011, 29:245-54; *Antibody Engineering* (Kontermann RE and Dübel S eds., Springer, 2nd edition)).

[00243] For the former approach (#1), candidate antibodies are selected from particular clonal families identified from relevant donors as described in, e.g., the General Materials and Methods section. The invention can be applied as described to the appropriate B-cells (e.g. blasting B-cells) from the appropriate human donor or animal to discover or identify candidate antibodies. For example, for a cancer therapeutic antibody candidate, the appropriate human donor can be a patient who has successfully suppressed cancer progression via an immune response; or for a particular diagnostic antibody candidate, the appropriate donor can be a patient who has autoantibodies against the diagnostic marker or a mouse immunized against the marker; or for an antibody reagent tool candidate, the appropriate donor can be a mouse, a rabbit, a goat, , a rat, a horse, a chicken, a dog, or other animal immunized with the target molecule and/or epitope that the antibody reagent is meant to recognize. Sequences and selection of antibodies for expression and testing can be performed as described in the General Materials and Methods section. Such applications of the technology can provide candidate antibodies often obtained via more laborious and time-consuming methods (e.g. hybridoma technology, virus-induced immortalization of B cells, etc).

[00244] For the latter approach (#2), a subset of, or the entire set of paired heavy and light chain sequences from a one or more human or animal antibody repertoires, obtained as in #1, are used to seed expression libraries containing identification regions to track sample origin and original cognate pairs from the sample when a library and/or a selected/enriched subset of a library is sequenced using a next generation sequencing platform. Variable regions and framework region information can be incorporated into one or more antibody display library formats to discover candidate antibodies. Variable regions of Ig genes can be cloned and incorporated into expression vectors using methods described in the subsection, "Cloning and expression of cloned light and heavy chain immunoglobulin pairs" in the General Materials and Methods section. For example, fragments and/or domains from cognate pair heavy and light chains obtained as in #1 can be used to seed Fab yeast (Weaver-Feldhaus JM, Lou J, Coleman JR, et al., FEBS Lett, 2004, 564:24-34) or phagemid (Kashyap AK, Steel J, Oner AF, et al., Proc

Natl Acad Sci U S A, 2008, 105:5986-91) libraries with identification region tracking of each chain to the proper, original B cell of origin regardless of combinatorial matching of different heavy and light chains into non-endogenous (non-cognate) pairings. The cognate pair heavy and light chains obtained as in #1 can also be used with other display platforms, beyond phagemid or yeast, and can be used with other antibody derivative expression constructs beyond Fab fragment expression constructs [*Antibody Engineering* (Kontermann, RE and Dübel S eds., Springer, 2nd edition)]. In an alternate application of the identification regions, identification regions can be added to already existing display libraries to provide the benefits of identification region tracking and error correction of next generation sequencing data. Depending on the library type, format, and expression/display system, identification regions can be incorporated using PCR reactions or reverse transcriptase followed by PCR reactions (see, e.g., the subsection, "Sequencing of paired light and heavy chain immunoglobulin genes from single B-cells" in the General Materials and Methods section).

[00245] Candidate antibodies, whether from B cell repertoires (see, e.g., General Materials and Methods section) or display expression library "repertoires" (Kashyap AK, Steel J, Oner AF, et al., Proc Natl Acad Sci U S A, 2008, 105:5986-91; Weaver-Feldhaus JM, Lou J, Coleman JR, et al., FEBS Lett, 2004, 564:24-34; Ravn U, Gueneau F, Baerlocher L, et al., Nucleic Acids Res, 2010, 38:e193; *Antibody Engineering* (Kontermann, RE and Dübel S eds., Springer, 2nd edition), are identified by expressing and testing the antibody or antibody-derivative molecules, or libraries of molecules, in assays for binding against desired antigen/target(s) and/or epitope(s) or in assays for testing of functional consequence in an in vivo or in vitro (including ex vivo samples/preparations) setting. Published reports have described the use of identification regions to track the donor source of antibody sequences obtained from a B-cell repertoire for use in an expression library (e.g. Kashyap AK, Steel J, Oner AF, et al., Proc Natl Acad Sci U S A, 2008, 105:5986-91). The identification region technology described herein, uniquely provides useful improvements upon such identification region usage. The invention: 1) provides a means to track, not only each donor, but each donor's B cells for cognate pairing of heavy and light chains; 2) provides a means to index back to the original B cell sample for retrieval of more sample for cloning and/or testing; 3) provides a means of tracking heavy and light chain origin despite non-cognate combinatorial pairings within the expression library; 4) provides a means of tracking heavy and light chain origin across rounds of selection-enrichment (e.g. when monitoring sequence evolution during pool selection in vitro, such as in Ravn U, Gueneau F, Baerlocher L, et al., Nucleic Acids Res, 2010, 38:e193).

[00246] Identification of the most frequently represented heavy or light chain sequences in a B cell immune response repertoire, and combining heavy and light chain pairs based on rank order frequency of the individual chains, has been shown to be a viable way to identify some candidate antibodies, despite the fact that the cognate pair information is not retained in the next generation sequence analyses when performed in this manner (Reddy ST, Ge X, Miklos AE, et al., Nat Biotechnol, 2010, 28:965-9). The invention also allows for this type of frequency analysis methodology, but can further provide a means to use next generation sequencing to assess the frequency of actual antibodies in the repertoire, not simply isolated, independent heavy or light chains.

[00247] Furthermore, the invention provides at least three improvements of significant utility beyond frequency analysis: 1) because the cognate pairing of heavy and light chains can be tracked, the discovery of actual antibodies from the immune response and the actual antibody clonal families produced by the B cells in the immune response can be identified (a clone involves a specific, cognate pair of heavy and light chains that co-evolved from the same cell progenitors and information about natural pairings within the affinity maturation process would improve upon approaches described in the literature to analyze immune responses using next generation sequencing [e.g. Wu X, Yang ZY, Li Y, Hogerkorp CM, et al., Science, 2010, 329:856-61]); 2) identification regions provide the means to minimize, or even eliminate, the effect on sequence analyses of sequencing errors common to next generation sequencing platforms (see, e.g., subsection, "Other sequencing data analysis options" in the General Materials and Methods section); and 3) identification regions provide the ability to link and track $\geq$2 sequences co-expressed at the single cell level.

[00248] For those candidate antibodies that have been identified as having desirable binding properties to an antigen, target, or epitope, or that have a desired functional effect, more antibodies from the respective clonal family can be cloned and expressed (see, e.g., General Materials and Methods section) to test for the presence of similar, but potentially more optimal, antibodies or antibodies that are the same in binding or functional properties but contain other difference of import to the final product profile.

[00249] For cases in which candidates are identified from display expression libraries, identification regions can provide a means to identify antibodies of potentially similar sequence to candidates by identifying sequences that were not selected in screening enrichment but which contain identification regions of the identified candidates and thus are derived from the same original heavy and/or light chains that seeded the library. Antibodies that are lost in rounds of in vitro selection, but are similar to selected, candidate antibodies can be recovered or "rescued"

for further analysis as potential candidates. Such rescue can obviate the effect of bias in expression or assays systems that may miss useful and functional antibodies (Ravn U, Gueneau F, Baerlocher L, et al., Nucleic Acids Res, 2010, 38:e193).

[00250] Once candidates with desired binding and/or functional properties are identified, they can then be advanced to relevant assays and other assessments based on the desired, downstream, product profile. For therapeutic antibodies intended for use in passive immunization, candidates are advanced to assays and preclinical testing models to determine the best candidates for clinical testing in humans or for use in animal health, including, but not limited, assessments of properties such as stability and aggregation, formulation and dosing ease, protein expression and manufacturing, species selectivity, pharmacology, pharmacokinetics, safety and toxicology, absorption, metabolism and target-antibody turnover, as well as immunogenicity [*See, e.g.,* Lynch CM, Hart BW, and Grewal IS, MAbs, 2009, 1: 2–11; Chapman K, Pullen N, Coney L, et al., mAbs, 2009, 1, 505-516; S. Dübel, *Handbook of Therapeutic Antibodies: Technologies, Emerging Developments and Approved Therapeutics* (John Wiley & Sons, 2010); Therapeutic *Monoclonal Antibodies: From Bench to Clinic*, (Z. An ed., John Wiley & Sons, 2009) *Antibody Engineering* (Kontermann RE and Dübel S eds., Springer, 2nd edition)]. Thus many candidates are selected because the majority will be insufficient for therapeutic testing in humans with respect to at least one of the many properties that need to be assessed prior to human testing (i.e. attrition). Clonal families can be mined, e.g., as described above, for candidates similar to ones already characterized, but possibly harboring differences regarding one or more of the properties that are assessed in preclinical work. Specific antibody engineering strategies may need to be employed to optimize for certain properties [*Antibody Engineering* (Kontermann RE and Dübel S eds., Springer, 2nd edition)].

[00251] For diagnostics, the invention can be used to identify antibodies, TCRs, and clonal families produced by infection or vaccination for use in the detection of infectious agents (Selvarajah S, Chatterji U, Kuhn R, et al., 2012, 6:29-37; Berry JD, Vet J, 2005, 170:193-211), as well as for any non-infectious disease, pathological condition, or medical treatment or therapy. Such antibodies, TCRs, and/or clonal families can provide useful diagnostic probes for biomarkers or provide immune system information about the disease state of, or effect of treatment on, a human or animal. As such, specific antibodies or TCRs, or specific clonal families of either immune receptor class can provide utility for diagnostic tools and personalized medicine. In another application to diagnostics, known disease or treatment response biomarkers can be used as immunogens to immunize mice or other animals from which B cells are harvested to identify antibodies (see, e.g., General Materials and Methods section) against

the biomarker which could subsequently be used in diagnostic tests, such as ELISAs or other immunoassays (Selvarajah S, Chatterji U, Kuhn R, et al., 2012, 6:29-37). Once identified as having potential diagnostic utility, candidate antibodies, TCRs, and/or clonal families can be advanced to assays, models, and possibly trials relevant to the desired profile of the diagnostic product [Berry JD, Vet J, 2005, 170:193-211; *Diagnostic and Therapeutic Antibodies* in Methods in Molecular Medicine, Vol. 40 (George AJT and Urch CE eds., Humana Press); *Antibody Engineering* (Kontermann RE and Dübel S eds., Springer, 2nd edition) ; Colwill K, Renewable Protein Binder Working Group, and Gräslund S, Nat Methods, 2011, 8:551-8; Pershad K, Pavlovic JD, Gräslund S, et al., Protein Eng Des Sel, 2010, 23:279-88.]. Specific antibody engineering strategies may need to be employed to optimize for certain properties [*Antibody Engineering* (Kontermann RE and Dübel S eds., Springer, 2nd edition)].

[00252] For research tool antibodies, candidates identified, e.g., as described above, can be advanced to test how they perform in the research application for which the research tool antibody is intended (e.g. immunoprecipitation; immunoblotting; immunostaining and histology; immunoaffinity purification; capture-, and sandwich-, and detection immunoassays; for example as described in *Antibodies: A Laboratory Manual*, E Harlow and D Lane (Cold Spring Harbor Laboratory Press, 1988). Validation criteria will be based on the final intended research use (Colwill K, Renewable Protein Binder Working Group, and Gräslund S, Nat Methods, 2011, 8:551-8; Pershad K, Pavlovic JD, Gräslund S, et al., Protein Eng Des Sel, 2010, 23:279-88). Specific antibody engineering strategies may need to be employed to optimize for certain properties [*Antibody Engineering* (Kontermann RE and Dübel S eds., Springer, 2nd edition)].

Application to vaccine discovery and development

[00253] The invention can be used to identify antibodies, TCRs, and clonal families of each of these immune receptor classes to a vaccine challenge in a human or animal. Specific antibodies can be used as probes to identify the vaccine component(s) recognized by the antibody and the clonal family to which the antibody that was used as a probe belongs. This information about antibody and clonal families can be used to make assessments about the proportions or strength of the immune response targeting particular antigens and/or epitopes of the vaccine. The assessment of antibody immune responses to different vaccine components can be complemented with information collected about the concomitant TCR repertoire response to the vaccine, (see, e.g., subsections, "For other cell types" and "PCR of other immunoglobulin heavy chains and T-cell receptor (TCR) chains" in the General Materials and Methods section). This information can be subsequently used to understand which components of the vaccine, or

which variants of a vaccine, or what adjuvants produce effective or more optimal responses from an immune system of a human or animal (Haynes BF, Gilbert PB, McElrath MJ, et al., N Engl J Med, 2012, 366:1275-86). The approach can also be used to compare individuals or populations in their response to a vaccine.

[00254] Similar analyses can be performed to identify and assess the antibodies, TCRs, and the clonal families produced in response to an actual pathogen and which may correlate with clinical outcomes of interest, such as protective responses to infection (for example, identification of antibodies from survivors of a severe influenza pandemic, Yu X, Tsibane T, McGraw PA, et al., Nature, 2008, 455:532-6; or identification of specific antibodies from HIV-infected individuals with broadly HIV-neutralizing sera, Wu X, Yang ZY, Li Y, Hogerkorp CM, et al., Science, 2010, 329:856-61 and Walker LM, Phogat SK, Chan-Hui PY, et al., Science, 2009, 326:285-9). Identification of such correlates of protection can be compared to the response produced by the vaccine and/or specific vaccine components as described above and the two datasets can be compared to assess the ability of the vaccine to produce immune responses that correlate to desired outcomes seen in cases of actual infection.

[00255] Thus, the invention provides a useful means of obtaining a surrogate readout of disease protection and vaccine response, via antibody and/or TCR repertoire sequence analysis, before a human or animal is challenged with actual infection. Once a clonal family has been identified as binding a particular antigen or epitope, the identification of antigens or epitopes targeted by other immune response repertoires is possible without doing assays in cases where the same or similar clonal families are found across repertoires. Thus, in those cases where enough information about a clonal family and it antigen/epitope binding is known (see, e.g., subsection, "Screening of expressed human antibodies" in the General Materials and Methods section), sequence analysis alone of newly analyzed repertoires can provide a readout of the antigens of that repertoire for the known clonal families that it contains. This application can provide a useful means to monitor responses across one, a few, or many subjects in vaccine clinical trials and to monitor immunity and infectious disease relationships for one, a few, or many people on a population level.

[00256] Furthermore, antibody, TCR, and clonal families that correlate with protection from a pathogen can be used to identify the specific antigens and sets of antigens (including both known and novel antigens) that mediate protective and/or effective immune responses against the pathogen. Identification of the antigens targeted in effective immune responses can be used to guide the selection of antigens to be included in vaccines that are expected to produce protective antibody- and TCR-mediated responses in immunized humans or animals.

[00257] Antibodies, TCR, and clonal families that do not bind known antigens in assays become candidates for identifying potentially novel antigens and/or epitopes of the pathogen against which the antibodies and/or TCRs provide protection. Antibodies known to not bind already known antigens can be used as probes in combination with immunoseparation and mass spectroscopy to identify the previously unidentified antigen or epitope (Zhu YZ, Cai CS, Zhang W, et al., PLoS One, 2010, 5:e13915; and see, e.g., subsections, "Immunoprecipitation of staph antigens with antibodies derived from staph-infected patients" and "Mass spectrometry identification of peptides" in the General Materials and Methods section). Such novel antigens or epitopes can be used as vaccine components that are expected to produce or contribute to the production of protective antibody- and TCR-mediated responses in immunized humans or animals.

[00258] In addition to facilitating development of vaccines for microbial pathogens, the antibody, TCR and clonal families can also be used to develop tumor vaccines. Humans or animals that mount an immune response against a cancer or pre-cancerous cells can yield antibodies, TCR, and clonal families that can be used to identify individual and combinations of antigens that can be incorporated into preventative or therapeutic vaccines for cancer.

Methods for producing one or more polynucleotides of interest

[00259] In some aspects, a method includes obtaining a cDNA library comprising a plurality of cDNAs associated with a plurality of samples obtained from one or more subjects, wherein each cDNA is associated with a single sample in the plurality of samples, and wherein each cDNA associated with each sample is present in a separate container; adding an adapter molecule to the cDNA associated with each sample, wherein the adapter molecule comprises a sample identification region and an adapter region, wherein the sample identification region is coupled to the adapter region, and wherein the sequence of the sample identification region of each adapter molecule is distinct from the sequence of the sample identification region of the other adapter molecules added to each cDNA in the library; and allowing the adapter region to attach to each cDNA in the library to produce the one or more polynucleotides of interest.

[00260] In some aspects, obtaining the cDNA library comprises obtaining the plurality of samples and processing the samples to prepare the cDNA library. In some aspects, obtaining the cDNA library comprises receiving the cDNA library directly or indirectly from a third party that has processed the plurality of samples to prepare the cDNA library.

[00261] In some aspects, the adapter molecule further comprises a universal primer region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample

identification region. In some aspects, each cDNA region comprises an mRNA polynucleotide hybridized to a cDNA polynucleotide.

[00262] In some aspects, each sample comprises a cell. In some aspects, the cell is a B cell. In some aspects, the B cell is a plasmablast, memory B cell, or a plasma cell. In some aspects, each sample comprises a plurality of cells.

[00263] In some aspects, each adapter region is attached to each cDNA via binding, e.g., G:C binding.

[00264] In some aspects, the adapter molecule is single-stranded, and further comprising incorporating the adapter molecule into each cDNA by allowing an enzyme to make the adapter molecule double-stranded. In some aspects, the adapter molecule is incorporated into each cDNA to produce the polynucleotide of interest by an MMLV H^- reverse transcriptase.

[00265] In some aspects, methods can include amplification steps such as PCR and other amplification reactions generally known in the art.

Methods for linking and barcoding polynucleotides of interest

[00266] In some aspects, the method includes the linking of two polynucleotide sequences of interest, e.g., an antibody light chain (LC) and heavy chain (HC) from a single sample, and providing one or more barcode or sequence identification sequences. In this aspect, there is provided a physical linkage between the polynucleotide sequences of interest as well as one or more barcodes to provide an identifier to allow polynucleotide sequences derived from a particular source or sample to be determined, e.g., single cell, sample well, single sample, etc. Single samples can comprise one or more B-lineage cells or other cell types. Examples of methods to link two polynucleotide sequences of interest are known in the art, for example, WO 99/16904, WO 93/03151, and US 7,749,697, which are hereby incorporated by reference. Among other advantages associated with the use of barcodes on linked polynucleotide sequences include facilitation of high-throughput sequencing and mapping of a sequence back to an original sample so that it can be re-sequenced and PCR cloned to express the polynucleotide sequences, e.g., HC and LC immunoglobulin polynucleotides. Some of the high-throughput sequencing technologies exhibit sequencing error rates of 1-10+%, and the use of barcodes enables repeat sequencing of templates to facilitate bioinformatic error correction. This is particularly important for distinguishing sequencing errors from gene variations, such as those in immunoglobulin polynucleotides. Specifically, it can be difficult to ascertain if closely related sequences are in fact distinct sequences or if they instead represent artifacts produced by sequencing errors. Barcodes, by enabling analysis of repeat sequencing of individual templates thereby enable sequencing error correction, thus providing determination of whether sequences

are distinct vs. artifacts from sequencing error(s). In one embodiment, the polynucleotide sequences are immunoglobulin HC and LC sequences that have diverged due to somatic hypermutation, and differ by only 1 nucleotide.

[00267] In this aspect, physically linked and barcoded structures as shown in Figure 15 are generally obtained. Figure 15 illustrates the physical linkage of two nucleic acid segments, A and B (e.g., two cDNAs). A barcode (BC) is appended to any one of ends or in the linker connecting A and B. This physical linkage of A and B, as well as, addition of the barcode is accomplished through any of a number of means, including by ligation, recombination, amplification, or overlap-extension, or a combination of these methods, as described in greater detail below. Optionally, additional barcodes can be added to the structure shown in Figure 15, to provide compound barcoding to enable sequencing of a large number of linked polynucleotides using a lesser number of barcodes. Also, it will be appreciated that depending on the particular strategy used to link the two nucleic acid segments, any relative orientation of the segments can be obtained, with respect to sense and antisense orientations, i.e., the segments, such as cDNAs, can be joined head to tail, head to head, or tail to tail.

[00268] Barcodes can be added to the polynucleotide sequences before, during or after physical linkage using methods known in the art. These methods include, but are not limited to, for example, ligation methods, such as blunt end ligation of barcode adaptors, and by the annealing and ligation of compatible ends, such as those generated by homopolymeric tailing, restriction enzyme digestion of a linker, or 3' tailing of cDNA by reverse transcriptase. Barcodes can also be added in amplification reactions using suitable primers carrying barcode sequences. Barcodes can also be added in a reverse transcription reaction using suitable primers containing the barcode sequence. Barcodes can also be added by incorporation into oligonucleotides used to link the genes of interest together, through overlap-extension tails or other methods, such that they are located between the two genes of interest. Accordingly, using these methods, barcodes can be incorporated onto the ends of physically linked polynucleotide sequences or into the linker sequences joining the two polynucleotide sequences of interest.

[00269] In one embodiment, the linkage is accomplished through the use of overlap-extension (see Figure 16). In general, overlap extension tails are complementary sequences that are added to polynucleotide sequences of interest to be joined. Annealing of overlap-extension tails appended to the polynucleotide molecules of interest allow them to be linked (see Figures 17, 18, 19, and 20). As described below, overlap-extension tails can be added through a number of well known methods including, but not limited to, polynucleotide synthesis reactions, such as nucleic acid amplification and reverse transcription, ligation, and recombination. Because of the

variety of methods available to effect linkage, it will be recognized that different relative orientations of the polynucleotide segments can be obtained, with respect to sense and antisense orientations, i.e., the segments, e.g., antibody heavy and light chains, can be joined head to tail, head to head, or tail to tail.

[00270] In one embodiment, overlap-extension tails enable the linkage of polynucleotide sequences generated during a polynucleotide synthesis reaction. For example, overlap extension tails can be introduced during the course of polynucleotide synthesis reactions, such as amplification or reverse transcription reactions, by using primers carrying an overlap-extension tail. Alternatively, ligation reactions can be used. As shown in Figures 17, 18, 19, and 20, after annealing of complementary overlap extension tails, the DNA is filled-in in a 5' to 3' direction during the extension phase of a polynucleotide synthesis reaction, such as PCR, to generate a double stranded polynucleotide with the two polynucleotides of interest physically joined.

[00271] In some embodiments, an overlap-extension RT-PCR method allows the sequences to be linked simultaneously as the reaction proceeds in a single tube, thus eliminating the need for intermediate purification. In some embodiments, an overlap extension tail comprises a barcode sequence.

[00272] Figure 17 illustrates generally one example of the use of overlap-extension tails to join polynucleotide sequences encoding antibody light and heavy chains and to provide at least one barcode. Other methods useful for linking two polynucleotide sequences of interest are discussed below. In this example, after polynucleotide synthesis, e.g., reverse transcription, has occurred, the use of a LC gene specific PCR primer containing a barcode, optional sequencing primer site, and optional restriction site (RE1) allows these elements to be added to the end of the resulting PCR product. Primers specific for LC (in one embodiment the V_L region) and HC (in one embodiment the V_H region) with extension overlaps and encoding an optional restriction site (RE3) are indicated. In one embodiment, the LC comprises the rearranged VJ with or without a short segment of the constant region, and the HC comprises the rearranged V(D)J with or without a short segment of constant region of the heavy chain. In one embodiment, the overlap-extension primers also contain a barcode sequence. A reverse primer specific for HC containing an optional RE2 is also used. As amplification with these primers proceeds, a nucleic acid with the linked structure shown is generated with a barcode at one end. Products from reactions conducted in single samples can be easily integrated into the other work flows disclosed herein. For example, an optional second barcode can be added and used in conjunction with the first barcode to further enable multiplexing to identify large numbers of

sequences using a relative minimum number of barcodes. For sequencing, a single barcode may be sufficient.

[00273] Variations of the general scheme shown in Figure 17, examples of which are illustrated herein, will be apparent. For example, the barcode can be placed at the other end of the final product, or at both ends, or between the polynucleotides. Furthermore, a barcode can be included as part of the extension overlap region (e.g., on either side of RE3 or the barcode can be split by the RE3 sequence).

[00274] LC (including V_L sequences) and HC (including V_H sequences) sequences can be derived through a variety of means. For example, they can be generated through reverse transcription of mRNA with either oligo dT or gene specific primers. In one embodiment, a reverse transcription and subsequent amplification reactions are performed simultaneously, i.e., an RT-PCR reaction, to arrive at the final product. When reverse transcription is used, the extension overlap region, as well as other elements, such as restriction sites, sequencing primer sites, or universal sequences, can be added via the annealing of an adaptor comprising one or more G residues to the one, two, three, or more C residues generated by the 3' tailing of cDNA generated in the reverse transcription reaction as shown, for example, in Figures 18, 19, and 20. Template switching by the reverse transcriptase allows an extension overlap region (and other sequence elements) to be added to the cDNA. For example, as shown in Figure 18, when taking advantage of the 3' tailing and template switching activities of reverse transcriptase, a first adaptor can be used to add an extension overlap sequence and a barcode to a first polynucleotide of interest, while a second adaptor with a sequence complementary to the overlap-extension of the first adaptor can be added to a second cDNA of interest. The complementary extension overlap sequences anneal during a subsequent nucleic acid synthesis reaction, such as PCR, to join the two polynucleotides of interest. Extension from the point of overlap results in a double stranded DNA molecule in which the two polynucleotides of interest are linked with a barcode between them. Variations that allow the generation of two internally located barcodes between two linked polynucleotide sequences are shown in Figures 19 and 20.

[00275] Other methods for joining or linking the polynucleotide sequences of interest include by ligation. In this embodiment, the primer mix used for the amplification is designed such that the amplified target sequences can be cleaved with appropriate restriction enzymes, and covalent linkage by DNA ligation can be performed. Following amplification with such a primer mix, the restriction enzymes needed to form compatible ends of the target sequences, are added to the mixture. Target sequences are then ligated together with a ligase. No purification of the PCR

products is needed prior to either the restriction enzyme digest or ligation steps, although purification may be performed.

[00276] In another embodiment, the polynucleotide sequences of interest can be linked by recombination. In this approach, the amplified polynucleotide sequences of interest can be joined using identical recombination sites. Linkage is performed by adding the appropriate recombinase to facilitate recombination. Suitable recombinase systems include Flp recombinase with a variety of FRT sites, Cre recombinase with a variety of lox sites, integrase ΦC31 which carries out recombination between the attP site and the attB site, the β-recombinase-six system as well as the Gin-gix system. Linkage by recombination has been exemplified for two nucleotide sequences (V_H linked with V_L) (Chapal, N. et al. 1997 BioTechniques 23, 518-524), hereby incorporated by reference.

[00277] Accordingly, in one aspect, the method comprises amplifying by PCR or RT-PCR amplification, nucleotide sequences of interest using a template derived from an isolated single cell or a population of isogenic cells and (1) effecting linkage of the amplified nucleotide sequences of interest and (2) adding one more barcodes to the linked polynucleotide sequences. The method comprises an optional step of performing an additional amplification of the linked products to, for example, add additional barcodes, restriction sites, sequencing primer sites, and the like.

[00278] In another aspect, a method of producing a library of barcoded linked pairs comprising antibody heavy and light chains from single cells from a donor is provided. This aspect comprises providing a lymphocyte-containing cell fraction from a donor, which is optionally enriched for a particular lymphocyte population from said cell fraction. Further, a population of isolated single cells is obtained by distributing cells from the lymphocyte-containing cell fraction, or the enriched cell fraction, individually among a plurality of vessels, containers, or wells. Multiplex molecular amplification (e.g., multiplex RT-PCR amplification) of the variable region encoding sequences contained in the population of isolated single cells is performed and linkage of pairs of heavy and light chains and barcode addition is affected, wherein an individual pair is derived from a single cell. Further, in different embodiments, the method can comprise two optional steps: in the first step, the individual isolated single cell in the population of single cells can be expanded to a population of isogenic cells prior to performing multiplex RT-PCR amplification, thereby providing a plurality of vessels, containers, or wells with a population of isogenic cells (one population of isogenic cells in one vessel, container, or well). Another optional step encompasses performing an additional amplification of the linked light and heavy chain encoding sequences. This additional amplification step can be used to

simply increase the amount of the linked nucleic acid, or to add a first or second barcode sequence or other sequence elements to the linked nucleic acid.

[00279] In some aspects, the multiplex RT-PCR amplification can be performed either as a two-step process, where reverse transcription (RT) is performed separate from the multiplex PCR amplification (or alternative multiplex molecular amplification), or as a single-step process, where the RT and multiplex PCR amplification steps are performed with the same primers in one tube.

[00280] The reverse transcription (RT) is performed with an enzyme containing reverse transcriptase activity resulting in the generation of cDNA from total RNA, mRNA or target specific RNA from an isolated single cell. Primers which can be utilized for the reverse transcription include oligo-dT primers, random hexamers, random decamers, other random primers, or primers that are specific for the nucleotide sequences of interest. In some embodiments, such primers can contain elements such as barcodes, universal priming sites, restriction sites, sequencing primer sites, and the like.

[00281] The two-step multiplex RT-PCR amplification procedure allows for the cDNA generated in the RT step to be distributed to more than one vessel allowing for the storage of a template fraction before proceeding with the amplification, if desired. Additionally, the distribution of cDNA to more than one tube, allows for the performance of more than one multiplex PCR amplification of nucleic acid derived from the same template. This two-step approach can for example be used to amplify and link heavy chain variable region and kappa light chain variable region encoding sequences in one tube and heavy chain variable region and lambda light chain variable region encoding sequences in a different tube utilizing the same template. A single cell usually only expresses one of the light chains. However, it will often be easier to perform the reactions simultaneously instead of awaiting the result of one of the reactions before performing the other. Further, the amplification of both kappa and lambda serves as an internal negative control, since it would be expected that only kappa or lambda would amplify from a single cell.

[00282] In the single-step multiplex RT-PCR procedure, reverse transcription and multiplex PCR amplification is carried out in the same vessel, container, or well. All the components necessary to perform both the reverse transcription and the multiplex PCR in a single step are initially added into the vessels, containers, or wells and the reaction is performed. Generally, there is no need to add additional components once the reaction has been started. The advantage of single-step multiplex RT-PCR amplification is that it reduces the number of steps necessary to generate the barcode linked nucleotide sequences of the present invention even further. This is

particularly useful when performing multiplex RT-PCR on an array of single cells, where the same reaction is carried out in a plurality of vessels. Generally, the composition needed for the single-step multiplex RT-PCR comprises a nucleic acid template, an enzyme with reverse transcriptase activity, an enzyme with DNA polymerase activity, deoxynucleoside triphosphate mix (dNTP mix comprising dATP, dCTP, dGTP and dTTP) and a multiplex primer mix. The nucleic acid template is preferably total RNA or mRNA derived from an isolated single cell either in a purified form, as a lysate of the cell, or as contained in the intact cell.

[00283] In one aspect, the methods generate libraries of linked and barcoded polynucleotides of interest. In some aspects, the plurality of polynucleotide compositions in a polynucleotide library can comprise at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 3000, at least 10,000, at least 30,000, at least 100,000, at least 300, 000, at least 1,000,000, at least 3,000,000, at least 10,000,000, at least 30,000,000, or more members. In other aspects, the plurality of polynucleotide compositions in a polynucleotide library can comprise at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 3000, at least 10,000, at least 30,000, or more genes of a cell sample's whole transcriptome. In other aspects, the plurality of polynucleotide compositions in a polynucleotide library comprises at least 1, at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 1,000,000,000 or more of the different antibody species present in the blood of an individual. These the antibody species can be expressed by plasmablasts, plasma cells, memory B cells, long-lived plasma cells, naïve B cells, other B lineage cells, or combinations thereof.

[00284] The linked and barcoded polynucleotide compositions generated by the methods disclosed above can advantageously be subjected to high throughout, multiplexed sequencing, preferably, using NextGen sequencing platforms as described herein.

[00285] The linked and barcoded polynucleotide compositions generated by the methods disclosed above can also used for cloning, producing polypeptides of interest, and screening as disclosed herein.

Methods of producing one or more polynucleotides of interest for sequencing

[00286] In some aspects, the method includes obtaining a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a universal primer region, a sample identification region, an adapter region, and an amplicon region derived from a single sample, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample

identification region of each polynucleotide derived from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library derived from one or more samples distinct from the first single sample; and amplifying the polynucleotide library with a set of primers to produce the one or more polynucleotides of interest for sequencing, wherein the one or more polynucleotides of interest for sequencing comprises a first sequencing region, a first plate identification region, a universal primer region, a sample identification region, an adapter region, an amplicon region derived from a single sample, and a second sequencing region.

[00287] In some aspects, a method further includes sequencing the one or more polynucleotides of interest. In some aspects, the sequencing is 454 sequencing.

[00288] In some aspects, sequencing includes longer sequencing reads such that the forward and reverse sequencing reads overlap enough to enable reconstruction of the entire, approximately 600 base pair (bp) sequence of, e.g., antibody light chains (LCs) (where exact sequence length can depend on the length of the 5' untranslated region (UTR)), and approximately 700 bp sequence of the heavy chains (HCs). Therefore, in some aspects, any sequencing technology that can yield sequencing reads of at least 350-400bp and thereby achieve the overlap included for sequence assembly can be used, and sequencing technologies that enable 600-700+ bp reads would allow one to sequence using just a forward primer (sequencing from the 5' end).

[00289] Any technique for sequencing nucleic acid known to those skilled in the art can be used. DNA sequencing techniques include classic dideoxy sequencing reactions (Sanger method) using labeled terminators or primers and gel separation in slab or capillary electrophoresis. In a preferred embodiment, next generation (NextGen) sequencing platforms are advantageously used in the practice of the invention. NextGen sequencing refers to any of a number of post-classic Sanger type sequencing methods which are capable of high throughput, multiplex sequencing of large numbers of samples simultaneously. Current NextGen sequencing platforms, such as those described in greater detail below, are capable of generating reads from multiple distinct nucleic acids in the same sequencing run. Throughput is varied, with 100 million bases to 600 giga bases per run, and throughput is rapidly increasing due to improvements in technology. The principle of operation of different NextGen sequencing platforms is also varied and can include: sequencing by synthesis using reversibly terminated labeled nucleotides, pyrosequencing, 454 sequencing, allele specific hybridization to a library of labeled oligonucleotide probes, sequencing by synthesis using allele specific hybridization to a library of labeled clones that is followed by ligation, real time monitoring of the incorporation of

labeled nucleotides during a polymerization step, polony sequencing, single molecule real time sequencing, and SOLiD sequencing. Sequencing has been demonstrated by sequential or single extension reactions using polymerases or ligases as well as by single or sequential differential hybridizations with libraries of probes. These reactions have been performed on many clonal sequences in parallel including demonstrations in current commercial applications of over 100 million sequences in parallel. These sequencing approaches can thus be used to study the repertoire of T-cell receptor (TCR) and/or B-cell receptor (BCR) and other sequences of interest.

[00290] The sequencing techniques can generate at least 1000 reads per run, at least 10,000 reads per run, at least 100,000 reads per run, at least 500,000 reads per run, or at least 1,000,000 reads per run.

[00291] The sequencing techniques can generate about 30 bp, about 40 bp, about 50 bp, about 60 bp, about 70 bp, about 80 bp, about 90 bp, about 100 bp, about 110, about 120 bp per read, about 150 bp, about 200 bp, about 250 bp, about 300 bp, about 350 bp, about 400 bp, about 450 bp, about 500 bp, about 550 bp, about 600 bp, about 650 bp, or about 700 bp or more bp per read.

[00292] The sequencing techniques can generate at least 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700 or more nucleotides per read.

[00293] A sequencing technique that can be used, for example, Helicos True Single Molecule Sequencing (tSMS) (Harris T. D. et al. (2008) Science 320:106-109). In the tSMS technique, a DNA sample is cleaved into strands of approximately 100 to 200 nucleotides, and a polyA sequence is added to the 3' end of each DNA strand. Each strand is labeled by the addition of a fluorescently labeled adenosine nucleotide. The DNA strands are then hybridized to a flow cell, which contains millions of oligo-T capture sites that are immobilized to the flow cell surface. The templates can be at a density of about 100 million templates/cm^2. The flow cell is then loaded into an instrument, e.g., HeliScope^{TM} sequencer, and a laser illuminates the surface of the flow cell, revealing the position of each template. A CCD camera can map the position of the templates on the flow cell surface. The template fluorescent label is then cleaved and washed away. The sequencing reaction begins by introducing a DNA polymerase and a fluorescently labeled nucleotide. The oligo-T nucleic acid serves as a primer. The polymerase incorporates the labeled nucleotides to the primer in a template directed manner. The polymerase and unincorporated nucleotides are removed. The templates that have directed incorporation of the fluorescently labeled nucleotide are detected by imaging the flow cell surface. After imaging, a cleavage step removes the fluorescent label, and the process is repeated with other fluorescently

labeled nucleotides until the desired read length is achieved. Sequence information is collected with each nucleotide addition step.

[00294] Another example of a DNA sequencing technique that can be used is 454 sequencing (Roche) (Margulies, M et al. 2005, Nature, 437, 376-380). 454 sequencing involves two steps. In the first step, DNA is sheared into fragments of approximately 300-800 base pairs, and the fragments are blunt ended. Oligonucleotide adaptors are then ligated to the ends of the fragments. The adaptors serve as primers for amplification and sequencing of the fragments. The fragments can be attached to DNA capture beads, e.g., streptavidin-coated beads using, e.g., Adaptor B, which contains 5'-biotin tag. The fragments attached to the beads are PCR amplified within droplets of an oil-water emulsion. The result is multiple copies of clonally amplified DNA fragments on each bead. In the second step, the beads are captured in wells (pico-liter sized). Pyrosequencing is performed on each DNA fragment in parallel. Addition of one or more nucleotides generates a light signal that is recorded by a CCD camera in a sequencing instrument. The signal strength is proportional to the number of nucleotides incorporated.

[00295] Pyrosequencing makes use of pyrophosphate (PPi) which is released upon nucleotide addition. PPi is converted to ATP by ATP sulfurylase in the presence of adenosine 5' phosphosulfate. Luciferase uses ATP to convert luciferin to oxyluciferin, and this reaction generates light that is detected and analyzed.

[00296] Another example of a DNA sequencing technique that can be used is SOLiD technology (Applied Biosystems). In SOLiD sequencing, genomic DNA is sheared into fragments, and adaptors are attached to the 5' and 3' ends of the fragments to generate a fragment library. Alternatively, internal adaptors can be introduced by ligating adaptors to the 5' and 3' ends of the fragments, circularizing the fragments, digesting the circularized fragment to generate an internal adaptor, and attaching adaptors to the 5' and 3' ends of the resulting fragments to generate a mate-paired library. Next, clonal bead populations are prepared in microreactors containing beads, primers, template, and PCR components. Following PCR, the templates are denatured and beads are enriched to separate the beads with extended templates. Templates on the selected beads are subjected to a 3' modification that permits bonding to a glass slide.

[00297] The sequence can be determined by sequential hybridization and ligation of partially random oligonucleotides with a central determined base (or pair of bases) that is identified by a specific fluorophore. After a color is recorded, the ligated oligonucleotide is cleaved and removed and the process is then repeated.

[00298] Another example of a sequencing technology that can be used is SOLEXA sequencing (Illumina). SOLEXA sequencing is based on the amplification of DNA on a solid surface using fold-back PCR and anchored primers. Genomic DNA is fragmented, and adapters are added to the 5' and 3' ends of the fragments. DNA fragments that are attached to the surface of flow cell channels are extended and bridge amplified. The fragments become double stranded, and the double stranded molecules are denatured. Multiple cycles of the solid-phase amplification followed by denaturation can create several million clusters of approximately 1,000 copies of single-stranded DNA molecules of the same template in each channel of the flow cell. Primers, DNA polymerase and four fluorophore-labeled, reversibly terminating nucleotides are used to perform sequential sequencing. After nucleotide incorporation, a laser is used to excite the fluorophores, and an image is captured and the identity of the first base is recorded. The 3' terminators and fluorophores from each incorporated base are removed and the incorporation, detection and identification steps are repeated.

[00299] Another example of a sequencing technology that can be used includes the single molecule, real-time (SMRT™) technology of Pacific Biosciences. In SMRT, each of the four DNA bases is attached to one of four different fluorescent dyes. These dyes are phospholinked. A single DNA polymerase is immobilized with a single molecule of template single stranded DNA at the bottom of a zero-mode waveguide (ZMW). A ZMW is a confinement structure which enables observation of incorporation of a single nucleotide by DNA polymerase against the background of fluorescent nucleotides that rapidly diffuse in an out of the ZMW (in microseconds). It takes several milliseconds to incorporate a nucleotide into a growing strand. During this time, the fluorescent label is excited and produces a fluorescent signal, and the fluorescent tag is cleaved off. Detection of the corresponding fluorescence of the dye indicates which base was incorporated. The process is repeated.

[00300] Another example of a sequencing technique that can be used is nanopore sequencing (Soni G V and Meller A. (2007) Clin Chem 53: 1996-2001). A nanopore is a small hole, of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it results in a slight electrical current due to conduction of ions through the nanopore. The amount of current which flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore represents a reading of the DNA sequence.

[00301] Another example of a sequencing technique that can be used involves using a chemical-sensitive field effect transistor (chemFET) array to sequence DNA (for example, as

described in US Patent Application Publication No. 20090026082). In one example of the technique, DNA molecules can be placed into reaction chambers, and the template molecules can be hybridized to a sequencing primer bound to a polymerase. Incorporation of one or more triphosphates into a new nucleic acid strand at the 3' end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors. In another example, single nucleic acids can be attached to beads, and the nucleic acids can be amplified on the bead, and the individual beads can be transferred to individual reaction chambers on a chemFET array, with each chamber having a chemFET sensor, and the nucleic acids can be sequenced.

[00302] Another example of a sequencing technique that can be used involves using an electron microscope (Moudrianakis E. N. and Beer M. Proc Natl Acad Sci USA. 1965 March; 53:564-71). In one example of the technique, individual DNA molecules are labeled using metallic labels that are distinguishable using an electron microscope. These molecules are then stretched on a flat surface and imaged using an electron microscope to measure sequences.

[00303] In some aspects, obtaining the polynucleotide library comprises preparing the polynucleotide library in a laboratory. In some aspects, obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has prepared the polynucleotide library.

Methods for analyzing sequencing data

[00304] In some aspects, the method includes obtaining a dataset associated with a plurality of polynucleotides, wherein the dataset comprises sequencing data for the plurality of polynucleotides, wherein each polynucleotide in the plurality of polynucleotides comprises a sample identification region, and wherein each sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the sample identification region of each polynucleotide derived from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the plurality of polynucleotides derived from one or more samples distinct from the first single sample; and analyzing the dataset to match together polynucleotides with identical sample identification regions, wherein a match indicates that the polynucleotides originated from the same sample.

[00305] In some aspects each polynucleotide in the plurality of polynucleotides further comprises a first plate identification region, wherein each combination of each first plate identification region and sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the first plate identification region of each

polynucleotide derived from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the plurality of polynucleotides derived from one or more single sample sets distinct from the first set of single samples, and further comprising analyzing the dataset to match together polynucleotides with identical first plate identification regions and identical sample identification regions, wherein a match between both regions indicates that the polynucleotides originated from the same sample.

[00306] In some aspects, both polynucleotides include a variable region. In some aspects, one polynucleotide includes a variable region. In some aspects, neither polynucleotide includes a variable region.

[00307] In some aspects, obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of polynucleotides to experimentally determine the dataset. In some aspects, obtaining the dataset comprises receiving the dataset directly or indirectly from a third party that has sequenced the plurality of polynucleotides to experimentally determine the dataset. In some aspects, the dataset is stored on an electronic storage medium. In some aspects, the dataset is transferred over the Internet.

[00308] In some aspects, the method is implemented on a computer, e.g., it is a computer-implemented method.

[00309] In some aspects, the single sample is a single cell. In some aspects, the single sample comprises a single cell. In some aspects, the single sample comprises a single B cell. In some aspects, the single sample comprises a plurality of B cells. In some aspects, the single sample comprises a single B cell and one or more other cells.

[00310] In some aspects, data generated from sequencing (e.g., 454 sequencing) can be analyzed by 454 GS FLX data analysis software, and sequences with poor-quality scores can be filtered out. Good-quality sequences can then be subdivided according to their sample identification region (and in some embodiments the combination of their sample identification region and plate identification region) by using a script in Python before sequences are assembled using bioinformatics approaches, for example, by using Newbler. Because reverse reads can have only a second plate identification region in some aspects, it is possible that sequence assembly could occur between forward and reverse reads of sequences from different cells. For circumventing this potential problem, the heavy- and light-chain V(D)J usage of both forward and reverse reads can first be identified using HighV-QUEST. Sequences can then be further grouped according to their V(D)J usage before being assembled. In some aspects, sequence assembly can be intolerant of nucleotide mismatches, thereby preventing assembly of forward and reverse reads from different cells that share the same V(D)J usage. In some aspects,

sequences can then be clustered together based on their V(D)J usage by using a computer program.

[00311] In some aspects, bioinformatics methods may be used to identify groups of sequences forming clonal families and subfamilies, and thereby immunoglobulin sequences of interest. Such bioinformatics methods involve measurements of sequence similarity. Such bioinformatics methods may be used to identify sequences of interest derived from an individual human, derived from one or more humans, derived from one or more humans with a condition, or derived from one or more humans with different conditions.

[00312] In some aspects, related immunoglobulin heavy and/or light chain sequences can be identified through computational phylogenetic analysis of the homology between the immunoglobulin heavy chain and/or light chain V(D)J sequences. In some aspects, standard classification methods (i.e. clustering) of the sequences representing the individual or combinations of the V, D, and/or J gene segments and/or other sequences derived from the immunoglobulin heavy chain and/or light chain can be used to identify clonal families or subfamilies (for example, by using ClustalX).

[00313] As used herein "clonal family" refers to a plurality of immunoglobulin sequences each having V, D, and/or J regions, wherein each sequence is a mutated version of the same germline immunoglobulin sequence having a V, D, and/or J region or the germline immunoglobulin sequence having the V, D, and/or J region. In some aspects, the plurality is a plurality of heavy chain sequences. In some aspects, the plurality is a plurality of light chain sequences. In some aspects, the plurality is a plurality of paired heavy and light chain sequences. In some aspects, each sequence has V, D, and J regions. In some aspects, each sequence has V and D regions. In some aspects, each sequence has D and J regions. In some aspects, each sequence has V and J regions. In some aspects, each sequence has a V region. In some aspects, each sequence has a D region. In some aspects, each sequence has a J region. In some aspects, the one or more mutations are located within the V, D, and/or J regions. In some aspects, the one or more mutations are located between the V, D, and/or J regions.

[00314] In some aspects, a set of antibodies whose heavy chains all use the same V and J gene segments are a clonal family. In some aspects, a set of antibodies whose heavy chains all use the same V and J gene segments and whose sum of the length of P/N nucleotides and D nucleotides are of the same length are a clonal family. In some aspects, a set of antibodies whose heavy chains all use the same V, D and J gene segments are a clonal family. In some aspects, a set of antibodies whose heavy chains all use the same V, D and J gene segments, and whose P/N nucleotides between the V and D gene segments are the same length, and whose P/N

nucleotides between the D and J gene segments are the same length, are a clonal family. In some aspects, a set of antibodies whose heavy chains all use the same V and J gene segments and whose light chains all use the same V and J gene segments are a clonal family. In some aspects, a set of antibodies whose heavy chains all use the same V and J gene segments and whose sum of the length of P/N nucleotides and D nucleotides are of the same length, and whose light chains all use the same V and J gene segments and whose P/N nucleotides are of the same length, are a clonal family. In some aspects, a set of antibodies whose heavy chains all use the same V, D and J gene segments and whose light chains all use the same V and J gene segments are a clonal family. In some aspects, a set of antibodies whose heavy chains all use the same V, D and J gene segments, and whose P/N nucleotides between the V and D gene segments are the same length, and whose P/N nucleotides between the D and J gene segments are the same length, and whose light chains all use the same V and J gene segments, and whose P/N nucleotides between the V and J gene segments are the same length, are a clonal family.

[00315] Methods for constructing clonal families

[00316] The V, D and J usage for a T cell receptor (TCR) or an immunoglobulin variable gene query sequence can be determined by identifying the germline V, D (if applicable) and J gene segments most likely to have given rise to the sequence. D segments are present in some TCR and immunoglobulin sequences (e.g. TCRβ, TCRδ and antibody heavy chain sequences) but not others (e.g TCRα, TCRγ and antibody light chain sequences). The following description includes D segments but the same approaches can be applied to variable region sequences that lack D segments. In all cases the determination of V(D)J usage uses a reference database of germline V, D and J gene segment sequences such as IMGT/GENE-DB (Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D256-61.).

[00317] In one approach to determination of V(D)J usage, the query sequence is compared serially to each V, D and J germline gene segment separately and the most similar gene segment of each type (V, D or J) is selected as the most likely to have given rise to the query sequence. V-QUEST and High V-QUEST are examples of this approach (Giudicelli V, Chaume D, Lefranc MP. IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W435-40.; Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W503-8.). V-QUEST first generates

pairwise alignments for the query sequence and each V gene segment sequence. Then it generates pairwise alignments for the query sequence region downstream of the deduced 3' end of the V segment and each J gene segment sequence. If a D segment is present, V-QUEST then generates pairwise alignments for the query sequence region found between the regions matching V and J segments and each D gene segment sequence. V-QUEST can also infer the boundaries of the V-D, V-J and/or D-J junction regions.

[00318] In another approach to determination of V(D)J usage, the combination of germline V, D and J segments most likely to have given rise to the query sequence is identified in a single step rather than in three separate steps for V, D and J respectively. This approach has the advantage that the identification of one type of segment (V, D or J) can take into account information about the potential matches to the other two types of segments. For example, the best matching D segment might depend upon which V segment match is being considered. SoDA is an example of this approach (Volpe JM, Cowell LG, Kepler TB. SoDA: implementation of a 3D alignment algorithm for inference of antigen receptor recombinations. Bioinformatics. 2006 Feb 15;22(4):438-44.). SoDA first selects candidate V, D and J segments. It generates pairwise local alignments for the query sequence and each V gene segment sequence and then keeps only the V segments with alignments meeting a score threshold. It repeats these steps for the J segments and D segments. Then an optimal alignment is generated for each possible combination of candidate V, D and J segments. The alignments are generated using the same general dynamic programming approach widely used in sequence alignment (Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol., 48,443–453.), but allowing for the insertion of additional nucleotides at the V-D, V-J and/or D-J junctions. Such insertion commonly takes place during the biological process of V(D)J recombination. In sequence alignment by dynamic programming there are typically penalty scores associated with insertions, deletions and mismatches. However, in this approach to determining V(D)J usage, no penalties are applied for the insertion of nucleotides at the junctions between segments. TheV(D)J combination yielding the highest-scoring alignment is selected to indicate the V(D)J usage for the query sequence. This approach can also identify the boundaries of junction sequence regions.

[00319] From the clonal families and subfamilies, a variety of approaches can be used to select specific clones for expression of their encoded paired heavy and light chain immunoglobulin genes and characterization of their binding properties. In some aspects, the highest frequency clones from clonal families and/or clonal subfamilies as well as other

representative clones from clonal families and subfamilies are expressed and screened for their binding properties. Clones may also be randomly selected, from all clones, from all or select clonal families, and/or from all or select clonal subfamilies for expression and characterization of their binding characteristics. Clones may also be selected based on possessing larger numbers of variations in the variable region of the antibody. A phylogenetic tree may be constructed, and clones may be selected based on features of the phylogenetic tree, for example by descending the tree always choosing the branch with the largest number of leaf nodes underneath.

[00320] In some aspects, a method further includes selecting one or more polynucleotides for cloning.

Methods for identifying a second polynucleotide of interest based on selection of a first polynucleotide of interest

[00321] In some aspects, the method includes obtaining a dataset associated with a plurality of polynucleotides, wherein the dataset comprises sequencing data for the plurality of polynucleotides, wherein each polynucleotide in the plurality of polynucleotides comprises a sample identification region, and wherein each sample identification region on each polynucleotide is unique to a single sample thereby associating each polynucleotide in the plurality of polynucleotides with a distinct single sample, wherein the sequence of the sample identification region of each polynucleotide derived from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the plurality of polynucleotides derived from one or more samples distinct from the first single sample; and selecting a first polynucleotide of interest associated with a first single sample from the dataset and identifying a second polynucleotide of interest in the first single sample based on the sample identification region of the first polynucleotide of interest.

[00322] In some aspects, each polynucleotide in the plurality of polynucleotides further comprises a first plate identification region, wherein each combination of each first plate identification region and sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the first plate identification region of each polynucleotide derived from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the plurality of polynucleotides derived from one or more single sample sets distinct from the first set of single samples, and further comprising identifying a second polynucleotide of interest in the first single sample based on the sample identification region and first plate identification region of the first polynucleotide of interest.

[00323] In some aspects, both polynucleotides include a variable region. In some aspects, one polynucleotide includes a variable region. In some aspects, neither polynucleotide includes a variable region.

[00324] In some aspects, the method is implemented on a computer, e.g., it is a computer-implemented method.

[00325] In some aspects, the first single sample comprises a B cell. In some aspects, the first single sample comprises a single B cell and one or more other cells. In some aspects, the first single sample comprises a plurality of B cells. In some aspects, the first single sample comprises a B cell, wherein the first polynucleotide of interest comprises an antibody heavy chain nucleotide sequence, and wherein the second polynucleotide of interest comprises an antibody light chain nucleotide sequence. In some aspects, the first single sample comprises a B cell, wherein the first polynucleotide of interest comprises an antibody light chain nucleotide sequence, and wherein the second polynucleotide of interest comprises an antibody heavy chain nucleotide sequence.

[00326] In some aspects, obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of polynucleotides to experimentally determine the dataset. In some aspects, obtaining the dataset comprises receiving the dataset directly or indirectly from a third party that has sequenced the plurality of polynucleotides to experimentally determine the dataset. In some aspects, the dataset is stored on an electronic storage medium.

Methods of producing one or more polynucleotides of interest for cloning

[00327] In some aspects, the method includes obtaining a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a universal primer region, a sample identification region, an adapter region, and an amplicon region derived from a single sample, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide derived from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library derived from one or more samples distinct from the first single sample; and amplifying the polynucleotide library with a set of primers to produce the one or more polynucleotides of interest for cloning, wherein the one or more polynucleotides of interest for cloning comprises a first restriction site region, a universal primer region, a sample identification region, an adapter region, an amplicon region derived from a single sample, and a second restriction site region.

[00328] In some aspects, obtaining the polynucleotide library comprises preparing the polynucleotide library in a laboratory. In some aspects, obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has prepared the polynucleotide library.

[00329] In some aspects, a methods further include cloning one or more polynucleotides, e.g., into a vector disclosed herein.

Methods of producing a molecule of interest

[00330] In some aspects, the method includes obtaining a host cell comprising a polynucleotide of interest; and culturing the host cell under conditions sufficient to produce the molecule of interest.

[00331] In some aspects, obtaining the host cell comprises preparing the host cell comprising the polynucleotide in a laboratory. In some aspects, obtaining the host cell comprises receiving the host cell comprising the polynucleotide directly or indirectly from a third party that has prepared the host cell.

[00332] In some aspects, the molecule of interest is a polypeptide. In some aspects, the molecule of interest is an antibody. In some aspects, the molecule of interest is a human monoclonal antibody.

[00333] In some aspects, the method further includes collecting the molecule of interest.

[00334] In some aspects, it is desirable to "refold" certain polypeptides, e.g., polypeptides comprising one or more ABP components or the ABP itself. In certain embodiments, such polypeptides are produced using expression systems discussed herein. In certain embodiments, polypeptides are "refolded" and/or oxidized to form desired tertiary structure and/or to generate disulfide linkages. In certain embodiments, such structure and/or linkages are related to certain biological activity of a polypeptide. In certain embodiments, refolding is accomplished using any of a number of procedures known in the art. Exemplary methods include, but are not limited to, exposing the solubilized polypeptide agent to a pH typically above 7 in the presence of a chaotropic agent. An exemplary chaotropic agent is guanidine. In certain embodiments, the refolding/oxidation solution also contains a reducing agent and the oxidized form of that reducing agent. In certain embodiments, the reducing agent and its oxidized form are present in a ratio that will generate a particular redox potential that allows disulfide shuffling to occur. In certain embodiments, such shuffling allows the formation of cysteine bridges. Exemplary redox couples include, but are not limited to, cysteine/cystamine, glutathione/dithiobisGSH, cupric chloride, dithiothreitol DTT/dithiane DTT, and 2-mercaptoethanol (bME)/dithio-bME. In certain

embodiments, a co-solvent is used to increase the efficiency of refolding. Exemplary cosolvents include, but are not limited to, glycerol, polyethylene glycol of various molecular weights, and arginine.

[00335] In certain embodiments, one substantially purifies a polypeptide, e.g., a polypeptide comprising one or more ABP components or the ABP itself. Certain protein purification techniques are known to those of skill in the art. In certain embodiments, protein purification involves crude fractionation of polypeptide fractionations from non-polypeptide fractions. In certain embodiments, polypeptides are purified using chromatographic and/or electrophoretic techniques. Exemplary purification methods include, but are not limited to, precipitation with ammonium sulphate; precipitation with PEG; immunoprecipitation; heat denaturation followed by centrifugation; chromatography, including, but not limited to, affinity chromatography (e.g., Protein-A-Sepharose), ion exchange chromatography, exclusion chromatography, and reverse phase chromatography; gel filtration; hydroxyapatite chromatography; isoelectric focusing; polyacrylamide gel electrophoresis; and combinations of such and other techniques. In certain embodiments, a polypeptide is purified by fast protein liquid chromatography or by high pressure liquid chromotography (HPLC). In certain embodiments, purification steps can be changed or certain steps can be omitted and still result in a suitable method for the preparation of a substantially purified polypeptide.

[00336] In certain embodiments, one quantitates the degree of purification of a polypeptide preparation. Certain methods for quantifying the degree of purification are known to those of skill in the art. Certain exemplary methods include, but are not limited to, determining the specific binding activity of the preparation and assessing the amount of a polypeptide within a preparation by SDS/PAGE analysis. Certain exemplary methods for assessing the amount of purification of a polypeptide preparation comprise calculating the binding activity of a preparation and comparing it to the binding activity of an initial extract. In certain embodiments, the results of such a calculation are expressed as "fold purification." The units used to represent the amount of binding activity depend upon the particular assay performed.

[00337] In certain embodiments, a polypeptide comprising one or more ABP components or the ABP itself is partially purified. In certain embodiments, partial purification can be accomplished by using fewer purification steps or by utilizing different forms of the same general purification scheme. For example, in certain embodiments, cation-exchange column chromatography performed utilizing an HPLC apparatus will generally result in a greater "fold purification" than the same technique utilizing a low-pressure chromatography system. In certain

embodiments, methods resulting in a lower degree of purification can have advantages in total recovery of polypeptide, or in maintaining binding activity of a polypeptide.

[00338] In certain instances, the electrophoretic migration of a polypeptide can vary, sometimes significantly, with different conditions of SDS/PAGE. See, e.g., Capaldi et al, Biochem. Biophys. Res. Comm., 76: 425 (1977). It will be appreciated that under different electrophoresis conditions, the apparent molecular weights of purified or partially purified polypeptide can be different.

Methods of screening

[00339] In some aspects, a molecule of interest is screened for activity. In some aspects, the molecule of interest is an ABP. In some aspects, the molecule of interest is an antibody.

[00340] In some aspects, methods of screening the libraries disclosed herein are used to identify ABPs capable of binding to a desired target. Any *in vitro* or *in vivo* screening method that allows for selection of an ABP from a library, based upon the ABP binding to a target molecule, is contemplated.

[00341] In one embodiment, a library may be screened using an art recognized in vitro cell-free phenotype-genotype linked display. Such methods are well known in the art and are described, for example, in U.S. Pat. Nos. 7,195,880; 6,951,725; 7,078,197; 7,022,479; 6,518,018; 7,125,669; 6,846,655; 6,281,344; 6,207,446; 6,214,553; 6,258,558; 6,261,804; 6,429,300; 6,489,116; 6,436,665; 6,537,749; 6,602,685; 6,623,926; 6,416,950; 6,660,473; 6,312,927; 5,922,545; and 6,348,315. These methods involve transcription of protein in vitro from a nucleic acid in such a way that the protein is physically associated or bound to the nucleic acid from which it originated. By selecting for an expressed protein with a target molecule, the nucleic acid that codes for the protein may also be selected.

[00342] To improve the expression of scFv proteins, the above referenced in vitro screening assays may include the addition or removal of certain reagents. In one embodiment, protein disulphide isomerase enzymes may be added to the in vitro expression system to improve the production of functional scFv molecules. In another embodiment, a mild oxidizing agent (for example, GSSG (oxidized glutathione)/GSH (reduced glutathione), for example 100 mM GSSG/10 mM GSH) may be added to in vitro translation reaction mixture of the scFv proteins to allow intra-chain disulphide bond formation in the VH and VL regions of the scFv molecule. In another embodiment, reducing agents (for example, dithiothreitol (DTT)) may be removed from the in vitro translation reaction mixture of the scFv.

[00343] In another embodiment, one or more labeled amino acids, or derivatives thereof, may be added to the in vitro translation system such that the labeled amino acid(s) becomes incorporated into the resultant antibody. Any art recognized labeled amino acid is contemplated, for example, a radiolabelled amino acid, for example, ^{35}S-labelled methionine or cysteine.

[00344] In one embodiment, the in vitro screening assays may include that after in vitro selection of an antibody or plurality of antibodies the mRNA that is physically associated with the antibody or plurality of antibodies may be reverse transcribed to generate cDNA encoding said antibody or plurality of antibodies. Any suitable method for reverse transcription is contemplated, for example, enzyme mediated, for example, Moloney murine leukemia virus reverse transcriptase.

[00345] The screening methods may include amplification of the nucleic acid that encodes antibodies that bind specifically to a desired target. In one embodiment, mRNA that is physically associated with an antibody or plurality of antibodies may be amplified to produce more mRNA. Any art recognized method of RNA replication is contemplated, for example, using an RNA replicase enzyme. In another embodiment, mRNA that is physically associated with an antibody or plurality of antibodies is first reverse transcribed into cDNA before being amplified by PCR. In one embodiment, PCR amplification is accomplished using a high fidelity, proof--reading polymerase, for example, the KOD1 thermostable DNA polymerase from Thermococcus kodakaraensis or Platinum Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, Calif.). In another embodiment, PCR amplification may be performed under conditions that result in the introduction of mutations into amplified DNA, i.e., error-prone PCR.

[00346] Screening methods may also include that the stringency of the target-binding screening assay be increased to select for antibodies with improved affinity for target. Any art recognized methods of increasing the stringency of an antibody-target interaction assay are contemplated. In one embodiment, one or more of the assay conditions may be varied (for example, the salt concentration of the assay buffer) to reduce the affinity of the antibody molecules for the desired target. In another embodiment, the length of time permitted for the antibodies to bind to the desired target may be reduced. In another embodiment, a competitive binding step may be added to the antibody-target interaction assay. For example, the antibodies may first be allowed to bind to a desired immobilized target. A specific concentration of non-immobilized target may then be added, which serves to compete for binding with the immobilized target such that antibodies with the lowest affinity for antigen are eluted from the immobilized target, resulting in enrichment for antibodies with improved antigen binding

affinity. In an embodiment, the stringency of the assay conditions may further be increased by increasing the concentration of non-immobilized target that is added to the assay.

[00347] Screening methods may also include multiple rounds of selection to enrich for one or more antibodies with improved target binding. In one embodiment, at each round of selection further amino acid mutations may be introduced into the antibodies using art recognized methods. In another embodiment, at each round of selection the stringency of binding to the desired target may be increased to select for antibodies with increased affinity for a desired target.

[00348] Screening methods may include purification of RNA-antibody fusion proteins from the components of an in vitro translation system. This may be accomplished using any art recognized method of separation. In one embodiment, the RNA-antibody fusion proteins may be separated by chromatography using a polydeoxythimidine (polydT) resin. In another embodiment, the RNA-antibody fusion proteins may be separated by chromatography using an antibody specific for an epitope present in the antibody component of the RNA-antibody fusion protein. In an embodiment, the epitope may be an amino acid sequence tag, for example, FLAG or HA tags, incorporated into the amino acid sequence of the antibody component of the RNA-antibody fusion protein, for example, at the N-terminal, C-terminal or in the inter variable region linker.

[00349] Selection of antibodies from a library may include the use of immobilized target molecules. In one embodiment, the target molecule may be directly linked to a solid substrate for example, agarose beads. In another embodiment, the target molecule may first be modified, for example, biotinylated and the modified target molecule may be bound via the modification to a solid support, for example, streptavidin-M280, neutravidin-M280, SA-M270, NA-M270, SA-MyOne, NA-MyOne, SA-agarose, and NA-agarose.

[00350] In some aspects, fluorescently-labeled antigens are used to single cell sort only plasmablasts or other B lineage cells with reactivity against specific, labeled antigens. In other aspects, fluorescently-labeled antigens are used to enrich for plasmablasts or other B lineage cells with reactivity against specific, labeled antigens, before single cell sorting occurs. In some aspects, fluorogenic or chromogenic molecules may be used to identify and sort B lineage cells. In some aspects, desired plasmablasts or other B lineage cells may be isolated by magnetic-activated cell sorting (MACS) or even by panning. Products resulting are generally monoclonal antibodies, against a variety of targets, including but not restricted to: cancer antigens, cytokines, chemokines, growth factors, secreted proteins, cell surface and other antigens to deplete cell

types of specific interest, microbes, bacteria, mycobacteria, parasites, and viruses. Other screening methods are described in the Examples section below.

Computer implementation

[00351] In some aspects, one or more methods described herein can be implemented on a computer. In one embodiment, a computer comprises at least one processor coupled to a chipset. Also coupled to the chipset are a memory, a storage device, a keyboard, a graphics adapter, a pointing device, and a network adapter. A display is coupled to the graphics adapter. In one embodiment, the functionality of the chipset is provided by a memory controller hub and an I/O controller hub. In another embodiment, the memory is coupled directly to the processor instead of the chipset.

[00352] The storage device is any device capable of holding data, like a hard drive, compact disk read-only memory (CD-ROM), DVD, or a solid-state memory device. The memory holds instructions and data used by the processor. The pointing device may be a mouse, track ball, or other type of pointing device, and is used in combination with the keyboard to input data into the computer system. The graphics adapter displays images and other information on the display. The network adapter couples the computer system to a local or wide area network.

[00353] As is known in the art, a computer can have different and/or other components than those described previously. In addition, the computer can lack certain components. Moreover, the storage device can be local and/or remote from the computer (such as embodied within a storage area network (SAN)).

[00354] As is known in the art, the computer is adapted to execute computer program modules for providing functionality described herein. As used herein, the term "module" refers to computer program logic utilized to provide the specified functionality. Thus, a module can be implemented in hardware, firmware, and/or software. In one embodiment, program modules are stored on the storage device, loaded into the memory, and executed by the processor.

[00355] Embodiments of the entities described herein can include other and/or different modules than the ones described here. In addition, the functionality attributed to the modules can be performed by other or different modules in other embodiments. Moreover, this description occasionally omits the term "module" for purposes of clarity and convenience.

Kits

[00356] A kit can include a polynucleotide, a polynucleotide library, a vector, and/or a host cell disclosed herein and instructions for use. The kits may comprise, in a suitable container, a polynucleotide, a polynucleotide library, a vector, and/or a host cell disclosed herein, one or

more controls, and various buffers, reagents, enzymes and other standard ingredients well known in the art.

[00357] The container can include at least one well on a plate comprising one or more wells. The container can include at least one vial, test tube, flask, bottle, syringe, or other container means, into which a polynucleotide, a polynucleotide library, a vector, and/or a host cell may be placed, and in some instances, suitably aliquoted. Where an additional component is provided, the kit can contain additional containers into which this component may be placed. The kits can also include a means for containing the polynucleotide, a polynucleotide library, a vector, and/or a host cell and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained. Containers can include labeling with instructions for use and/or warnings.

EXAMPLES

[00358] The examples are offered for illustrative purposes only, and are not intended to limit the scope of any embodiment of the present invention in any way. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

[00359] Various methods can employ, unless otherwise indicated, conventional methods of protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. *See, e.g.,* T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols A and B(1992); *Current Protocols in Molecular Biology* (2002- ; Wiley; Online ISBN: 9780471142720; DOI: 10.1002/04711142727); *Current Protocols in Immunology* (2001- ; Wiley; Online ISBN: 9780471142737; DOI: 10.1002/0471142735).

General Materials and Methods

[00360] **Blood collection and isolation of PBMCs**

[00361] All human samples were collected after informed consent and under Investigational Review Board (IRB)-approved human subject protocols. Blood was collected in heparin tubes (Beckton Dickinson and Company, catalog #BD366664) or in CPT tubes (Beckton Dickinson

and Company, catalog BD362761) tubes. For processing of the heparin tubes, one milliliter of blood was transferred into a microfuge tube and spun down at 12,000 rpm for 3 minutes, plasma was collected and frozen at -80°C (for later testing for antibody reactivities), the remainder of the blood was layered over Ficoll and centrifuged in a Beckman Coulter Allegra X-15R benchtop centrifuge with a SX4750 Swinging Bucket Rotor at 800 g for heparin tubes for 20 min at room temperature, with minimal acceleration and without use of the brake, and the peripheral blood mononuclear cell (PBMC) layer was collected. Alternatively, CPT tubes were directly centrifuged at 1,500g for 20 min at room temperature, with minimal acceleration and without use of the brake, and the PMBC layer was collected. The collected PBMCs were then washed twice with PBS before use.

[00362] PBMCs may also be frozen for future use and isolation of B-cells, memory B-cells, plasmablasts, plasma cells, or other B-cell populations. One method for freezing PBMCs involves resuspending the PBMCs in 90% fetal bovine serum (FBS) and 10% dimethyl sulfoxide (DMSO) in cryovials, and then slowly freezing the cells contained in the vials overnight at -80°C in a Mr. Frosty (Sigma C1562-1EA). The vials of frozen cells were then transferred for long-term storage in liquid nitrogen, and can be thawed at a later date for the isolation of individual B-cells and for the high-throughput sequencing of paired immunoglobulin genes. Thawed cells were incubated in a media containing an excess of DNase I, usually 25ug/ml (Sigma D4513) till the end of the 1st sort to prevent cell clumping.

Isolation and enrichment of cells and cell subpopulations

[00363] ***Plasmablasts.*** For some samples, PBMCs were first enriched for plasmablasts by using a modified Plasma Cells Isolation Kit II (Miltenyi 130-093-628). This is an optional step. This yielded fewer total cells for subsequent sorting, resulting in shorter sort times. This was used primarily when multiple samples needed to be single-cell sorted on the same day. It is also possible to use different kits to enrich for different B-cell populations (*see* below). For every $5x10^7$ PBMCs, cells were suspended in 200 µL of ice-cold MACS buffer (PBS with 0.5% FBS). 50 uL of non-plasma cell biotin-antibody cocktail was added, and cells were incubated in the fridge (4°C) for 10 minutes. 100 µL of MACS buffer, 100 µL of non-plasma cell microbead cocktail, and 50 µL of CD56 microbeads were added and incubated in the fridge for an additional 10 minutes. Cells were then washed with 7 mL of MACS buffer, centrifuged at 300g for 5 minutes at 4°C, resuspended in 500 µL of MACS buffer, and run on an equilibrated LS column in a magnetic field. The column was washed with 4 x 3 mL of MACS buffer and enriched cells were in the negative fraction.

[00364] ***Memory B-cells.*** CD19+ microbeads (Miltenyi 130-050-301) and CD27+ microbeads (130-051-601) may be used to enrich for memory B-cells before cell sorting, to shorten sort times. Other enrichment methods, such as Memory B-cell isolation kit (Miltenyi 130-093-546), may also be used, provided that they enrich for $CD19^+CD27^+$ cells. For every $5x10^7$ PBMCs, 300 µL of ice-cold MACS buffer is used for resuspension. 100 µL of CD19 microbeads and 100 µL of CD27 microbeads are then added, and the sample is incubated at 4°C for 15 minutes. Cells are then washed with 7 mL of MACS buffer, centrifuged at 300g for 5 minutes at 4°C, and resuspended in 500 µL of MACS buffer. Cells are then run through an equilibrated LS column in a magnetic field, and washed with 2x 3 mL of MACS buffer. The LS column is then removed from the magnetic field, and the cells are washed out with 5 mL of MACS buffer to elute the enriched cells.

[00365] ***Total B-cells.*** CD19+ microbeads (Miltenyi 130-050-301) may be used to enrich for total B-cells before cell sorting, e.g., to shorten sort times. Other enrichment methods may also be used, provided that they enrich for $CD19^+$ cells. For every $5x10^7$ PBMCs, resuspend cells in 400 µL of ice-cold MACS buffer. Add 100 µL of CD19+ microbeads and incubate in the fridge (4°C) for 15 minutes. Cells are then washed with 7 mL of MACS buffer, centrifuged at 300g for 5 minutes at 4°C, and resuspended in 500 µL of MACS buffer. Cells are then run through an equilibrated LS column in a magnetic field and washed with 2x 3 mL of MACS buffer. The LS column is then removed from the magnetic field, and the cells are eluted with 5 mL of MACS buffer, yielding the enriched cells.

[00366] ***Other cell types.*** Although not necessary, MACS enrichment of the desired cell population can shorten sort times. Other cell populations, including plasma cells, other B-cell populations and non-B-cell populations may also be enriched using MACS or other systems using the appropriate reagents. For example, total T-cells may be enriched using CD3+ microbeads, and effector T-cells and helper T-cells isolated using CD8+ and CD4+ microbeads, respectively. CD45RO microbeads may be used to isolate memory T-cells and, in conjunction with CD8+ or CD4+ beads, used to isolate memory effector or memory helper T-cells, respectively.

Single-Cell Sorting

[00367] MACS enrichment is not required for sorting, but MACS enrichment for plasmablasts may be performed to shorten sort times. If PBMCs have undergone MACS enrichment, an aliquot of unenriched PBMCs (~1 million cells) is also analyzed in tandem, allowing the baseline plasmablast percentage in the sample to be determined. For sorting

plasmablasts, cells were stained with manufacturer-recommended volumes of CD3-V450 (BD 560365), IgA-FITC (AbD Serotec STAR142F), IgM-FITC (AbD Serotec STAR146F) or IgM-PE (AbD Serotec STAR146PE), CD20-PerCP-Cy5.5 (BD 340955), CD38-PE-Cy7 (BD 335808), CD19-APC (BD 340437) and CD27-APC-H7 (BD 560222) in 50 μL of FACS buffer (PBS or HBSS with 2% FBS) on ice for 20 minutes in the dark. Some cells may also be stained with IgG-PE (BD 555787), CD138-PE (eBioscience 12-1389-42), or HLA-DR-PE (BD 555812) together with IgM-FITC instead. For simultaneous sorting of plasmablasts, memory and naive B-cells, the following staining scheme was used: IgD-FITC (Biolegend 348205), IgG-PE (BD 555787), CD20-PerCP-Cy5.5, CD38-PECy7, IgM-APC (BD 551062), CD27-APC-H7, IgA-biotin (AbD Serotec 205008) followed by Strepavidin-eFluor710 (eBioscience 49-4317-82) and CD19-BV421 (Biolegend 302233). Memory B-cells have also been sorted either as $CD19^+CD27^+IgG^+$ or $CD19^+CD20^+IgG^+$, naive B-cells have been sorted as $CD19^+IgD^+IgM^+$. IgA^+ plasmablasts have also been sorted, and are defined as $CD19^+CD20^-$ $CD27^+CD38^{++}IgA^+IgM^-$. Other cell surface markers may also be used, as long as the B-cell or other cell population is phenotypically identifiable using cell surface markers, the population can be single-cell sorted. *See* below. Cells were then washed once with 2 mL of FACS buffer and resuspended at an appropriate volume for FACS. Cells were first sorted on a BD Aria II into a 5 mL round bottom tube. Typically, purities of >80% were achieved from the first sort. Single cells were sorted into the first 11 columns of a 96-well PCR plate containing 6.65 μL of a hypotonic buffer (10mM Tris-HCl pH 7.6) containing 2 mM dNTPs (NEB N0447L), 5 μM oligo$(dT)_{20}$VN, and 1 unit of Ribolock (Fermentas EO0384), an RNase inhibitor. As a negative control, the last column was left devoid of cells. For IgG^+ plasmablasts, the gating (selection of cells) strategy was $CD19^+CD20^-CD27^+CD38^{++}IgA^-IgM^-$. Sorted plates were sealed with aluminum plate sealers (Axygen PCR-AS-600) and immediately frozen on dry ice and stored at -80°C.

[00368] ***Single-cell sorting gating strategies***

[00369] *B-cells.* For B-cells, the gating approach comprises sorting for one or more of the following markers: IgM, IgG, IgA, IgD, CD19, or CD20. For total IgG^+ B-cells, the gating approach comprises sorting for IgG^+. For total IgA^+ B-cells, the gating approach comprises sorting for IgA^+. For total IgM^+ B-cells, the gating approach comprises sorting for IgM^+.

[00370] *Activated B cells.* Activated B cells include B cells that have been stimulated through binding of their membrane antigen receptor to its cognate antigen and/or have received T cell help from T cells recognizing epitopes derived from the same macromolecular antigen. Activated B cells can be identified by a variety of properties including increased cell size (e.g.

"blasting B cells"; see below), expression of cell surface marker or markers, expression of intracellular marker or markers, expression of transcription factor or factors, exiting the gap 0 (G0) phase of the cell cycle, progressing through the cell cycle, production of cytokines or other factors, and/or the down regulation of certain cell surface marker or markers, intracellular marker or markers, transcription factor or other factor. One method of identifying an activated B cell is to combine detection of a B cell marker such as CD19 or immunoglobulin with a marker of activation such as increased cell size or volume, the cell surface activation marker CD69, or progression through the cell cycle based on cell-permeable acridine orange DNA stain or another cell cycle analysis.

[00371] *Blasting B cells.* "Blasting B cells" are B cells that are activated and increased in size relative to resting B cells. Blasting B cells include the plasmablast population as well as other populations of activated B cells, and blasting B cells are physically larger in size than resting B cells. Blasting B cells can be single-cell sorted using several different approaches, including gating (selection) of B cells based on their physically being larger based on cell diameter, cell volume, electrical impedance, FSC, the integral (area) of a FSC pulse (FSC-A), FSC height (FSC-H), forward scatter pulse width (FCS-W), side scatter (SSC), side scatter pulse area (SSC-A), side scatter height (SSC-H), side scatter width (SSC-W), autofluorescence and/or other measures of cell size.

[00372] In flow cytometry, forward scatter (FSC) is measured using a light beam in line with the stream of cells and provides information regarding the proportional size and diameter of each cell. Using FSC one can select B cells with FSC greater than the median FSC of resting B cell, for example an FSC-A or FSC-H 5% greater than resting B cells, 10% greater than resting B cells, 15% greater than resting B cells, 20% greater than resting B cells, 30% greater than resting B cells, 40% greater than resting B cells, 50% greater than resting B cells, 60% greater than resting B cells. By analyzing calibration beads of specific sizes, one can use FSC to determine the relative size of B cells relative to the calibration beads. By doing so, one can specifically gate on and thereby select B cells that possess diameters of about 8um, >8 um, >9 um, >10 um, >11 um, >12 um, >13 um, > 14 um, >15 um, > 16 um, >17 um, >18 um, >19 um, or >20 um.

[00373] Another measurement of cell size is cell volume. The "gold standard" for cell volume uses the Coulter principle which is based on an electronic measurement (Tzur et al, PLoS ONE, 6(1): e16053. doi:10.1371/journal.pone.0016053, 2011). Although the method of sorting by droplet charging and deflection was first used in a device that measured cell volume by impedance, the currently available flow cytometers make only optical measurements. FSC measurements, specifically the FSC-A (FSC integral area) are commonly used to assess cell

size, although FSC measurements can be influenced by the refractive index differences between particles and fluid (Tzur et al, PLoS ONE, 6(1): e16053. doi:10.1371/journal.pone.0016053, 2011). Some have shown that volume estimation can be improved by combining optical parameters, including FSC-W, SSC and 450/50-A auto fluorescence (Tzur et al, PLoS ONE, 6(1): e16053. doi:10.1371/journal.pone.0016053, 2011).

[00374] For example, selection of activated B cells based on increased size can be achieved through identifying B cells using a marker such as CD19 and assessing size through FSC or FSC-A. Other B cell markers and/or parameters for assessment of size are described herein.

[00375] *Plasmablasts.* For isolation of plasmablasts, the gating approach comprises sorting for $CD19^{+}CD38^{++}$ B-cells. For isolation of IgG^{+} plasmablasts, the gating approach comprises sorting for $CD19^{+}$ $CD38^{++}IgA^{-}IgM^{-}$ B-cells. For isolation of IgA+ plasmablasts, the gating approach comprises sorting for $CD19^{+}CD38^{++}IgA^{+}$ B-cells. For isolation of IgM+ plasmablasts, the gating approach comprises sorting for $CD19^{+}CD38^{++}IgM^{+}$ B-cells. In addition, other gating strategies can be used to isolate a sufficient number of plasmablasts to carry out the methods described herein. Plasmablasts were also isolated using the following marker expression patterns $CD19^{low/+}$, $CD20^{low/-}$, $CD27^{+}$ and $CD38^{++}$. Although use of all these markers generally results in the purest plasmablast population from single cell sorting, not all of the above markers need to be used. For example, plasmablasts may also be isolated using the following gating strategies: forward scatter high (FSC^{hi}) for larger cells, $FSC^{hi}CD19^{lo}$ cells, FSC^{hi} and $CD27^{+}$, $CD38^{++}$, or $CD20^{-}$ cells. Combination of any of these markers or other markers found to be able to distinguish plasmablasts from other B-cells will generally increase the purity of sorted plasmablasts, however any one of the above markers alone (including FSC^{hi}) can distinguish plasmablasts from other B-cells, albeit with a lower purity.

[00376] *For memory B-cells.* For IgG^{+} memory B-cells, the gating approach comprises sorting for $CD19^{+}$ $CD27^{+}IgG^{+}$ or $CD19^{+}$ $CD20^{+}IgG^{+}$. For IgA^{+} memory B-cells, the gating strategy comprises $CD19^{+}CD27^{+}IgA^{+}$ or $CD19^{+}CD20^{+}IgA^{+}$. For IgM^{+} memory B-cells, the gating strategy comprises $CD19^{+}CD27^{+}IgM^{+}$or $CD19^{+}CD20^{+}IgM^{+}$.

[00377] *For other cell types.* As long as the B-cell, T-cell, or other cell population is phenotypically identifiable using cell markers, it can be single-cell sorted. For example, T-cells can be identified as $CD3^{+}$ or TCR^{+}, naïve T-cells identified as $CD3^{+}CD45RA^{+}$, memory T-cells identified as $CD3^{+}CD45RO^{+}$. Effector and helper T-cells can be identified as $CD3^{+}CD8^{+}$ and $CD3^{+}CD4^{+}$, respectively. Cell populations can be further subdivided by using combinations of markers, such as $CD3^{+}CD4^{+}CD45RO^{+}$ for memory helper T-cells.

[00378] Reverse transcription with adaptor molecules

[00379] Single-cell sorted plates were thawed on ice and briefly centrifuged before use. Plates were incubated in the thermal cycler at 55°C for 3 minutes, 42°C for 2 minutes, and indefinitely at 4°C. Plates were briefly centrifuged again and carefully opened to avoid the formation of aerosols. 1 μL of a 10 μM solution of the appropriate adapter molecule (each adapter molecule generally has a sample identification region (sample-ID)) was added to each well, with all negative control wells (containing RNA preservative buffer alone, or non-B-cells) receiving identical adapter molecules. 2.35 μL of a mix containing 0.75 μL H_2O, 1 μL of 10x M-MuLV RT buffer (NEB B0253S), 0.6 μL of 50mM $MgCl_2$, 0.25 μL of Ribolock (40 U/μL), and 0.125 μL of Superscript III (200 U/μL) (Invitrogen 18080-085) was added and mixed by pipetting. Plates were briefly centrifuged and incubated at 42°C for 120 minutes to 8 hours using a thermal plate shaker and then kept at -20°C. After the reaction, RT products from all wells were pooled in a microfuge tube. Pooled RT products were then extracted with phenol-chloroform-isopropyl alcohol with ~0.1% 8-hydroxychloroquine (Sigma 77617), and then extracted with chloroform extraction in gel-lock phase tubes (5 PRIME 2302820). RT products were then concentrated and desalted by 5-minute spins at 14 000g with Amicon Ultra-0.5 30kDa (Millipore UFC503096) or Ultra-0.5 100kDa (Millipore UFC510096), followed by a 5 min spin at 14 000g with TE (10mM Tris-HCl pH 7.6 with 1mM EDTA) and a final 5-minute spin at 14 000g with EB (Qiagen 19086). RT products were eluted by inverting the Amicon Ultra column in a new centrifuge tube and centrifuging at 1000g for 2 minutes. At this point, RT products were kept at -20°C or -80°C.

[00380] Touchdown PCR

[00381] For 454 sequencing runs 1 and 2, the Touchdown PCR method was used as follows. For some samples in PCR runs 3 and 4, the PCR method was changed, leading to increased numbers of paired heavy and light chains. This change is detailed under the sub-section "Non-touchdown PCR" below.

[00382] For both the 1st PCR and the nested 2nd PCR, Phusion Hot Start II DNA polymerase (NEB F-549L) was used in the provided GC buffer. For IgG, primers and adapter molecules are shown in Table 1. Sample-ID sequences are shown in Table 2. Plate-ID sequences are shown in Table 3. *See* also Figs. 3 and 9. Reaction conditions included a final $MgCl_2$ concentration of 1.8 mM, 200 μM dNTPs, 0.2 μM for all primers, 0.2 U of Phusion polymerase, varying amounts of DMSO as an additive and 2 μL of template in a final volume of 25 μL. For the 1st PCR, lambda and kappa light chains and the gamma heavy chain were amplified in different wells, and

DMSO is used at a final concentration of 8%, 5% and 10%, respectively. Forward primers used in the 1st PCR were the FW long primer1 and the FW short primer1. Because the FW long primer1 added a plate identification region (plate-ID) to the 5' end of amplicon regions (amplicons), FW long primer1 containing different plate-IDs was added to different samples. Gene-specific reverse primers were used to amplify the kappa, lambda, and gamma chains were kappa GSP1, lambda GSP1, and gamma GSP1, respectively. Cycling conditions for the 1st PCR included an initial denaturation step at 98°C for 30'', followed by 2 cycles of 98°C for 10'', 72°C for 25''; 7 touchdown cycles of 98°C for 10'', 71.5°C to 68.5°C for 15'' and 72°C for 20'' with a drop of 0.5°C for each subsequent annealing step; 30 cycles of 98°C for 10'', 68°C for 15'' and 72°C for 20'', followed by a final extension at 72°C for 5' and hold at 4°C indefinitely. Products from 1st PCR were diluted 100x in TE and 2 µL used for the nested 2nd PCR. For the 2nd PCR, 5% DMSO was used as an additive in all samples. Forward primer is the FW primer2 and reverse primers were the RV primer2 and the GSP long primer2. Kappa GSP long primer2, lambda GSP long primer2 and gamma long primer2 were used to amplify their respective amplicons. Because the GSP long primer2 also added the plate-ID to the 3' end of the amplicons, a different GSP long primer2 with plate-specific plate-IDs was added to each pooled-plate sample. Cycling conditions for the nested 2nd PCR included an initial denaturation step of 98°C for 30'', 30-40 cycles of 98°C for 10'', 67°C for 15'', and 72°C for 20'', followed by a final extension of 72°C for 5' and hold at 4°C indefinitely.

[00383] Non-Touchdown PCR

[00384] For the non-touchdown PCR, conditions were identical to the touchdown PCR unless otherwise stated. The 1st PCR cycling parameters were an initial denaturation of 95°C for 5', 15-25 cycles of 98°C 30", 62°C 30", 72°C 30", a final extension of 72°C 5' and hold at 4°C indefinitely. 1st PCR was a multiplex PCR, where all 3 gene-specific reverse primers, the kappa, lambda, and gamma constant regions reverse primers were used in conjunction at 0.2, 0.2 and 0.24 µM, respectively. All other primers used were the same as in touchdown PCR. The gene-specific primers can be those used in touchdown PCR and also any one of those designated as suitable for 1st PCR (Table 6). DMSO was used at a final concentration of 5%; 0.1mg/ml of BSA (NEB B9001S), and ET-SSB (NEB M2401S) may also be added at 1:100 for the PCR reaction. During the 1st PCR, 4-6ul of cDNA template was used in a total 80 or 90ul reaction volume. Each PCR1 reaction was split into eight or nine 10ul reactions, each occurring in a different well. 1st PCR was pooled again after the PCR and diluted 100x in TE0.1, and 2ul used for 2nd PCR. 2nd PCR is a separate reaction for each gene-specific primer (not multiplex), and the reaction mix was identical to the touchdown 2nd PCR except for the following: any of the

gene-specific constant region primers designated as working for the 2nd PCR may be used (Table 6), primers were used at either 0.2μM or 0.4μM throughout, or the gene-specific primers were used at 0.2μM and the rest used at 0.4μM. 0.1mg/ml BSA was added to the reaction and ET-SSB may also be used at 1:100. The 2nd semi-nested PCR cycling parameters were an initial denaturation of 95°C for 5', 20-35 cycles of 98°C 30", 67°C 30", 72°C 30", a final extension of 72°C 5' and hold at 4°C indefinitely. The total number of PCR cycles for 1st and 2nd PCR combined was typically between 50-60 cycles for the non-touchdown PCR. As different pooled-wells undergoing the PCR cycling tend to use different number of cycles to obtain a reasonable amount of DNA product (typically between 1-12ng/ul), 4 different PCR cycles were carried out for each 2nd PCR, e.g. 23, 26, 30 and 33 cycles, 5ul run on a 2% agarose gel and compared. Based on qualitative judgment of the amount of PCR product, only PCR product from one of the 2nd PCR cycle numbers was used for each pooled-well 2nd PCR in preparing for the 454 sequencing run.

[00385] For PCR of other immunoglobulin heavy chains in humans, immunoglobulin heavy and light chains in mice and TCR chains in humans and mice, PCR conditions are identical to the non-touchdown PCR section above except that 1st PCR is non-multiplex, with each cDNA being individually amplified. The following 3' primers in Tables 10 and 11 are used in PCR1 and 2.

[00386] Preparing for 454 XLR70 sequencing run

[00387] For the 1st and 2nd 454 runs, sequencing primers (Titanium Primers A and B, respectively) for a 454 Titanium sequencing run were added onto the amplicons during the 1st and nested 2nd PCRs. 5 μL of each amplicon were run on an agarose gel with a mass DNA ladder (Fermentas SM0383), an image was taken, and band intensities were analyzed and quantified with AlphaFC Imager software (Cell Biosciences). 5 ng of each of the kappa, lambda, and gamma amplicons were separately pooled, run on a 0.8% agarose gel, and visualized with GelGreen (Biotium 41005). Bands of the appropriate sizes (~600bp for kappa and lambda, and ~750bp for gamma) were cut and purified with MinElute Gel Extraction kit (Qiagen 28606), according to manufacturer's instructions with slight modifications. Briefly, agarose gel was melted in QG buffer without heating, and the additional QG wash step is done. PE wash buffer was allowed to sit for 5 minutes before spinning. An additional PE wash step was also performed. Samples were eluted with 25 uL of EB buffer. Samples were also cleaned once with SPRI beads using a ratio of 1:0.65 for DNA volume: bead volume for 454 2nd run. DNA concentration was determined with Picogreen DNA assay kit (Invitrogen P11496), and samples were pooled such that DNA concentration of gamma:kappa:lambda is 2:1:1. Pooled

samples were at a concentration of >0.5ng/μL and were shipped to a 454 DNA sequencing facility for 454 sequencing.

[00388] For the 3rd and subsequent 454 sequencing runs, the protocol was changed. The amplicons were still separately pooled to normalize DNA quantities from each PCR reaction, but first underwent an SPRI bead cleanup to remove small DNA fragments according to manufacturer's instructions. The amplicons were run on a 3% agarose gel, and the appropriate bands cut and purified with MinElute Gel Extraction kit as before. Thereafter, the amplicons underwent another 2 rounds of SPRI bead cleanup to remove even more small DNA fragments, and quantitated with Picogreen, quality checked with Nanodrop to ensure the OD260/280 ratio was >1.8 and 1μl run on a gel to ensure there were no small DNA fragments. Lambda and kappa amplicons were pooled in a 1:1 ratio, gamma was used as-is. DNA was then diluted to 1 x 10^9 copies per 454's instructions, and sent to sequencing facility (Roche) for emPCR at 1cpb and sequenced; gamma heavy chain in one region and the pooled light chains in the other region of the picotiter plate.

[00389] **Preparing for 454 XL+ sequencing run**

[00390] Currently the 454 XL+ sequencing run does not support the Lib-A sequencing kit that was used for the XLR70 run. XL+ currently only supports the Lib-L kit, which is unidirectional sequencing. To adapt our protocol to do XL+ sequencing, the protocol for XLR70 run is followed, but after the gel cleanup step, each amplicon (kappa, lambda and gamma) underwent 2 separate PCRs, each 5 cycles long to add on the Lib-L A and B adaptors. PCR conditions are as follows: Phusion polymerase is used, with 5x GC buffer and a final concentration of 5% DMSO. Primers are used at 0.2uM. 0.1mg/ml BSA is added to the reaction. The PCR cycling parameters are an initial denaturation of 95°C for 5', 20-35 cycles of 98°C 30", 67°C 30", 72°C 30", a final extension of 72°C 5' and hold at 4°C indefinitely. Two PCRs are done for each amplicon: 5LIB-LA and 3LIB-LB in one PCR, and 5LIB-LB and 3LIB-LA in the other PCR. The adaptors are added such that each amplicon becomes either 5'-LibA-amplicon-LibB-3' or 5'-LibB-amplicon-LibA-3'. These amplicons have either the LibA "A" or "B" adaptors on the 5' end (and the corresponding "B" or "A" adaptor on the 3' end), which allows for bidirectional sequencing. Amplicons with the new Lib A adaptors then undergo 3 rounds of SPRI bead cleanup before following the protocol for XLR70 runs to quantitate and quality check the DNA before diluting it to 1 x 10^9 copies and sending to a 454 sequencing facility (Roche) for emPCR at 1cpb and sequencing.

[00391] **Preparing for PacBio sequencing run**

[00392] For PacBio sequencing run, touchdown PCR was employed as above. DNA pooling and cleanup was done as the section above on "preparing for 454 XLR70 run." To obtain sufficient DNA (500ng) for sequencing requirements, a minimum of 1ug of DNA was pooled for gel and SPRI cleanup. Picogreen quantitation and $1x10^9$ dilutions were not done as it was not required for PacBio sequencing. If insufficient DNA was obtained from the 2nd PCR, the 2nd PCR and pooling steps were repeated until sufficient DNA was obtained. A minimum of 500ng of cleaned-up DNA was sent to PacBio sequencing facility for sequencing.

[00393] **Other sequencing approaches**

[00394] The methods disclosed herein are not dependent on 454 or PacBio sequencing. Lambda and kappa light chains are ~600bp and gamma heavy chain ~700bp. Thus, what is generally desired is the ability to have longer sequencing reads such that the forward and reverse sequencing reads overlap enough to enable reconstruction of the entire, approximately 600 bp sequence of the light chains (LCs) (exact sequence length depends on the length of the 5' untranslated region (UTR)), and approximately 700 bp sequence of the heavy chains (HCs). Therefore, any sequencing technology that can yield sequencing reads of at least about 350-400bp and thereby achieve the overlap used for sequence assembly can be utilized, and sequencing technologies that enable app. 600-700+ bp reads would allow one to sequence using just the forward (Fw) primer (sequencing from the 5' end).

Sequences

[00395] The sequence data for the above runs was received from the relevant facility and processed as described below.

[00396] **Sequence Nomenclature**

[00397] Each sequence in the sequence listing which corresponds to a sequencing read, sequence assembly or amino acid translation from a sequence has an identifier. Each such identifier has 9 fields separated by a period, ".". The fields are numbered from 1 to 9 and give the following information:

1. Read ID. A Read ID assigned by the software associated with the sequencing technology used to determine the read, or "NA" if the sequence is not a raw read.
2. Plate Number. A plate number that the sequence is associated with. See Table 12 (plate to sample mapping table) for corresponding biological sample information.
3. Sample ID. Sample ID indicating the well that the sequence is associated with. Sample ID numbers are between 1 and 89 inclusive. *See* Table 2 for correspondence between Sample ID and Well Name.
4. Well Name. Well name containing the well that the sequence is associated with. The well name corresponds to the usual 96 well plate name, e.g. D07. The well name and the Sample ID are equivalent ways of specifying a particular well on a plate.

5. Contig ID. The contig ID distinguishes different sequences associated with a well from a given assembly and chain type.
6. Platform. The platform field indicates the sequencing technology that the sequence is derived from. The possible values for platform are 454, Sanger, and PacBio.
7. Chain Type. The chain type field indicates whether the sequence is associated with a set of heavy chain antibody sequences, light chain antibody sequences, or a set containing both heavy and light chain antibody sequences. Possible values are "heavy", "light" or "CMB".
8. Run ID. An identifier for a set of reads on a particular platform.
9. Sequence Type. The type of the sequence. Possible values are "raw" for raw sequencing technology reads, "nb", "urt", "multim50", "zerom50" or "pb" for assembled reads (see Assembly of Sequences section), or "nb-aa", "urt-aa", "multim50-aa", "zerom50-aa" or "sanger-aa" for amino acid sequences derived from the various nt assembly consensus sequences.

[00398] Preparation of sequences for analysis

[00399] Data generated from the 454 sequencing were analyzed by 454 GS FLX data analysis software, and filter-passed high quality sequences were returned. Due to the stringency of the default amplicon filter used by 454 GS FLX data analysis software, filter stringency may need to be relaxed to obtain sufficient long reads. One way is to follow the suggestions in 454 technical bulletin APP No. 001-2010. Changing <vfScanAllFlows> from "TiOnly" to "False" of the amplicon filter can lead to a large increase in filter-passed sequences of good quality. Another option is to change the <vfTrimBackScaleFactor> to a lower number. For 454 run 1, standard shotgun processing was used and for run 2, <vfScanAllFlows> was changed to "False", and standard amplicon pipeline processing was used for runs 3 and 4.

[00400] Data generated from Pacific Biosciences sequencing was received from Pacific Biosciences as Circular Consensus Sequence reads with associated quality scores.

[00401] Assignment of Sequences to Wells

[00402] cDNA from samples was sequenced with either 454 or Pacific Biosciences sequencing technology. The reads are those in the Sequence Listing whose Sequence Type is "raw". The sequencing reads were analyzed and either assigned to a source plate and well or discarded.

[00403] Plate and well assignments for reads were made by comparing the observed read sequence to the possible plate identification region, universal primer region, and sample identification region sequences using regular expressions. The comparison was done in three stages using the regular expressions listed in Tables 13, 14 and 15.

[00404] In stage 1, analysis of possible plate identification regions, a read was checked against all of the regular expressions listed in column "Plate Identification Region Regular Expression" in Table 13, requiring a match to begin with the first nucleotide of the sequence. If

no match was found the read was discarded and plate/well assignment continued with the next available read to process, if any. If a match was found, the sequence was assigned the corresponding ID from the "Plate ID" column as its plate ID. The nucleotides of the read matching to the plate regular expression were recorded for use during later stages of matching and during assembly.

[00405] In stage 2, analysis of the universal primer region, a read was checked against the "universal primer regular expression" "CACGACCGGTGCTCGATT+AG", requiring a read match to begin with the first nucleotide following the last read nucleotide matching to the Plate Regular Expression. If the read did not match the universal primer regular expression, the read was discarded and plate/well assignment continued with the next available read to process, if any. Otherwise the nucleotides of the read matching to the universal primer regular expression were recorded for use during the last stage of matching and during assembly.

[00406] In stage 3, analysis of possible sample identification regions, a read was checked against all of the regular expressions listed in column "Sample Identification Region Regular Expression" in Table 14, requiring a match to begin with the first nucleotide following the last read nucleotide matching to the universal primer regular expression. If no match was found the read was discarded and plate/well assignment continued with the next available read to process, if any. If a match was found and the sample ID column contained only a single identifier, the sample ID of the read was assigned to be the ID found in the sample ID column. If the sample ID column contained more than a single identifier, those identifiers were considered "candidate sample IDs". The read was then checked sequentially against all of the regular expressions listed in column "Sample Identification Region Regular Expression" of Table 15 where at least one of the corresponding sample IDs in the "Sample ID" column of Table 15 matched with a candidate sample ID. If the read matched the regular expression, and the match began with the first nucleotide after the last read nucleotide matching to the universal primer regular expression, the right-most identifier from the candidate sample IDs was assigned as the read's sample ID. Otherwise, the right-most identifier was removed from the list of candidate sample IDs and the process repeated with the smaller list of candidate sample IDs until either a match was found, or, if no matching regular expression was found in the list of regular expressions in Table 15 then the last candidate sample ID (that is, the left-most in the original list of candidate sample IDs) was assigned as the sample ID for the read.

[00407] Reads that were discarded during the plate ID and sample ID assignment process were not included in the sequence listing.

[00408] **Assembly of Sequences**

[00409] All sequence reads assigned to a sample ID associated with a well were assembled to produce consensus sequences. These consensus sequences correspond to the heavy and light chain mRNA sequences expressed in the sorted cells.

[00410] Sequences were assembled with Newbler 2.5 (Runs 1 and 2), and Newbler version 2.6 and/or Mira version 3.4.0 for other sequences.

[00411] Sequences in the listing with a Platform field of "454", a Chain Type field of "mixed", a Run ID of "1" or "2" and a Sequence Type of "nb" are contigs resulting from an assembly using newbler. To assemble these sequences, sff output files from 454 sequencing, which contain both sequences and quality scores for each nucleotide, were read into Python using the Biopython package and sequences subdivided according to their compound barcodes (sample-ID + plate-ID) as described above and output into separate sff files. These files were then reparsed by sfffile (provided by GS FLX data analysis software) into sff files with file headers understood by Newbler, a sequence assembler provided in the GS FLX data analysis software suite, using the "-force", "-cdna" and "-urt" options. Newbler then assembled forward reads with shared compound barcodes. Because reverse reads have only a 3' plate-ID, it is possible that sequence assembly could occur between forward and reverse reads of sequences from different cells. For circumventing this potential problem, the heavy- and light-chain V(D)J usage of both assembled forward and unassembled reverse reads can be first identified using HighV-QUEST (http://imgt.cines.fr/HighV-QUEST/index.action). Sequences can then further grouped according to their V(D)J usage before being assembled again with Newbler using one assembled forward read and the reverse reads that share the same V(D)J usage. This can be repeated for all assembled forward reads. Sequence assembly can also be done to be intolerant of nucleotide mismatches, thereby preventing assembly of forward and reverse reads from different cells that share the same V(D)J usage. This way, inappropriate sequence assembly of reverse reads between highly similar sequences from different cells can be largely avoided.

[00412] Sequences in the listing with a Platform field of "454", a Chain Type field of "heavy" or "light", a Run ID of "3" or "4" and a Sequence Type of "nb" are contigs resulting from assemblies of 454 reads, executing Newbler with this command line: runAssembly -cdna -o output seqs.fastaq where seqs.fastq contained a single well's trimmed reads in FastQ format.

[00413] Any wells for Run ID 3 or Run ID 4 reads which Newbler did not create exactly one heavy chain contig or exactly one light chain contig were reanalyzed by assembling with mira. Sequences in the listing with a Platform field of "454", a Chain Type field of "heavy" or "light", a Run ID of "3" or "4" and a Sequence Type of "multim50" or "zerom50" are contigs resulting from these assemblies, executing mira with this command line: mira --project=seqs --

job=denovo,est,accurate,454 454_SETTINGS -ED:ace=yes -AL:egp=no -CL:pvlc=yes --fastq –notraceinfo

[00414] A file named seqs_in.454.fastq contained a single well's trimmed reads in FastQ format.

[00415] For wells from Run ID 3 or Run ID 4 reads where neither Newbler nor mira created contigs using the above assembly commands, a different Newbler command was executed. Sequences in the listing with a Platform field of "454", a Chain Type field of "heavy" or "light", a Run ID of "3" or "4" and a Sequence Type of "urt" are contigs resulting from these assemblies, where Newbler was executed with this command line: runAssembly -cdna -ud -urt -o output seqs.fastaq where seqs.fastq contained a single well's trimmed reads in FastQ format.

[00416] Sequences in the listing with a Platform field of "PacBio", a Chain Type of "heavy" and a Sequence Type of "pb" are contigs resulting from assemblies of reads from the PacBio platform, executing mira with the command line: mira --project=seqs --job=denovo,est,accurate,454 454_SETTINGS -ED:ace=yes -AL:egp=no -CL:pvlc=yes --fastq –notraceinfo

[00417] A file named seqs_in.454.fastq contained a single well's trimmed reads in FastQ format.

[00418] **Amino Acid Sequences**

[00419] Sequences in the listing with a Platform Field of "454" and a Sequence Type of "nb-aa", "urt-aa", "multim50-aa" or "zerom50-aa" are amino acid sequences determined by translating the nucleotide sequences of assemblies of 454 reads as described under "Assembly of Sequences".

[00420] Sequences in the listing with a Platform Field of "PacBio" and a Sequence Type of "pb-aa" are amino acid sequences determined by translating the nucleotide sequences of assemblies of Pacifc Biosciences reads as described under "Assembly of Sequences".

[00421] Sequences in the listing with a Sequence Type of "sanger-aa" are amino acid sequences determined by directly translating reads determined by Sanger sequencing.

[00422] **Other sequencing data analysis options**

[00423] The workflow of data analysis described above can be used to accurately determine the heavy- and light-chain sequences of each cell. However, this information is not absolutely necessary for our "selection screening" approach (*see* "screening of expressed human antibodies"). For the selection screen to work, we first cluster paired antibody sequences into clonal families, on the basis of their heavy chain V(D)J usage and light chain VJ usage. Therefore, we do not require the full sequence of the immunoglobulin heavy and light chains,

and can use sufficient sequence information to determine V(D)J usage. Therefore, we can tolerate sequencing errors and can use lower-quality reads generated by 454 or any other sequencing technology. Sequence assembly of the forward read is not generally an issue because all sequences can be first grouped according to their compound barcode before being assembled. Because each compound barcode comes from one sample/cell, there is only one 'correct' sequence for each immunoglobulin chain with the same compound barcode. Sequencing errors in different strands can then be averaged out because it is unlikely that all sequencing errors occur at the same bases, meaning that taking the consensus base sequence will give the most accurate sequence. In cases of ambiguity, bases with high Phred quality scores will generally be chosen instead. Because 454 trims sequence reads from the 3' end until only higher quality reads remain, this can result in very short reads. With our method, we can tolerate lower-quality reads and thereby use much longer reads generated by 454 (400-500bp). With these longer reads, we can identify V(D)J usage without requiring assembly of the forward read with the reverse read, thereby making the 3' plate-ID non-essential in some aspects. Furthermore, the latest generation of 454 sequencing can sequence up to a mean of 746bp and a mode of 800bp. Thus, sequencing from just the forward reads can be sufficient to cover the entire heavy- and light-chain immunoglobulin amplicons, also making the 3' plate-ID non-essential, in some aspects, because assembling forward with reverse reads is no longer required.

Selection and cloning of antibodies

[00424] After assembly, heavy and light chain sequences were analyzed to select antibodies for characterization. Antibodies were selected based on predicted V(D)J germline usage and inspection of evolutionary trees derived from the antibody sequences. The selected antibodies were cloned, expressed, and characterized in different assays.

[00425] V(D)J Assignment

[00426] Heavy and light chain sequences from Runs 1 and 2 were analyzed with V-QUEST (Brochet, X. et al., *Nucl. Acids Res.* 36, W503-508 (2008)), software that compares an antibody chain sequence to a database of known alleles of germline sequences and predicts the germline alleles used in the antibody, how the germline sequences were recombined, and mutations in the antibody relative to the germline. Table 18 shows results of a V-QUEST V(D)J assignment for the antibodies which were chosen for further characterization; the same data were obtained for all other sequence assemblies from Run 1 and Run 2 as well. Some sequences from Run 3 and Run 4 assemblies were analyzed with SoDA (Volpe, Cowell and Kepler, *Bioinformatics* (2006) 22 (4): 438-444), software similar to V-QUEST.

[00427] If the patient's genome has been sequenced, the genome sequence data can be used as the germline sequences for the VDJ assignment analysis, further improving the ability to reliably identify somatic hypermutations in the patient's antibody sequences.

[00428] **Evolutionary Trees and Clonal Families**

[00429] The nucleotide sequences corresponding to the mature peptides of heavy chains from Runs 1 and 2 were separated into sets that corresponded to the patient from which they were derived. From these individual sets the software clustalx2 (Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007) *Bioinformatics*, 23, 2947-2948) was used to generate an alignment and tree using default settings for all parameters.

[00430] Evolutionary trees of sequences from a patient can also be constructed from the set of sequences from individual clonal families The putative progenitor antibody heavy and light chain sequences for the family can be inferred and added to the set of sequences if they are not already in the set. An evolutionary tree for the set can be constructed using, for example, Maximum Parsimony, Maximum Likelihood, or any suitable algorithm, and the tree can be rooted at the sequence of the progenitor antibody. The tree can be constructed on the basis of the heavy chains alone, or the light chains alone, or preferably by constructing the tree based on the individual heavy and light chains simultaneously, so that the tree represents the co-evolution of the heavy and light chains.

[00431] **Antibody Selection**

[00432] For each patient, the table of V-QUEST results was reviewed in conjunction with the trees built from Run 1 or Run 2 sequences (viewed in TreeViewX: http://darwin.zoology.gla.a.c.uk/~rpage/treeviewx). Representative sequences were selected to cover the different families of VDJ present based on the V-QUEST data, inspecting the corresponding sequences on the tree to choose sequences that appeared to be representative of the clade. Typically one sequence was selected from each clade. Some of the selected sequences came from families with many members, but some were also selected from families with few members or one member. The selected sequences are described in Table 18. For each antibody in Table 18, the column "Antibody" is the same as the text following the "-" in the Contig ID field of the name associated with the sequence in the Sequence Listing.

Cloning and expression of cloned light and heavy chain immunoglobulin pairs

[00433] Vectors

[00434] One system is a Neomycin and dihydrofolate reductase (DHFR) selectable vector system modified from Invitrogen vectors pcDNA3.3 and pOptivec. An alternative system is the Lonza GS system, in which the amplifiable, selectable marker is glutamine synthetase (GS) (*see* below). Sequences encoding the immunoglobulin kappa light chain, lambda light chain, and gamma heavy chain are inserted into vectors. Kozak consensus and leader sequences are already present in the clones, and thus do not need to be engineered into the vectors. The constant regions are synthesized to contain 5'-flanking restriction sites and one or more other internal restriction sites. For facilitating the cloning of varied immunoglobulin heavy and light chains, inserts are engineered with multiple restriction sites that increase the possibility that the clone itself will not contain the restriction site and therefore not be cut internally. The inserts have two different 8-cutter restriction sites at the 5' end of the insert region and at two different restriction sites engineered into constant regions. The 5' restriction sites are FseI and PacI for both light chains, and AscI and AsiSI for the gamma heavy chain. Restriction sites engineered into the constant region themselves are NheI and XhoI for both light chains, and EcoRI and SacII for the gamma heavy chain. See Table 16 for the sequence of the constant region inserts containing the restriction sites. Heavy or light chain clones from the 1st PCR reaction are then subjected to a 2nd round of PCR with cloning primers, which have 5' flanking restriction sites that are incorporated into the clones. Appropriate restriction enzymes are used for cutting the expression vectors and the clones, which now have complementary ends and are ligated together using T4 DNA ligase. Both the Invitrogen and Lonza GS vector systems contain an amplifiable selection marker. This marker is DHFR in the Invitrogen system and GS in the Lonza GS system. Under selection pressure from the appropriate selector (methotrexate for DHFR and L-methionine sulfoximine for the glucose synthetase (GS)), genes linked to the selection marker are amplified together with it. With more copies of the immunoglobulin genes, there is greater secretion of antibodies. This is useful when large amounts of antibody need to be purified for subsequent *in vivo* screening for neutralizing antibodies.

[00435] Cloning and expression

[00436] Assuming that the highest-affinity plasmablasts are selected for during germinal center maturation, we expect that the highest-affinity clonal family also has the highest number of clones. Furthermore, the highest-affinity clone within each clonal family will also be the most frequent clone within that family. On the basis of these assumptions, we choose to express the highest-frequency clone from the 5 largest clonal families from each patient sample in some

aspects. Clones are amplified from the 1st PCR cDNA, where all samples from the same plate, each containing a single cell, have been pooled together. The forward primer contains the sample-ID and therefore amplifies only DNA barcoded with that particular sample-ID. This is highly specific because sample-IDs contain nucleotide differences between one another. Some sample-IDs can have identical cloning forward primers and clones amplified by these primers must subsequently be distinguished from each other by bacterial colony selection too. Both forward and reverse primers contain flanking restriction sites (first and second restriction site regions) that allow the clone to integrate (with coding frame aligned) into a vector that already contains a kappa, lambda, or gamma constant region. See Tables 4 and 5 for cloning primer sequences. Light chains are cloned into modified pcDNA3.3, and heavy chains into modified pOptivec, or both chains are cloned into the Lonza GS dual-expression vector. Mammalian cells are either doubly transfected with separate expression vectors encoding the immunoglobulin heavy and light chain genes, or singly transfected with a dual-expression vector containing both the heavy and light chain genes. Supernatants containing the secreted antibodies are then collected and screened for the desired properties.

[00437] In some instances, variable regions of Ig genes may be cloned by DNA synthesis, and incorporating the synthesized DNA into the vector containing the appropriate constant region using restriction enzymes and standard molecular biology. During synthesis, the exact nucleotide sequence need not be followed as long as long as the amino acid sequence is unchanged, unless mutagenesis is desired. This allows for codon optimization that may result in higher expression levels. This also allows for adding in restriction sites for the purpose of cloning. Non-translated sequences such as 5' UTR and barcode sequences need not be synthesized and leader sequences can also be swapped for other signal peptide sequences known for higher expression levels. These result in an Ig nucleotide sequence that can be very different from the high-throughput reads but give identical amino acid sequenced when expressed.

[00438] In some instances, the sample-ID barcode adaptor added on during reverse transcription may already incorporate a restriction enzyme site. This results in an adaptor with a restriction site 3' of the sample-ID barcode in the PCR amplicon pool. During cloning with cloning primers, desired amplicons are amplified from a plate-specific amplicon pool using 5' primers that are complementary to the sample-ID barcode sequences, and chain specific 3' primers (for the kappa, lambda and gamma chains). 3' primers will add on 3' restriction sites. 5' primers do not need to add restriction sites as the 5' primer already contains a restriction site 3' of the well-ID barcode. Following this amplification, restriction enzymes are used to cut the amplicon for ligation into the vector containing the constant region insert. During the restriction

enzyme digest, sequences added on to the 5' end of the Ig gene sequences, such as barcodes and universal sequences are cut as they are 5' of the 5' restriction site.

[00439] Alternative methods for cloning and expression

[00440] In another aspect, variable regions of Ig genes may be cloned by DNA synthesis, and incorporating the synthesized DNA into the vector containing the appropriate constant region using restriction enzymes and standard molecular biology. During synthesis, the exact nucleotide sequence need not be followed as long as long as the amino acid sequence is unchanged, unless mutagenesis is desired. This allows for codon optimization that may result in higher expression levels. This also allows for adding in restriction sites for the purpose of cloning. Non-translated sequences such as 5' UTR and barcode sequences need not be synthesized, leader sequences can also be swapped for other signal peptide sequences known for higher expression levels. These result in an Ig nucleotide sequence that can be very different from the high-throughput reads but give identical amino acid sequenced when expressed.

[00441] In another aspect, the well-ID barcode adaptor added on during reverse transcription may already incorporate a restriction enzyme site. This results in an adaptor with a restriction site 3' of the well-ID barcode in the PCR amplicon pool. During cloning with cloning primers, desired amplicons are amplified from a plate-specific amplicon pool using 5' primers that are complementary to the well-ID barcode sequences, and chain specific 3' primers (for the kappa, lambda and gamma chains). 3' primers will add on 3' restriction sites. 5' primers do not need to add restriction sites as the 5' primer already contains a restriction site 3' of the well-ID barcode. Following this amplification, restriction enzymes are used to cut the amplicon for ligation into the vector containing the constant region insert. During the restriction enzyme digest, sequences added on to the 5' end of the Ig gene sequences, such as barcodes and universal sequences are removed as they are 5' of the 5' restriction site.

[00442] Cloning of heavy and light chains into Lonza vectors

[00443] *Cloning immunoglobulin constant regions*

[00444] Lonza vectors were obtained through Stanford University's academic licensing agreement with Lonza. Kappa and lambda light chains were inserted into vector pEE12.4 and the gamma heavy chain was inserted into vector pEE6.4. Heavy and light chain sequences were cloned in two steps: first, the constant regions were cloned in, followed by the 5' end of the immunoglobulin chains (the leader and V(D)J sequences). Constant regions inserts were gene-synthesized by Intergrated DNA Technologies (IDT), and contained appropriate silent mutations for gene optimization and incorporation of restriction sites. Inserts were obtained from IDT in their proprietary pIDTSmart vector. Insert sequences are in Table 17. IgG1 was used as the

gamma heavy chain constant region. The alleles used were Km3, Mcg$^-$Ke$^-$Oz$^-$ and G1m3 for kappa, lambda and gamma chains respectively. To incorporate the constant regions into the Lonza vectors, the Lonza vectors and the pIDTSmart-constant region inserts were all individually transformed into competent dam$^-$dcm$^-$ *E. coli* and plasmids were purified using Qiagen miniprep kit following manufacturer's instructions. Plasmids were then digested using HindIII and BclI at 37°C for 1 hour and run on a 1.2% agarose gel at 150V for 1 hour. Digested Lonza vectors and the constant region inserts were gel purified and ligated (pEE12.4 with Km3 or Mcg$^-$Ke$^-$Oz$^-$ light chains and pEE6.4 with G1m3 gamma1 heavy chain) using T4 DNA ligase for 10 minutes at RT in a ratio of 3:1 insert:vector. T4 DNA ligase was then inactivated at 70°C for 8 minutes and 5ul of the ligation mix was transformed into heat shock competent TOP10 cells using standard molecular biology techniques. Colonies were picked and insertion was verified via Sanger sequencing.

[00445] *Cloning immunoglobulin variable regions*

[00446] Next, pEE12.4 containing either lambda or kappa light chain was digested using AscI and XmaI for 5 hours at 37°C and gel purified. pEE6.4 containing the gamma 1 heavy chain was digested using AscI and AgeI for 5 hours at 37°C and gel purified. Selected amplicons were selectively amplified using well-ID specific forward primers and constant region-specific reverse primers (Tables 4 and 5) from the specific plate-ID 100x dilution of the 1st PCR. The forward primers had the restriction site AscI on the 5' end of the primer, and the reverse primers contained the XmaI or AgeI restriction site for the light and heavy chain primers constant region primers respectively. PCR cycling was done using an initial denaturation at 98°C for 30 seconds, and 35 to 45 cycles at 98°C for 10 seconds, 68°C for 15 seconds and 72°C for 20 seconds. The final extension was 72°C at 5 minutes and on hold at 4°C indefinitely. The PCR products were purified using PCR_{u96} ultrafiltration plates from Millipore, following manufacturer's instructions. Following this, PCR products were double digested for 3 hours at 37°C using AscI and XmaI for the light chains and AscI and AgeI for the gamma 1 heavy chain. Digested products were then run on a 2.5% low melt agarose with gelgeen (Biotium) and visualized under blue light. Gel slices containing bands at the appropriate sizes were excised. In gel ligation was performed by melting the gel slices at 65°C and adding the appropriate digested and Antarctic phosphatase (NEB)-treated Lonza vectors containing the constant region insert, and incubating with T4 DNA ligase for 1-3 hours at RT. Heat-shock competent bacteria were then transformed and plated on ampicillin agar. 6 clones per construct were picked and grown in 2xLB (2x concentrated Luria-Bertani broth). Miniprep was performed using Millipore's Multiscreen 96-well filter plates according to manufacturer's instructions

(www.millipore.com/techpublications/tecg1/tn004) to obtain plasmid DNA. Colony PCR was performed using an initial denaturation of 95°C for 5 minutes, and 40 cycles of 95°C for 1 minute, 50°C for 2 minutes, 72°C for 1 minute, and a final extension at 72°C for 5 minutes and holding t 4°C indefinitely. Clones with the appropriate insert were sent for Sanger sequencing at Sequetech, Mountain View, CA, USA. Clone VDJ identification was done using IMGT HIGHV-Quest and the correct clones were kept as both bacterial stock (stored at -80°C with 15% glycerol) and plasmids.

[00447] Expression of monoclonal antibodies in 293T

[00448] Transient dual transfections of paired pEE12.4-lightchain and pEE6.4-heavychain constructs were done using Lipofectamine 2000 following manufacturer's protocol. Transfections have been done in 48-well, 24-well, 6-well, 60mm and 100mm dishes. In brief, 293T cells were cultured in DMEM+10% ultralow IgG FBS (Invitrogen) to prevent bovine IgG from competing with secreted human IgG at the downstream protein A purification step. 293T cells were cultured for 20 passages before new aliquots liquid N_2 were thawed and used. For 48-well plate transfections, each well was seeded the day before with $8x10^4$ cells, and allowed to grow to ~90% confluency the next day. 50ng each of heavy and light chain constructs were incubated in Optimem media for a final volume of 50ul, and Lipofectamine 2000 was also separately incubated with 50ul of Optimem media. Both incubations were from 5-25 minutes. Lipofectamine 2000 and the constructs were then mixed by gentle pipetting and incubated for 20 minutes before adding to 293T cells and gently mixed. Media was changed the next day and culture supernatants were collected every other day (e.g., Monday, Wednesday, and Friday) for 2 weeks. For transfections of other sizes, the following amounts of constructs and Lipofectamine 2000 were used: for 24-well plate transfections, 100ng of each construct were used with 1.25ul of Lipofectamine 2000. For 60mm dishes, 625ng of each construct were used with 12.5ul of Lipofectamine 2000. For 100mm dish transfections, 3ug of each construct were used in 37.5ul of Lipofectamine 2000.

[00449] Anti-human IgG ELISA

[00450] In some instances, human IgG ELISA was done on the sample to quantitate the amount of expressed IgG in the culture supernatant, and the culture supernatant was used directly in downstream applications after normalizing the amount of antibody. Anti-human IgG ELISA quantitation kit was purchased from Bethyl Laboratories and performed according to manufacturer's instructions. In brief, 100ul of capture antibody was coated on Nunc Maxisorp plates overnight at 4°C and washed 5x with PBST (PBS with 0.05% Tween20). Wells were blocked with 1% BSA in PBS for an hour at RT, then washed 5x with PBST. Wells were then

incubated with the appropriate standard dilutions from the kit or diluted culture supernatants for 1 hr at RT, then washed 5x with PBST. 100ul of diluted HRP detection antibody was added to each well and incubated at RT for an hour, then washed 5x with PBST. 50ul of TMB substrate solution was added and the reaction stopped with 50ul of stop solution. Absorbance was read on a SpectraMax M5 spectrophotometer at 450nm and standard curves generated with a 4-parameter curve. Antibodies were kept at 4°C in PBS with 0.1% sodium azide as a preservative.

[00451] Protein A-IgG purification of expressed monoclonal antibodies

[00452] In other instances, antibodies were first purified from culture supernatant and quantitated using BCA before use. In brief, culture supernatants were collected 3x a week for 2 weeks into 50ml tubes and stored at 4°C with 0.1% sodium azide as an additive until protein A-IgG purification. Culture supernatants were spun down and decanted to remove any cellular aggregates. 1M pH 9.0 Tris was added to culture supernatants to ensure pH is between 7.5-8.0 as determined by pH indicator strips. Protein A plus agarose beads (Pierce) were washed 2x with PBS before 400ul of a 50% slurry was added to culture supernatants and incubated at 4°C overnight on a rotator to ensure even mixing of the beads. Beads were recovered by spinning culture supernatant at 1000g for 5 minutes and pipetting out the beads from the bottom of the tubes into 5ml gravity flow columns. Beads were washed with 4x2ml of PBS before elution with 2x1.5ml of IgG elution buffer (Pierce), which is a low pH elution buffer, into an Amicon-4 100kDa concentrator column. Eluted antibody was immediately neutralized with 400ul of 1M Tris pH8.0. Antibodies were then concentrated by spinning at 1000g in Amicon-4 100kDa concentrators for 10 minutes, followed by a 2mL wash of PBS, and a 2ml wash of PBS with 0.1% sodium azide. Antibody concentrations were determined by BCA assay, and adjusted to 0.5mg/ml. Protein A Plus Agarose was regenerated by washing with 1x2ml of PBS, 3x2ml of IgG elution buffer and a wash with 3ml of PBS with 0.1% sodium azide and kept at 4°C. Columns can be regenerated up to 5 times.

Screening of expressed human antibodies

[00453] Screens for antibody-antigen binding

[00454] Selected antibodies (see cloning and expression section above) are first screened for their ability to bind to the antigen of interest, and then antibodies of the entire clonal family are expressed and screened for their ability that block or neutralize the antigen (see "functional screens" below). The IgG concentration in supernatants containing antibodies of interest is first determined by IgG ELISA, so that the same amount of IgG can be used for each sample in the antibody-antigen binding screen. In other cases, IgG was purified from the supernatant using

Protein A agarose beads, and quantified with BCA assay using bovine immunoglobulins as the standard. Purified IgGs were then normalized to the same concentration before use in screening antibodies. To screen antibodies that bind antigen, we perform an indirect ELISA. A 96-well plate is coated overnight with the antigen of interest, and excess antigen is then washed off. Supernatants containing antibodies of interest are then added to the wells and incubated for four hours, after which the wells are washed. As a positive control, known amounts of commercially available antibodies (from non-human species) that are specific for the antigen are added to separate wells containing the antigen. As a negative control, commercially available antibodies specific for an irrelevant antigen are added to separate wells containing the antigen of interest. An HRP-conjugated secondary antibody is then added to the wells, incubated for 30 minutes, and excess washed off. Tetramethylbenzidine (TMB) is then added to the plate, and the reaction allowed to proceed until color is observed in the positive-control wells. The reaction is then stopped with acid, and absorbance is measured. Specific well supernatants are deemed to contain antibodies that bind the antigen of interest if the absorbance readout they yield is significantly higher than that in the negative control.

[00455] Fluzone ELISA

[00456] Volunteers were administered the 2010/2011 season flu vaccine from Fluzone, which consists of 3 strains of inactivated virus, the A/California/7/2009, A/Perth/16/2009, B/Brisbane/60/2008 strains. Fluzone ELISAs were done to determine if monoclonal antibodies derived from vaccinated volunteers bind to the flu vaccine itself as an initial screen for expressed antibodies with binding activity. Fluzone vaccine was diluted 100x in a pH9 carbonate buffer and coated on Nunc Maxisorp plates either at RT for 1 hour or overnight at 4°C. Plates were then washed 5x with PBST (PBS w/ 0.05% Tween20) and blocked with PBS w/ 1% BSA for 1 hour at RT. 100ul of 100ng/ml of expressed flu antibodies were then added to wells at RT for 1 hour, before washing 5x with PBST and adding diluted HRP detection antibody (from Bethyl Labs human IgG ELISA quantitation kit) for 1 hour at RT. Plates were washed 5x with PBST and 50ul of TMB substrate were added. Color was allowed to develop for up to 30min before stopping the reaction with 50ul of stop solution. Plates were read on a SpectroMax M5 spectrophotometer at 450nm absorbance. Antibodies used in this assay are described in Table 19. Sequences of antibodies can be referred to in the master table, Table 18.

[00457] Surface plasmon resonance determination of flu antibody affinities

[00458] Binding of monoclonal antibodies (mAbs) to HA molecules was analyzed at 25°C using a ProteOn surface plasmon resonance biosensor (BioRad Labs). Expressed flu monoclonal antibodies derived from the plasmablasts of a flu-vaccinated donor (25 nM in pH

4.5 acetate buffer) were coupled to a GLC sensor chip with amine coupling at a target density of 800 resonance units (RU) in the test flow cells using EDAC-NHS chemistry. Unreacted active ester groups were quenched with ethanolamine. Purified recombinant hemagglutinin H3 (HA(ΔTM)(H3N2/Perth/16/2009) and H1 (HA(ΔTM)(A/California/07/2009)(H1N1)) were purchased from Immune Technology Corp. (New York, NY) and were diluted to 100, 50, 25, 12.5, 6.25, nM, along with a blank buffer control were injected at a flow rate of 30 μL/min with 120 seconds contact time and 2000 seconds dissociation time. Binding kinetics and data analyses were performed with BioRad ProteON manager software. Affinity measurements were calculated using the bivalent analyte algorithm, as HA consisted of several repeating units. The accuracy of the fitted curves was verified by checking that the $\chi 2$ values of the goodness of each fit was below 10% of the peak binding value (Rmax). Antibodies used in this assay are described in Table 20. Sequences of antibodies can be referred to in the master table, Table 18.

[00459] RA antibody reactivities on RA antigen microarrays

[00460] To print the antigen microarrays, antigens were diluted in phosphate buffered saline to 0.2 mg/mL and attached to ArrayIt SuperEpoxy slides using an ArrayIt NanoPrint Protein LM210 system. Slides were marked with a hydrophobic marker pap pen and blocked overnight in PBS with 3% fetal bovine serum and 0.05% Tween 20 at 4°C, gently rocked at 30 rpm. Arrays were probed with 400uL of 40ug/mL monoclonal antibody for one hour at 4°C, gently rocked at 30 rpm. Arrays were then washed and incubated in Cy3 conjugated anti-human IgG/IgM secondary antibody diluted at 1:2500 for 45 minutes at 4°C, gently rocked at 30 rpm. Following another wash, slides were scanned using a GenePix 4300A Microarray Scanner. GenePix 7 software was used to find the median florescence intensity of each feature and background.

[00461] To analyze the data, background fluorescence intensities were subtracted from each feature and expressed as the median value of the four antigen features on each array. Median intensities were log transformed with a cut-off value of 10. These values were subjected to hierarchical clustering using Cluster software to arrange antigens based on similarities to each other. The relationships were displayed as a heatmap using Java TreeView software. Antibodies used in this assay are described in Table 21. Sequences of antibodies can be referred to in the master table, Table 18.

[00462] Anti-histone 2A ELISA

[00463] A direct ELISA was used for detection of antibodies to histone 2A. Microtiter plates (Nunc Maxisorp) were coated with 100 μl of recombinant H2A in carbonate buffer, at a concentration of 20 μg/ml, and incubated at 4°C overnight. After blocking in PBS containing 1%

bovine serum albumin (BSA), RA patient-derived antibodies were used in a titration from 15ug/ml to 250 ug/ml in dilution buffer (PBS containing 0.1% BSA and 0.1% Tween-20), added to the plate in duplicate at 100 µl/well, and incubated for 2 hours at room temperature. The samples were then incubated for 1 hour at room temperature with a 1:5,000 dilution of a monoclonal, horseradish peroxidase–labeled goat anti-human antibody. The reaction was developed by application of 3,3′,5,5′-tetramethylbenzidine substrate (TMB) (Sigma-Aldrich) for 15 minutes and stopped by addition of 50 µl of 2N H2SO4. Relative quantification of antibodies against was performed by optical densitometry at 450nm using known seropositive RA serum as a positive control. Antibodies used in this assay are described in Table 22. Sequences of antibodies can be referred to in the master table, Table 18.

[00464] Anti-CCP2 ELISA

[00465] Anti-CCP2 ELISA was performed according the manufacturer's instructions (Eurodiagnostica, Malmo, Sweden). Briefly, antibodies derived from RA patients are diluted to approximately 125 ug/ml in dilution buffer (PBS containing 0.1% BSA and 0.1% Tween-20), added to the pre-blocked commercial CCP2 ELISA plate at 100 µl/well, and incubated for 2 hours at room temperature. The samples were then incubated for 1 hour at room temperature with a 1:5,000 dilution of a monoclonal, horseradish peroxidase–labeled goat anti-human antibody. The reaction was developed by application of 3,3′,5,5′-tetramethylbenzidine substrate (TMB) (Sigma-Aldrich) for 15 minutes and stopped by addition of 50 µl of 2N H2SO4. Relative quantification of antibodies against was performed by optical densitometry using standards provided by the vendor and known positive RA serum. Antibodies used in this assay are described in Table 22. Sequences of antibodies can be referred to in the master table, Table 18.

[00466] Anti-Rheumatoid Factor ELISA

[00467] For detection of antibodies to rheumatoid factor (RF), microtiter plates (Nunc Maxisorp) were coated with 10ug/ml of rabbit IgG in carbonate buffer, and incubated at 4°C overnight. After blocking in PBS containing 1% bovine serum albumin (BSA), RA patient derived antibodies were at 5ug/ml dilution buffer (PBS containing 0.1% BSA and 0.1% Tween-20), added to the plate in duplicate at 100 µl/well, and incubated for 2 hours at room temperature. The samples were then incubated for 1 hour at room temperature with a 1:5,000 dilution of a monoclonal, horseradish peroxidase–labeled goat anti-human antibody. The reaction was developed by application of 3,3′,5,5′-tetramethylbenzidine substrate (TMB) (Sigma-Aldrich) for 15 minutes and stopped by addition of 50 µl of 2N H2SO4. Relative quantification of antibodies against was performed by optical densitometry at 450nm using two

known RF+ control serum as a positive controls. Antibodies used in this assay are described in Table 23. Sequences of antibodies can be referred to in the master table, Table 18.

[00468] Immunohistochemistry of antibodies from lung adenocarcinoma patient on lung cancer tissue arrays

[00469] Two different types of tissue microarray slides were purchased from US Biomax. They were VLC 12 and BS0481. A variety of lung carcinomas tissue cores are included on the slides, including lung adenocarcinomas, and also normal lung tissue controls. Slides were heated in a citrate pH6.0 antigen retrieval buffer at 95-99oC for 40 minutes before allowing to cool to RT. Slides were pre-treated with 0.02% Triton-X and 0.6% H2O2 for 20 minutes. Slides were then blocked with 10% normal goat serum in TBST (TBS with 0.05% Tween20) for 2 hours before further blocking in 100ug/ml of F(ab) goat-anti-human IgG (Jackson Immunoresearch) overnight at 4oC. Slides then underwent an avidin/biotin block (Vector Laboratories) according to manufacturer's instructions. Slides were then incubated in 5 or 10ug/ml of expressed lung antibody for 1 hour at RT, washed 3x5 minutes with TBST, and then incubated with a biotinylated goat-anti-human secondary antibody for 20 minutes at RT. Slides were then washed 3x5 minutes with TBST and incubated with prepared Vectastain ABC reagent for 30 minutes at RT. Slides were then washed 3x5 minutes of TBST and stained with Vector Red (Vector Laboratories) and the color development tracked with a light microscope. After the appropriate staining time, the reaction was stopped by washing with distilled water and counterstained with hematoxylin. Slides were aqueous-mounted and photographed with a BX-51 microscope. Antibodies used in this assay are described in Table 24. Sequences of antibodies can be referred to in the master table, Table 18.

[00470] Flow cytometry determination of binding of antibodies expressed from lung adenocarcinoma patient to lung cancer cell lines

[00471] Lung cancer cell lines used were A549, H226, H441, H23, H1975, H1437, H2126, H1650 and H2009. HEK 293T cells were also used as a negative control. Cells were detached by incubating cells in PBS without Ca^{2+} and Mg^{2+} with 2mM EDTA for 1 hour at 37°C. This is to prevent damaging cell surface antigens that may be done with trypsinizing or any other proteolytic digest to detach cells. Cells were washed once with FACS buffer (HBSS with 2% FCS) before suspended in 50ul of FACS buffer and incubated with 10ug/ml, 3ug/ml, 1ug/ml, 0.2ug/ml of expressed lung antibodies to titrate the dose. The optimal concentration was found to be from the 0.2-1ug/ml range. Therefore, 1ug/ml, 0.5ug/ml, 0.25ug/ml of lung antibodies were used subsequently. Lung antibodies were incubated for 30 minutes at 4°C before washing with 2x 200ul of FACS buffer in 96-well plates. Anti-human IgG-PE was then added and

incubated for 15 minutes at 4°C in the dark. Samples were then washed 2x200ul with FACS buffer and resuspended in 200ul of FACS buffer and analyzed on a BD LSR II or LSR Fortessa. Sytox blue was used as a live/dead staining. Antibodies used in this assay are described in Table 24. Sequences of antibodies can be referred to in the master table, Table 18.

[00472] Staph Flow Cytometry

[00473] Fixed *S. aureus* particles (Wood strain) were obtained from Invitrogen. Wood strain is a strain that expresses minimal protein A by some of the bacteria. The particles were suspended in 50ul of FACS buffer at $10x10^6$ cells/50ul, and incubated with a titration of 10ug/ml, 5ug/ml or 1ug/ml of expressed antibodies derived from staph individuals for 1 hour minutes at 4°C. Fixed staph particles were then washed with twice with FACS buffer before incubating for 15min in the dark at 4°C with anti-human IgG-FITC antibody. Particles were then washed with 1ml of FACS buffer and resuspended in 200ul of FACS buffer for analysis on a BD LSR II or LSR Fortessa. Antibodies used in this assay are described in Table 25. Sequences of antibodies can be referred to in the master table, Table 18.

[00474] **Functional screens**

[00475] Blocking antibody to receptor-ligand interactions

[00476] To screen antibodies for their ability to block ligand-receptor interactions (e.g. cytokine-receptor interactions), we transfect 293T cells with a vector encoding the appropriate receptor. These 293T cells are also stably transfected with an NF-κB-dependent luciferase reporter—such that these stably transfected 293T cells express luciferase when NF-κB is activated. We then culture the transfected 293T cells with the appropriate ligand in the presence or absence of the candidate antibodies. 293T cells are finally assayed for luciferase expression by measuring luciferase-dependent light emission. Interaction between the ligand and its receptor, e.g. interaction between IL-17A and IL-17R, activates NF-κB. Blocking antibodies prevent NF-κB signaling by ligand-receptor binding and therefore abrogate the expression of luciferase. In cases where the ligand-receptor interaction does not activate NF-κB, other transcriptional response elements are used to drive the promoter of the luciferase gene, e.g. AP-1 response elements, etc.

[00477] Screening antibodies for their ability to inhibit cytokine function or inhibit a functional assay

[00478] Functional assays can also be used to screen for anti-cytokine antibodies in patient sera or in cloned and expressed antibodies. In this approach, the expressed human antibodies are tested for their ability to inhibit cytokine or other immune mediator induction of a cellular response.

[00479] Antibodies targeting bacteria, virally-infected cells, parasites or cancer cells

[00480] To screen for antibodies that kill or neutralize bacteria, virally-infected cells, parasites, or cancer cells, we culture the appropriate cell type either in the presence or absence of antibody, together with non-heat-inactivated serum (which contains complement factors). If the antibody is a neutralizing antibody, it will opsonize the bacteria, other microbes, or cancer cells and activate complement components that form the membrane attack complex (MAC), which induces cell death. To test neutralization, we run a fluorescent live/dead assay (Invitrogen), in which live and dead cells are stained with different fluorophores. Cells can be assayed for percentages of live and dead cells by using flow cytometry. The antibody that results in the highest percentage of dead cells will be a good candidate neutralizing antibody that will be further analyzed in *in vivo* screens.

[00481] Antibodies that neutralize viruses

[00482] To screen for antibodies that neutralize viruses, we perform a standard plaque-reduction assay or other *in vitro* cellular infection assays. Neutralizing antibodies are expected to decrease viral infection of cells. Candidate antibodies are then tested in an *in vivo* model.

[00483] Influenza microneutralization assay

[00484] Some expressed flu antibodies that showed binding activity to the Fluzone ELISA were sent to an external CRO, Virapur, LLC for microneutralization assays. In brief, two-fold dilutions of each antibody, starting from 100ug/ml, were mixed with an equal volume of approximately 100 $TCID_{50}$ infectious units of titered stock virus in quadruplicates in wells of a 96-well plate. Virus/antibody solutions were incubated for 2 hours and then the mixture was transferred to a 96-well plate containing 80% confluent MDCK cells. Cells, antibody and virus were incubated for an additional 2 hours at 37°C, after which virus was removed, monolayers rinsed and viral growth media added to each well. Wells were observed microscopically after 72 hours for the presence of influenza virus infection. Antibodies used in this assay are described in Table 26. Sequences of antibodies can be referred to in the master table, Table 18.

[00485] Staph Inhibition Assay

[00486] *S. aureus* were used when in log-phase growth. They were added to 96-well polypropylene plates, and anti-staph antibody from staph patients was added at 10ug/ml. Baby rabbit complement (Cedarlane) was added at manufacturer's recommended amount and mixed thoroughly. Plates were incubated at 37°C for 45 minutes before being diluted 1:10, 1:100 and 1:1000 and plated on 5% TSA blood agar plates and grown overnight. Bacterial CFUs were counted and tabulated the next day. Antibodies used in this assay are described in Table 27. Sequences of antibodies can be referred to in the master table, Table 18.

[00487] Immunoprecipitation of staph antigens with antibodies derived from staph-infected patients

[00488] Staph protein lysate was made by lysing S. aureus using B-Per Bacterial Protein Extraction Reagent (Pierce) with 100ng/ml of lysostaphin for 30 minutes at RT along with 1x Halt protease inhibitor, and separated from the insoluble fraction by centrifuging at 15000rpm in a microcentrifuge. Lysate was precleaned by incubating with protein G Dynabeads for 1 hour at RT. 5ug of antibody derived from Staph patient was bound onto protein G Dynabeads by incubating for 1 hour at RT. Protein G-bound antibodies were then incubated with precleaned staph lysate overnight at 4°C. Beads were then washed 3x with PBST (PBS with 0.1% Tween20) and heated with 5x reducing lane sample buffer (Thermo Scientific) at 95°C for 5 minutes before running an SDS-PAGE on a 4-12% Criterion Bis-Tris gel. Proteins were visualized with RAPIDStain Reagent (Calbiochem). Antibodies used in this assay are described in Table 28. Sequences of antibodies can be referred to in the master table, Table 18.

[00489] Mass spectrometry identification of peptides

[00490] Stained protein bands of interest were cut out of the gels, immersed in 10 m*M* ammonium bicarbonate containing 10 m*M* DTT and 100 m*M* iodoacetamide, treated with 100% acetonitrile, and then digested overnight at 37°C with 0.1 mg trypsin (Sigma-Aldrich) in 10 m*M* ammonium acetate containing 10% acetonitrile. The trypsinized proteins were identified with LCMS by using the Agilent 1100 LC system and the Agilent XCT Ultra Ion Trap (Agilent Technologies, Santa Clara, CA) as previously described (Lopez-Avila V, Sharpe O, Robinson WH: Determination of ceruloplasmin in human serum by SEC-ICPMS. Anal Bioanal Chem 2006, 386:180-7.). LCMS data was scanned against the SwissProt or NCBInr databases by using the SpectrumMill software (Agilent) for the detection of peptides used to identify proteins. Antibodies used in this assay are described in Table 29.

Example 1: High-throughput sequencing of paired heavy- and light-chain sequences from individual B cells.

[00491] We developed a method of adding compound barcodes (sample-ID + plate-ID) to sequences in order to unambiguously identify which sequences originated from the same well in a plate. We used this approach to sequence paired heavy chain and light chain immunoglobulin genes from individual B cells. Individual B cells can be sorted by flow cytometry from blood, bulk peripheral blood mononuclear cells (PBMCs), bulk B cells, plasmablasts, plasma cells, memory B cells, or other B cell populations (Figure 1).

[00492] First, B cells were single-cell-sorted into 96-well PCR plates, leaving one column of wells empty, as a negative control. Oligonucleotides containing different sample-ID barcodes

were added into different wells during reverse transcription (RT). After reverse transcribing the mRNA, the MMLV H$^-$ reverse transcriptase switches templates and transcribes the oligonucleotide, incorporating it and the sample-ID barcode into the 3' end of the 1st strand cDNA (Fig 2a). All cDNAs (barcoded with a sample-ID) from one plate were then pooled and subjected to two rounds of PCR. During the PCR, 454 sequencing primers (first and second sequencing regions) and plate-ID barcodes were added onto the 5' and 3' ends of the amplicon by using PCR primers with 5'-flanking barcode sequences. Amplicons (amplicon regions) from different plates now have n different plate-IDs, and the compound barcode comprising a plate-ID and a sample-ID unambiguously identifies sequences as coming from a particular cell, allowing pairing up of sequenced heavy- and light-chain genes (Fig 2b-c).

[00493] Figure 3 describes the general methodology used and the associated sequences. Primers and adapter molecules are shown in Table 1. Sample-ID sequences are shown in Table 2. Plate-ID sequences are shown in Table 3. Cloning primers are shown in Table 4, with the 3' sequence of cloning forward primers shown in Table 5.

[00494] We obtained PCR products at the expected sizes: ~600bp for the kappa and lambda light chains, and ~700bp for the gamma heavy chain (Figure 4a). Next, we sent the material for Sanger sequencing. We obtained sequences that were identified by NCBI BLAST as kappa, lambda, and gamma chains (data not shown). Further investigation of the DNA chromatogram showed a mix of several peaks starting at the sample-ID barcodes, showing that we successfully added sample-ID barcodes to cDNA from cells in different wells during RT and successfully amplified them in two subsequent rounds of PCR (Figure 4b). The Sanger sequencing chain from the 3' end also showed a mix of several peaks starting at the VJ junction, owing to amplification of genes from different cells, which differed after the VJ junction as a result of insertions and deletions and random recombination of different V and J genes. Furthermore, when we performed PCR with cloning primers specific for the well A1 sample-ID, we obtained a single peak rather than a mix of several peaks, showing that we indeed can amplify sequences from a specific cell within the pool (Figure 4c-d).

Example 2: Gating scheme for single cell sorting of plasmablasts.

[00495] Plasmablasts were defined as CD19$^+$CD20$^-$CD27$^+$CD38^{++} for this experiment. Figure 22 shows a gating scheme for flow cytometry sorting of single plasmablast cells into 96-well plates.

[00496] Single PBMCs were prepared and stained as described above. Cells were first gated on based on their FSC and SSC profile (data not shown). Live CD19$^+$ B cells were then gated on (left panel), and further narrowed down to CD20$^-$ B cells (2nd panel from left), and refined to

$CD27^{+}CD38^{++}$ cells. From this, IgG^{+} plasmablasts were determined as IgA^{-} and IgM^{-}, as IgG^{+} plasmablasts do not express cell surface IgG. This population was single cell sorted into 96-well plates.

Example 3: Plasmablasts are present in subjects undergoing immunological challenge.

[00497] Plasmablasts generally represent about 0.15% of B cells in healthy donors, but can range from about 3.3%-16.4% in subjects undergoing a variety of immunological challenges including infections (e.g., S. aureus and C. diff infections), cancer associated with non-progression (e.g., metastatic melanoma and metastatic adenocarcinoma of the lung in which patients following an intervention (chemotherapy in the case of the lung adenocarcinoma patient and ipilimumab therapy in the case of the metastatic melanoma patient) became long-term non-progressors associated with an active B cell response), and vaccinations (e.g., influenza).

[00498] Figure 23 shows that plasmablasts were present in and obtainable from a range of subjects for high-throughput sequencing of the paired antibody repertoire and characterization of the active humoral response antibody repertoire. This demonstrates that the methodologies disclosed herein can be used to obtain evolutionary trees of heavy and light (H&L) chains and use this information to clone and/or express antibodies for, e.g.: a) novel antigen discovery; b) to inform vaccine design – for example, using the immune system to inform us as to which are the known and novel antigens are likely useful for opsonization and phagocytosis and/or killing/inhibition of a pathogen or target of interest and, optionally, to put that into vaccine design; c) making neutralizing monoclonal antibodies, e.g., from vaccines; d) making binding monoclonal antibodies; e) making antibodies against microbial pathogens; and f) making antibodies against cancers. Examples of these are described in more detail below.

Example 4: Disease activity in CCP+ RA is correlated with circulating plasmablasts.

[00499] Having shown that our method can be used for sequencing of immunoglobulin genes in identifiable pairs, we used our method to investigate the antibody repertoire of plasmablasts in CCP+ RA patients . We obtained blood samples from consented RA patients and stained for plasmablasts by flow cytometry (Figure 5a). Circulating plasmablasts were expressed as a percentage of total PBMCs. We found that CCP+ RA patients have significantly higher peripheral blood plasmablast percentages than CCP- RA patients (Figure 5b). Furthermore, plasmablast percentages in CCP+ patients, but not in CCP- patients, correlated with disease activity ($r = 0.35$ and $p = 0.028$) (Figure 5c).

Example 5: Plasmablasts produced anti-CCP antibodies.

[00500] Although CCP+ patients have plasmablast percentages that correlate with disease activity, these patients could have an ongoing infection or other factors that elevated their circulating plasmablast percentages. To determine the specificity of circulating plasmablasts in CCP+ patients, RosetteSep-enriched B cells from patients were cultured in RPMI supplemented with 10% FBS. Other media supplements, such as anti-IgM, IL-6, BAFF etc., were not used so that plasmablasts would be the only cells secreting antibodies (with other B cells remaining inactivate). To confirm that only plasmablasts produce antibodies, we depleted some of the samples of plasmablasts (Figure 5d). B cells were then cultured for seven days before collecting the supernatant and running it on a Luminex peptide array. The array assays antibody reactivity to citrullinated peptides. Antibody reactivity was absent in supernatants of plasmablast-depleted samples compared to supernatants of mock-depleted B cells, suggesting that plasmablasts secrete significant amounts of anti-citrulline peptide autoantibodies (Figure 5e). Furthermore, when peptides with a mean fluorescent intensity (MFI) above 60 for each sample were counted, a strong correlation was found between circulating plasmablast percentages and the number of peptides to which antibodies reacted ($r = 0.90$ and $p = 0.0139$). An MFI of 60 was chosen as this was the threshold below which >99% of peptide reactivity falls in supernatants of plasmablast-depleted samples.

Example 6: 454 sequencing and analysis of sequences.

[00501] Plasmablasts from patients were single-cell sorted into 96-well plates as described above, and their RNA reverse transcribed and PCR amplified according to "Touchdown PCR" in materials and methods section such that they contained sample-ID (sample identification region) and plate-ID (plate identification region) barcodes, as described above. *See* Figure 3. Sequences of the cDNA were then obtained from a 454-sequencing facility (DNA Sequencing Center, Brigham Young University and 454 sequencing center, Roche).

[00502] Sequences were obtained from a first 454 sequencing run using the shotgun pipeline. Sequences of acceptable quality were obtained from a 2nd 454 sequencing run through a modified amplicon filter of the 454 GS FLX data analysis suite. The amplicon filter was modified to have <vfScanAllFlows> set to "false", and <vfBadFlowThreshold> changed to "6". Sequences from a third and fourth run were obtained using a standard 454 amplicon filter. Filter-passed sequences were then processed as described in "Assignment of sequences to wells" section in the materials and methods, and sequences in each well were individually assembled as described in "Assembly of sequences" in the materials and methods. Assembled sequences were then parsed with IMGT HighV-Quest to obtain identification of VDJ regions used.

[00503] After sequence assembly and V(D)J usage identification, ClustalX was used to cluster sequences into clonal families (Figure 6). Alternatively, sequences can be assembled using both forward and reverse reads. In this case, subdivided forward sequences are first assembled as above. Forward and reverse sequences V(D)J usage are then identified using HighV-QUEST and forward and reverse sequences subset according to plate-ID and V(D)J usage, and forward and reverse sequences assembled using Newbler. Because immunoglobulin sequences are largely similar, assembly from a smaller subset of sequences avoids potential problems of sequences from different cells being incorrectly paired.

Example 7: Clustering of sequences into evolutionary trees

[00504] Peripheral blood mononuclear cells (PBMCs) were isolated from human subjects with the indicated diagnoses or after vaccination. Plasmablasts were single-cell sorted into individual wells in 96-well plates, creating single-cell samples in each well, the mRNA in each well was then reverse transcribed, and then well contents were pooled and subjected to two rounds of PCR to amplify the immunoglobulin heavy and light chain cDNAs. The reverse transcription added an identifying sample-ID to all cDNAs generated from each single sample, and the first round and second rounds of PCR added plate-IDs and then 454 Titanium Primers A and B to every amplicon, respectively, as described in "Touchdown PCR and non-touchdown PCR" in the materials and methods section. The general methodology is outlined in Figure 3. Pooled amplicons were sequenced with 454 sequencing technology, reads of acceptable quality obtained as described above. Reads were assigned to wells and assembled as described in "Assignment of reads to well" and "Assembly of sequences" sections in the materials and methods. V(D)J segments in assembled sequences were then identified using HighV-QUEST. Identified heavy and light chain sequences with shared compound barcodes can then paired simply by putting assembled sequences with matching compound barcodes together.

[00505] Amplicons from individual human subjects were clustered based on these V(D)J segments, and sequences expressing the same V(D)J segments were classified as being from the same clonal family (Figure 7). Each pie chart represents the percentage of clones derived from individual plasmablasts from an individual human subject expressing identical V(D)J gene segments (*i.e.,* percentage of clones in each clonal family). Human subjects included those with sepsis (2 subjects), rheumatoid arthritis (3 subjects), lung cancer (1 subject), and after vaccination for influenza (1 subject). These subjects were chosen to show that clonal families can be isolated from plasmablasts of subjects undergoing both acute (sepsis and flu vaccine) and chronic conditions (rheumatoid arthritis and lung cancer).

[00506] The immunoglobulin heavy chain V(D)J sequences from the individual human subjects of Figure 7 were clustered using ClustalX and displayed using Treeview as unrooted radial trees (Figure 8). Each radial tree represents the heavy chain sequences derived from an individual human subject. For each radial tree, the terminal ends represent a unique sequence. The major branches represent clonal families, and the smaller branches represent clonal subfamilies that differ from one another by mutations that arose via junctional diversity (addition of P-nucleotides or N-nucleotides, or nucleotide deletion), somatic hypermutation and affinity maturation.

Example 8: Cloning, expression, and purification of antibodies.

[00507] All selected antibodies were cloned, expressed, and isolated as described above in the Materials and Methods section (*See* sections: Cloning of heavy and light chains into Lonza vectors; Expression of monoclonal antibodies in 293T; Anti-human IgG ELISA; and Protein A-IgG purification of expressed monoclonal antibodies). Purified antibodies were then used for further study, as discussed below.

Example 9: Characterization of antibodies from subjects following influenza vaccination.

[00508] Antibodies from humans administered an influenza vaccine were selected and isolated as described above. The antibodies selected for further characterization below are indicated in the appropriate sections.

Fluzone ELISA

[00509] Volunteers were administered the 2010/2011 season flu vaccine from Fluzone, which consists of 3 strains of inactivated virus, the A/California/7/2009, A/Perth/16/2009, B/Brisbane/60/2008 strains. Fluzone ELISAs were performed as described above to determine if monoclonal antibodies derived from vaccinated volunteers bind to the flu vaccine itself as an initial screen for expressed antibodies with binding activity. 14 of 31 antibodies bound to Fluzone ELISA (Figure 26) and a subset of these were subsequently selected and tested for binding activity to hemagglutinin using surface plasmon resonance. Antibodies characterized by Fluzone ELISA were: Flu14 - Flu23, Flu25 - Flu27, Flu29, Flu30, Flu34, Flu35, Flu37, Flu39 - Flu41, Flu43 - Flu46. S1 and S2 were used as negative controls.

Surface plasmon resonance determination of flu antibody affinities

[00510] Binding of monoclonal antibodies (mAbs) to HA molecules was analyzed at 25°C using a ProteOn surface plasmon resonance biosensor (BioRad Labs) as described above. Of the 14 antibodies that bound to Fluzone ELISA, 10 bound to H3, and 1 bound to H1, while 3 did not

bind (Figure 27). One of the non-binders, the H1 binder, and 4 other randomly chosen H3 binders were selected and sent to a contract research organization (CRO) to test neutralization activity in a microneutralization assay. Antibodies characterized by SPR were: Flu14 - Flu22, Flu26, Flu29, Flu34, Flu35, Flu46.

Influenza microneutralization assay

[00511] Some of the expressed flu antibodies that showed binding activity in the Fluzone ELISA were sent to an external CRO, Virapur LLC, for microneutralization assays as described above. The results of the assays showed that the antibody that bound to H1 in previous assays neutralized H1, while the antibodies that bound to H3 in previous assays neutralized H3. The non-binder did not neutralize influenza virus (Figure 28). Antibodies characterized by microneutralization assay were Flu15, Flu16, Flu18, Flu19, Flu20, Flu21.

CDR Variation

[00512] Flu antibodies were obtained as described above. Figure 25 shows a partial dendrogram blown-up for clarity (a). Clonal families are clearly visible and the shaded clonal family has the assigned V(D)J as shown in the grey box. Amino acid sequence across the CDRs (boxed region) for the heavy and light chains are shown in Figure 25 (b) and (c) respectively, showing some residue differences between the chains.

[00513] These above results demonstrate that evolutionary trees can be obtained using the compositions and methods described herein. Fully human monoclonal antibodies can be isolated from activated B cells, such as plasmablasts, of subjects undergoing acute conditions using the compositions and methods described herein. These fully human monoclonal antibodies can also be neutralizing antibodies using the compositions and methods described herein. The results also demonstrate that the compositions and methods disclosed herein can be used to isolate mAbs targeted against foreign antigens.

Example 10: Characterization of antibodies from subjects with RA.

[00514] Antibodies from humans suffering from rheumatoid arthritis (RA) were selected and isolated as described above. The antibodies selected for further characterization below are indicated in the appropriate sections.

RA antibody reactivities on RA antigen microarrays

[00515] Antibodies derived from RA patients were probed on an RA antigen array and florescence scanned with a GenePix machine as described in "RA antibody reactivities on RA antigen arrays" section in materials and methods. The identified relationships were displayed as

a heatmap using Java TreeView software (Figure 37). The antibodies characterized by this assay were: RA1, RA2, RA3, RA4, RA8 - RA13, RA16, RA19, RA22 and RA23. Flu14 and Flu26 were used as negative controls.

Anti-histone 2A ELISA

[00516] For detection of antibodies to H2A, a direct ELISA was performed as described in "Anti-histone 2A ELISA" in materials and methods section. Figure 35a shows the absorbance values detected for each antibody tested. The antibodies characterized in Figure 35a and in anti-CCP2 ELISA below were: RA1, RA2, RA4-RA16, RA19, RA23-RA24. Figure 36 shows the selected antibodies (RA1, RA2, RA8, RA9) on another independent ELISA using 30ug/ml of antibodies.

Anti-CCP2 ELISA

[00517] Anti-CCP2 ELISA was performed as described in "Anti-CCP2 ELISA" in materials and methods section. Figure 35b shows the absorbance values detected for each antibody tested.

Anti-Rheumatoid Factor ELISA

[00518] Antibodies derived from RA patients were used as the primary antibody in a direct ELISA and anti-human IgG-HRP was used as the secondary antibody, and visualized with TMB substrate. For detection of antibodies to rheumatoid factor (RF), Anti-RF ELISA was performed as described in "anti-rheumatoid factor" ELISA in the materials and methods section. Figure 34 shows that antibodies RA2 and RA3 showed reactivity. Antibodies characterized here were: RA1-RA6, RA8-RA12, RA14.

[00519] These above results demonstrate that antibodies can be isolated from activated B cells, such as plasmablasts, of subjects undergoing chronic conditions using the compositions and methods described herein. The results also demonstrate that the compositions and methods disclosed herein can be used to isolate mAbs targeted against self antigens.

Example 11: Characterization of antibodies from subjects with lung cancer.

[00520] Antibodies from a long-term non-progressor human suffering from metastatic lung adenocarcinoma were selected and isolated as described above. This human developed metastatic lung adenocarcinoma and was expected to succumb to cancer, however following chemotherapy this patient entered a state of long-term non-progression for over 4 years that was associated with plasmablasts constituting 3.1% of all peripheral blood B cells. The elevated peripheral blood plasmablast levels in this patient indicated that an ongoing immune response could be contributing to her long-term non-progression. The following antibodies were selected

for further characterization below: LC1, LC5-LC7, LC9-LC18. Flu16 was used as the negative control.

Immunohistochemistry of antibodies from lung adenocarcinoma patient on lung cancer tissue arrays

[00521] Immunohistochemistry using two different types of tissue microarray slides was performed as described in "Immunohistochemistry of antibodies from lung adenocarcinoma patient on lung cancer tissue arrays" in the materials and methods section. Our results demonstrated one of the expressed antibody bound to lung adenocarcinoma (Figure 32).

Flow cytometry determination of binding of antibodies expressed from lung adenocarcinoma patient to lung adenocarcinoma cell lines

[00522] Binding of antibodies to various lung cancer cell lines was performed as described in "Flow cytometry determination of binding of antibodies expressed from lung adenocarcinoma patient to lung cancer cell lines" in the materials and methods section. Our results showed that one antibody bound to lung adenocarcinoma cell lines and may be specific for lung adenocarcinomas (Figure 33).

[00523] These above results demonstrate that antibodies can be isolated from activated B cells, such as plasmablasts, of subjects undergoing chronic conditions such as cancer using the compositions and methods described herein. The results also demonstrate that the compositions and methods disclosed herein can be used to isolate mAbs targeted against self antigens.

Example 12: Characterization of antibodies from subjects with *Staphylococcus aureus* infection.

[00524] Humans with *S. aureus* infections, including a human with chronic *S. aureus* osteomyelitis with immune-mediated control of the infection in the absence of antibiotics, were used as sources for peripheral blood from which peripheral blood plasmablasts were stained and sorted. cDNA processing with barcoding, 454 sequencing, and bioinformatics analysis generated evolutionary trees of antibody repertoires in humans mounting effective immune responses against *Staph. aureus*. Antibodies from humans mounting effective immune responses against *S. aureus* infection were selected and isolated as described above. The antibodies selected for further characterization below are indicated in the appropriate sections.

Staph Flow Cytometry

[00525] Anti-staph antibodies were used to stain fixed *S. aureus* as described in "Staph Flow Cytometry" in the materials and methods. Our results showed that antibodies S6 and S11 bind to the surface of *S. aureus* and may be candidates for opsonization, resulting in phagocytosis and

killing/inhibition of *S. aureus* (Figure 29). The antibodies characterized in this assay were: S1 - S4, S6 - S13, with F26 as a negative control.

Staph Inhibition Assay

[00526] *S. aureus* in log-phase growth were combined with anti-staph antibody to determined the inhibitory activity of the antibodies as described in "Staph Inhibition Assay" in the materials and methods. Our results demonstrate that several of the antibodies cloned and expressed exhibited potent killing/inhibition activity on *S. aureus* (Figure 30). Antibodies characterized by this assay were S6 and S9, with LC1 as the negative control.

Immunoprecipitation of staph antigens with antibodies derived from staph-infected patients

[00527] Antibodies were used to immunoprecipitate various candidate staph antigens as described in "Immunoprecipitation of staph antigens with antibodies derived from staph-infected patients" of the materials and methods. Immunoprecipitated proteins were then identified with mass spectrometry as described below. Antibodies characterized by this assay were S1 - S13.

Mass spectrometry identification of peptides

[00528] Stained protein bands of interest were selected and subjected to mass spectrometry as described in "Mass spectrometry identification of peptides" of the materials and methods. The results identified either phenol-soluble modulin alpha 1 peptide or delta-hemolysin as the likely binding target for antibody S4. This demonstrates that the methods disclosed herein can be used to perform novel antigen discovery (Figure 31). Antibody characterized by this assay was S4.

[00529] These above results demonstrate that antibodies can be isolated from activated B cells, such as plasmablasts of subjects undergoing acute conditions such as a bacterial infection using the compositions and methods described herein. The results also demonstrate that the compositions and methods disclosed herein can be used to isolate mAbs targeted against foreign antigens and to determine the identity of antigens bound by selected antibodies.

Example 13: Blasting Cells and Plasmablast characterization.

[00530] Immunoglobulin sequences from B cells that are activated by an ongoing immune response can be used to produce an evolutionary tree of an ongoing immune response, as described above. This evolutionary tree is typically characterized by multiple clonal families representing activated B cells from multiple lines of descent. Sequences from naive B cells will not generally be able to be used to produce such an evolutionary tree, as they have not been activated and therefore provide little to no information on the active, ongoing immune response.

Activated B cells first become blasting cells, which are activated and are larger in size. These blasting cells then go on to become either memory B cells or plasma cells. In humans, although memory B cells and plasma cells result from an immune response, they join large pools of memory B cells and plasma cells that have resulted from responses to previous immunological insults, making it difficult to distinguish memory B cells and plasma cells against recent or previous immune responses. Therefore, in humans, blasting cells are a preferred candidate for sequencing to obtain evolutionary trees of an ongoing immune response. In research animals bred in controlled conditions (e.g., mice) however, blasting B cells, memory B cells, and plasma cells are all candidates for sequencing to obtain evolutionary trees as they are bred in a clean environment, making it possible for the majority of memory B cells and plasma cells after a rigorous immune response to be against the insult, especially after booster shots, as they should not have large memory or plasma cell populations that have seen any major immunological challenge before.

[00531] Similarly for T cells, in humans, the preferred cells to sequence to obtain an evolutionary tree of an ongoing immune response will be blasting T cells. For mice, activated, blasting, and memory T cells are all preferred candidates to sequence to obtain an evolutionary tree.

[00532] Blasting B cells are known to be larger than typical B cells. The size of a small lymphocyte, of which a resting B cell is one, is typically between 6-8 μm in size. Blasting lymphocytes (T and B cells) are typically between 8-14 μm in size. (See Figure 41, also Tzur et al, PLoS ONE, 6(1): e16053. doi:10.1371/journal.pone.0016053, 2011;Teague et al, Cytometry, 14:7 2005). Plasmablasts can have the following expression pattern: $CD19^{low/+}$, $CD20^{low/-}$, $CD27^{+}$ and $CD38^{high}$. Although use of all of these markers results in the purest population for single cell sorting, not all of the above markers need to be used to isolate plasmablasts.

[00533] As exemplified in Figure 39, plasmablasts can be gated on by using an FSC^{hi} for larger cells, resulting in a 37% pure plasmablast population. Gating on $FSC^{hi}CD19^{hi}$ cells gives 72% plasmablast purity. Gating on FSC^{hi} and $CD27^{+}$, $CD38^{hi}$, or $CD20^{-}$ gives 44, 80, and 71 percent plasmablast purity, respectively. Combination of any of these markers or other markers found to be able to distinguish plasmablasts from other B cells can be used to increase the purity of sorted plasmablasts, however any one of these markers alone can distinguish plasmablasts from other B cells, albeit with a lower purity.

Example 14: Alternative platform for sequencing and analysis.

[00534] Heavy chain reads from a single plate, plate 44, was prepared for PacBio sequencing run using the methods described in "Touchdown PCR" in the materials and methods to amplify

gamma heavy chain cDNA. 48 2nd PCRs were done to obtain sufficient DNA for PacBio run. Pooling and cleanup of DNA was done as described in "Preparing for PacBio sequencing run". DNA was sent to PacBio for prep and sequencing. CCS reads were obtained from PacBio and assigned to wells and assembled according to "Assignment of sequences to wells" and "Assembly of sequences" in the materials and methods. Results of the assignment is in Figure 38. This shows that our methods and compositions are not platform-specific for high-throughput sequencing.

Example 15: Sequencing and analysis on 454 XL+ runs

[00535] Sequencing can be adapted to 454 XL+ runs by following the method described in "Preparing for 454 XL+ sequencing run". This needs to be done as 454 XL+ runs currently only support Lib-L chemistry, while our 454 XLR70 runs utilize the Lib-A chemistry. This can also generally be adapted to situations where Lib-L chemistry is preferred to the typical Lib-A chemistry for amplicon sequencing on XLR70 runs. Reads from 454 XL+ runs can still be assigned to wells and assembled following the methods described in "Assignment of sequences to wells" and "Assembly of sequences". Reads from XLR70 and XL+ runs after 454 filtering can be used in identical fashion, i.e. downstream selection of antibodies for cloning and expression and assaying of antibody functional properties can still proceed as per Figures 6 and 9.

Example 16: Cloning of paired immunoglobulin genes.

[00536] Assuming that each clonal family recognizes the same epitope, and that sequence variance within each family is due to somatic hypermutation, we can first clone and express the most frequent clone of each clonal family for screening of antibodies that bind the antigen of interest (Figure 6). We use the most frequent clone because, during affinity maturation and selection in the germinal center, centrocytes that bind antigen with the highest affinity out compete other centrocytes for survival factors. Therefore, we expect the highest frequency clone to also have the highest binding affinity. Once a clone has been identified as an antibody capable of binding antigen, representative paired immunoglobulin sequences from the entire clonal family are then cloned, expressed, and screened for being neutralizing antibodies (Figure 6). This process may involve cloning and expression of sequences representing multiple sub-clones within the clonal family, or encoding antibodies of different isotypes within the clonal family, to enable direct testing and comparison of the binding and functional properties of specific clones representing the spectrum of antibodies contained within the clonal family. The specific clone exhibiting the desired binding and functional properties are then selected for further characterization and consideration for development as a therapeutic human antibody.

[00537] An alternative approach to selecting candidates for cloning from a family (or any other set of antibodies of interest) is to build a phylogenetic tree for the antibodies (rooted at the germline sequence in the case of a clonal family). Leaf nodes in such a tree correspond to the antibody family members. Candidates for cloning are then selected by descending from the top of the tree, always choosing the branch with the largest number of leaf nodes underneath (choosing randomly in case of a tie), then at the last node above the leaves choosing the leaf with the largest number of mutations, or choosing randomly in the event of a tie. Additional candidates could then be selected, if desired, by repeatedly selecting candidates until the desired number is achieved, as follows. For every node in the tree, if none of the leaves that are descendants of the node have been selected, count the number of leaves that are descendants. For the node with the largest such count (choosing randomly in case of a tie), descend, always choosing the branch with the largest number of descendant leaf nodes (choosing randomly in case of a tie between branches). Then, at the last node above the leaves, choose the leaf with the largest number of mutations or choose randomly in case of a tie.

[00538] Yet another approach to selecting candidates from a family of antibodies (or any other set of antibodies of interest) is to list the antibodies by descending number of non-silent mutations relative to germline and select from the list in order, thereby choosing the antibodies that are most evolved.

Example 17: Permanent transfection and expression of candidate human antibodies

[00539] Desired clones are selectively amplified from a pooled plate of sequences by using cloning primers specific to a given sample-ID; these primers also incorporate different 5' and 3' restriction sites into the clone. The restriction sites are then used for inserting the clone into vectors. Because the amplified clones may contain only a partial constant region sequence, vectors already contain either the kappa, lambda or gamma constant regions with the appropriate restriction sites needed for inserting the amplified clones in the open reading frame. Multiple restriction sites are engineered into the vector, because clones have variable sequences, to avoid the potential problem of the restriction site existing also in the clone itself, which would then also be cut by the restriction enzyme. This allows as many clones to be inserted as possible. Vectors used are either two separate vectors with different mammalian selectable markers (modified Invitrogen pcDNA3.3 or pOptivec vectors that contain constant-region gene with engineered restriction sites) or a dual-expression vector containing both the genes (Lonza GS system; pEE6.4 and pEE12.4). See Tables 16 and 17 respectively for sequences of the constant region inserts. Selection markers are amplifiable, such as dihydrofolate reductase (DHFR) in pOptivec or glutamine synthetase (GS) in Lonza GS system, to allow for gene amplification and

efficient production of antibodies for further screening purposes requiring large amounts of antibody (e.g. an *in vivo* screen). Mammalian cells are either transfected using a double transfection, with a light chain in one and the heavy chain in the other vector (modified pOptivec and pcDNA3.3), or a dual-expression vector (Lonza GS system) containing both genes.

[00540] *Modified Invitrogen Vectors.* Vectors are two separate vectors with different mammalian selectable markers and engineered restriction sites). pcDNA3.3 has a Neomycin resistance gene as a selectable marker, and pOptivec has a DHFR-selectable marker. CHO DHFR - cells are co-transfected with modified pcDNA3.3 and pOptivec under selection from Geneticin. Only DHFR - cells transfected with pOptivec, which contains a copy of DHFR, will survive, and the Neomycin resistance gene in pcDNA3.3 confers resistance to Geneticin, This allows for selection of cells that are successfully transfected with both vectors (containing the light chain in one vector and heavy chain in the other vector), and therefore will produce functional immunoglobulins.

[00541] *Lonza GS System.* Lonza GS system utilizes the vectors pEE12.4 and pEE6.4. Vector pEE12.4 contains the GS gene as the amplifiable selection marker, and pEE6.4 is a supplementary vector. The light chain will be cloned into one of the vectors and the heavy chain into the other vector. Thereafter, both vectors are cut with restriction enzymes and ligated together to form a single vector that can express both heavy and light chain genes on separate promoters. Therefore, is a dual-expression vector system, , allowing for expression of both genes from a single vector.CHO cells are transfected with the dual-expression vector under the selection of methionine sulfoximine. Transfected cells are thus selected for.

[00542] *Gene amplification*

[00543] Both dihydrofolate reductase (DHFR) and GS are amplifiable selection markers. Under selection pressure from increasing amounts of methotrexate and methionine sulfoximine respective, transfected cell lines that have duplicated genomic regions containing the DHFR and GS genes will survive because they are more resistant to the selection reagents. Genes near the selection markers, such as the inserted heavy- and light-chain immunoglobulin genes are also amplified, resulting in higher gene copies and greater production rates of immunoglobulins. Clones producing antibodies that have been found to have neutralizing properties in the *in vitro screens* (see below) are amplified so that more antibodies can be obtained for subsequent *in vivo* studies.

Example 18: Identifying the specificity of the expressed human antibodies

[00544] Antibody screening occurs in two stages. We are utilizing a novel 'selective screening' process, in which we first select appropriate clonal families to be used in the screen

for neutralizing antibodies. We screen the most frequent 1-3 clones of each clonal family for its ability to bind to the antigen. Our screen typically takes the format of an indirect ELISA, although flow cytometry may be used to identify cell-binding antibodies. This comprises first binding the appropriate antigen to an ELISA plate, then incubating it with supernatants containing the expressed antibodies. Antibodies that bind to the antigen are detected by a specific secondary antibody.

[00545] Once binding antibodies have been identified, the entire clonal family of that clone is cloned and expressed in the screening stage of the 'selection screen'. Although all antibodies in a clonal family are expected to bind to the same epitope, they may differ slightly in avidity of antigen binding and in their positioning over the antigen, differences that may affect the binding properties and/or neutralization ability of the antibodies; thus, in most cases, several different antibodies (possessing minor differences in their CDR3 regions) are expressed and screened for binding and neutralizing properties.

[00546] For neutralizing antibodies that target a specific ligand/receptor pair, 293T cells are first stably transfected with a signaling pathway reporter construct, such as a plasmid containing the luciferase gene linked to NF-kB transcription response elements. Activation of NF-kB in the transfected cell induces the expression of luciferase, whose levels can be determined in a luciferase assay. This measures NF-kB signaling activated by ligand-receptor binding. NF-kB is the signaling element of choice because most signaling events activate NF-kB. For assaying other signaling pathways, the luciferase gene promoter region contains the appropriate transcriptional binding site, such as that for AP-1, for example. Next, the 293T cells are transfected with the target receptor. 293T cells are then incubated with the ligand and binding antibodies of interest in 96-well plates. After 24 or 48 hours, a luciferase assay is done to determine expression of luciferase gene. Wells with neutralizing antibodies have minimal to no luciferase expression. Results are verified by Western blotting for phosphorylated signaling proteins in the NF-kB signaling pathway. A neutralizing antibody prevents ligand-receptor signaling; and consequently abrogates phosphorylation of signaling proteins.

[00547] For antigens present on live cells, such as cancer antigens and bacterial antigens, the *in vitro* neutralization assay takes the form of an assay that detects live/dead cells, and can be done in a high-throughput format. Cancer cells or bacteria are incubated in 96-well plates with a candidate antibody. A stain that distinguishes live from dead cells and is compatible with flow cytometry can then be applied to each well. Live and dead cells are stained with different fluorophores and screened using flow to give percentages of live and dead cells. Antibodies that pass the *in vitro* screen will then be screened *in vivo* for their neutralizing activity.

[00548] Virus *in vitro* neutralization assay may be conducted using a standard plaque neutralization assay. By doing plaque neutralization assays in 96-well plates, each well can be imaged using a microscope and plaque counting can be automated with image-analysis software. Neutralizing antibodies reduce plaque formation. These antibodies are then further screened *in vivo* for neutralizing activity.

[00549] See example 9 section "Fluzone ELISA", example 11 section "Flow cytometry ..." and example 12 section "Staph Flow Cytometry" for successful assays of binding activity using ELISA (example 7) and flow cytometry (examples 9 and 10). See example 9 section "Influenza microneutralization assay" for a successful assay of antibodies with neutralizing activity.

Example 19: Sequencing of B cells with more than one cell per well

[00550] Individual samples having multiple B cells are separately reverse transcribed in containers. Reverse transcription adds a sample-ID and a 5' universal primer region to all 1st strand cDNA. cDNA from all containers of a set of containers are pooled and undergo 2 rounds of PCR. Steps are as described in "Touchdown PCR and non-touchdown PCR" , "Preparing for 454 XLR70 sequencing run" in the materials and methods. Sequences for primers are also shown in Fig. 9. Note that regardless of which gene is amplified, the forward primers remain constant (b). After RT and 2 PCRs, amplicons from all container sets are pooled and 454-sequenced. Assignment to wells and assembly of sequences follow the protocol as described in "Assignment of sequences to wells" and "assembly of sequences" in the materials and methods. The combination of plate-IDs and sample-IDs allows for identification of sequences that originate from the same sample.

[00551] Even though there are multiple cells in a well, we can pair individual heavy chains with light chains. The heavy chains from B cells derived from a common progenitor will be clonally related, as will the light chains. Therefore, we can associate a heavy chain clonal family to a light chain clonal family by observing the correlation across wells. Once an association is established between the heavy chains of a clonal family and the light chains of a clonal family, pairs are assigned in each well by selecting the heavy chain that is a member of the heavy chain clonal family and a light chain that is a member of the light chain clonal family. The selection of the pair is unambiguous when only a single instance of the heavy chain family and a single instance of the light chain family is present in a well. After determining which heavy and light chains are associated with one another, evolutionary trees may be drawn and antibodies selected for downstream characterization of their functional properties.

Example 20: Sequencing of B cells with one or more cells per well

[00552] Samples could be sorted with one B cell per well in some plates, and multiple B cells per well in other plates, yet heavy and light chains could still be paired for those wells having more than one B cell. We examined the sequences generated from the flu vaccination patient of example 9 above, where some wells had more than one distinct heavy chain sequence assembly or more than one distinct light chain sequence assembly observed. RT, PCR, sequencing and assignment to wells and assembly of sequences followed the protocol in example 9 above. For determining which heavy and light chains were associated with each other, heavy chains were assigned to clonal families by grouping all heavy chains with the same V and J gene usage, and the same number of nucleotides between the end of the V gene segment and the beginning of the J gene segment. Light chains were assigned to clonal families by grouping all light chains with the same V and J gene usage, and the same number of nucleotides between the end of the V gene segment and the beginning of the J gene segment. Pairing relationships between heavy and light chains were first assigned for wells with exactly one heavy chain and one light chain, based on them sharing a well (i.e. having the same compound barcodes). Then, a score was computed for each possible pairing of a heavy chain clonal family with a light chain clonal family. The score was determined by counting the number of times a member of the heavy chain family and the light chain family share a well. Then, each heavy chain family was associated with the light chain family with which the highest score was achieved, or the heavy chain family was not associated to a light chain family if the highest score was achieved with more than one light chain family. Individual heavy and light chains were then paired by starting with the overall highest-scoring heavy chain family, and proceeding well by well through the family assigning pairs, then continuing on with the next heavy chain family. For a given heavy chain family, for each well, if there was a single heavy chain within the well that was a member of the heavy chain family, then the light chain from that well which belonged to the heavy chain family's associated light chain family was assigned to be the heavy chain's pair. If more than one such light chain existed, no pairing was assigned. This process of associating heavy chains with light chains was continued until all families and all chains within those families had been considered. If, for a given heavy or light chain, the process resulted in more than one candidate for pairing, both heavy and light chain were discarded. Evolutionary trees were generated from the paired chains, and antibodies selected for downstream characterization of their functional properties. A portion of the evolutionary tree is shown in Figure 25A.

Example 21: Use of sorted plasmablasts to generate human monoclonal antibodies

[00553] From a subject with a recent or current condition resulting in acute, subacute, or ongoing generation of circulating plasmablast, flow cytometry is performed on peripheral blood (either whole blood or peripheral blood mononuclear cells (PBMCs)) to identify the plasmablast population. This population of B cells is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. During RT, well-specific sample-ID adaptor oligonucleotides are added to the reaction. These adaptors have well-specific barcode sequences (sample-IDs) that can identify sequences as originating from different wells. Utilizing the 3' tailing and template-switching activity of MMLV H- reverse transcriptases, sample-IDs are added to the 3' end of the 1st strand cDNA. cDNA from each plate are pooled together. During the first round of PCR, a plate-specific FW long primer1 adds the plate-ID to the 5' end of the amplicons. Thus, FW long primer1 with different plate-IDs are added to different plates giving each PCR product an identifying barcode sequence. Gene specific reverse primers are used to amplify the kappa, lambda and gamma chains, they are kappa GSP1, lambda GSP1 and gamma GSP1 respectively. These primers bind to the constant region of the immunoglobulin genes. Products from the first round of PCR are diluted and used for a second nested PCR. FW primer2 is used as the forward primer and reverse primers kappa, lambda, and gamma GSP long primers are used to amplify their respective amplicons. Notably, the GPS long primer2 for each plate adds a common plate ID to the 3' end of each amplicon for each plate, thus each will end up having two plate-IDs and a sample-ID barcode. Further details for RT, 1st and 2nd PCR are found in "non-touchdown PCR" in the materials and methods. Multiple plates are then pooled according to the method detailed in "preparing for 454 XLR70 sequencing run" and subjected to high-throughput 454 DNA sequencing and individual sequences identified with their barcodes serving as identifiers of which heavy and/or light chain is obtained from each well thus providing a guide for matching individual variable heavy and light chains derived from the same initial cell, according to the methods detailed in "Assignment of sequences to wells" and "Assembly of sequences" in the materials and methods. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity (see Figures 6-8).

[00554] Candidate heavy and light chain genes from particular cells of origin are then cloned and expressed for screening of desired properties as in example 8. Once stably or transiently transfected, the expression of the paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants

containing secreted antibodies are then screened for desired properties including but not limited to antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 demonstrate how this was done using the compositions and methods herein to obtain human monoclonal antibodies against hemagglutinin from single cell sorted plasmablasts from an influenza-vaccinated human.

Example 22: Use of sorted unbiased or antigen-specific memory B cells to generate human monoclonal antibodies

[00555] From a subject with a documented or suspected exposure to an antigen of interest, FACS is performed on peripheral blood (either whole blood or isolate peripheral blood mononuclear cells, PBMCs) to identify the memory B cell population (defined as $CD19^{+}CD20^{+}CD27^{+}$). Additionally, memory B cells specific against antigens of interest may also be sorted by staining peripheral blood or PBMCs with memory B cell surface markers and with fluorophore-conjugated antigen(s) ($CD19^{+}CD20^{+}CD27^{+}antigen^{+}$). This population of cells is then sorted by FACS either as single cells or multiple cells into wells. The process described in detail in Example 21 is repeated to barcode and obtain sequences from 454 sequencing and assign sequences to wells and assemble sequences. HighV-QUEST is used to identify VDJ gene usage, and a few members of each clonal family on an evolutionary tree selected for cloning and expression as in example 8. Cloning and expression is done as detailed in example 8. Once transfected, the expression of the entire paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing antibodies will be screened for antigen specificity against the target antigen(s) of interest as well as functionality by appropriate functional assays. Figures 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining human monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 23: Use of sorted unbiased or antigen-specific total B cells to generate human monoclonal antibodies

[00556] From a subject with or without documented or suspected prior exposure an antigen of interest, FACS is performed on peripheral blood (either whole blood or isolate peripheral blood mononuclear cells, PBMCs) to identify the $CD19^{+}$ B cell population. This population of cells is then sorted by FACS either as single cells or multiple cells into wells. The process described in detail in Example 21 is repeated. The process described in detail in Example 21 is repeated to barcode and obtain sequences from 454 sequencing and assign sequences to wells and assemble

sequences. HighV-QUEST is used to identify VDJ gene usage, and a few members of each clonal family on an evolutionary tree selected for cloning and expression as in example 8. Once transfected, the expression of paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing expressed antibodies will be screened for antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figures 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining human monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 24: Use of plasma cells to generate human monoclonal antibodies

[00557] From a subject with or without documented or suspected prior exposure an antigen of interest, FACS is performed on peripheral blood (either whole blood or isolate peripheral blood mononuclear cells, PBMCs) or bone marrow cells to identify the $CD138^{+}$ plasma cell population. This population of cells is then sorted by FACS either as single cells or multiple cells into wells. The process described in detail in Example 21 is repeated to barcode and obtain sequences from 454 sequencing and assign sequences to wells and assemble sequences. HighV-QUEST is used to identify VDJ gene usage, and a few members of each clonal family on an evolutionary tree selected for cloning and expression. Once transfected, the expression of paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing expressed antibodies will be screened for antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining human monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 25: Use of blasting B cells to generate human monoclonal antibodies

[00558] From a subject with or without documented or suspected prior exposure an antigen of interest, FACS is performed on peripheral blood (either whole blood or isolate peripheral blood mononuclear cells, PBMCs) to identify the FSC^{hi} blasting B cell population. Blasting cells are activated B cells, and therefore are cells that have responded against the antigen and are actively proliferating. These B cells consist of clonal families and their paired heavy and light chains can be used to obtain evolutionary trees. Other markers of B cell activation, such as $CD69^{hi}$ and $CD44^{hi}$ may also be used in conjunction. Additionally DNA content, which may be stained using cell permeable DNA stains such as SYTO Blue (Invitrogen), to determine cells that are

activated, proliferating and in the cell cycle may also be used in conjunction to delineate blasting B cells. This population of cells is then sorted by FACS either as single cells or multiple cells into wells. The process described in detail in Example 21 is repeated to barcode and obtain sequences from 454 sequencing and assign sequences to wells and assemble sequences. HighV-QUEST is used to identify VDJ gene usage, and a few members of each clonal family on an evolutionary tree selected for cloning and expression. Once transfected, the expression of paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing expressed antibodies will be screened for antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figures 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining human monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 26: Use of murine B cells to generate monoclonal antibodies

[00559] A mouse is challenged with an antigen of interest, and may be given booster shots several times before sacrificing the mouse to obtain murine B cells. Murine B cells may be obtained from blood, from splenocytes or from the bone marrow. Flow cytometry is performed to obtain $CD19^+$ or $B220^+$ B cells. This population of B cells is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT, 1st and 2nd PCR is performed as detailed in "non-touchdown PCR" in the materials and methods. Mouse gene-specific primers are found in Table 11 and other primers used for RT and PCR are found in Table 1. Multiple plates are then pooled according to the method detailed in "preparing for 454 XLR70 sequencing run" and subjected to high-throughput 454 DNA sequencing and individual sequences identified with their barcodes serving as identifiers of which heavy and/or light chain is obtained from each well thus providing a guide for matching individual variable heavy and light chains derived from the same initial cell, according to the methods detailed in "Assignment of sequences to wells" and "assembly of sequences" in the materials and methods. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity.

[00560] Sequences for cloning can either be obtained through synthetic gene synthesis or amplified from the 1^{st} PCR products using cloning primers. The forward cloning primer is sample-ID specific and can amplify specific sequences from a pool of amplicons. The sequence for each heavy and light chain is then cloned into an expression vector containing

complementary restriction sites for those introduced by the cloning primers. Vectors also contain the either the heavy or light chain constant region, which the heavy or light chain sequences are cloned into (reading frame aligned) to produce the entire antibody. Vectors contain either heavy or light chain clones are then dual transfected into a mammalian expression system or alternately, both amplicons can be cloned into a dual expression vector to allow for a single transfection into mammalian cells.

[00561] Candidate heavy and light chain genes from particular cells of origin are then expressed using the for screening of desired properties as above. Once stably or transiently transfected, the expression of the paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing secreted antibodies are then screened for desired properties including but not limited to antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 27: Use of murine plasma cells to generate monoclonal antibodies

[00562] A mouse is challenged with an antigen of interest, and may be given booster shots several times before sacrificing the mouse to obtain murine B cells. Murine plasma cells may be obtained from blood, from splenocytes or from the bone marrow, although the splenocytes and bone marrow are typically used. Flow cytometry is performed to obtain $CD19^{low/-}B220^{low/-}CD138^{+}$ plasma cells. This population of plasma cells is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT, 1st and 2nd PCR is performed as detailed in "non-touchdown PCR" in the materials and methods. Mouse gene-specific primers are found in Table 11, and other primers used for RT and PCR are found in Table 1. Multiple plates are then pooled according to the method detailed in "preparing for 454 XLR70 sequencing run" and subjected to high-throughput 454 DNA sequencing and individual sequences identified with their barcodes serving as identifiers of which heavy and/or light chain is obtained from each well thus providing a guide for matching individual variable heavy and light chains derived from the same initial cell, according to the methods detailed in "Assignment of sequences to wells" and "assembly of sequences" in the materials and methods. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity.

[00563] Sequences for cloning can either be obtained through synthetic gene synthesis or amplified from the 1st PCR products using cloning primers as described in example 26. Candidate heavy and light chain genes from particular cells of origin are then expressed using the for screening of desired properties as above. Once stably or transiently transfected, the expression of the paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing secreted antibodies are then screened for desired properties including but not limited to antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 28: Use of unbiased or antigen-specific murine memory B cells to generate monoclonal antibodies

[00564] A mouse is challenged with an antigen of interest, and may be given booster shots several times before sacrificing the mouse to obtain murine B cells. Murine memory B cells may typically be obtained from splenocytes or lymph nodes. Flow cytometry is performed to obtain $CD19^+$ or $B220^+$ and $CD38^+IgG^+$ memory B cells. Other markers, such as CD45RO, may also be used. Antigen-specific memory B cells may also be visualized by staining with fluorophore-conjugated antigen and sorted for. This population of memory B cells is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT, 1st and 2nd PCR , followed by sequencing, assignment of sequences to wells and sequence assembly is performed as in example 26. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity.

[00565] Sequences for cloning can either be obtained through synthetic gene synthesis or amplified from the 1st PCR products using cloning primers as described in example 26. Candidate heavy and light chain genes from particular cells of origin are then expressed using the for screening of desired properties as above. Once stably or transiently transfected, the expression of the paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing secreted antibodies are then screened for desired properties including but not limited to antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 3 and 9 each provide one example of a general methodology for carrying out this

method. Figures 26-27 provide another example of obtaining monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 29: Use of murine short-lived plasmablasts to generate monoclonal antibodies

[00566] A mouse is challenged with an antigen of interest, and may be given booster shots several times before sacrificing the mouse to obtain murine B cells. Murine short-lived plasmablasts may typically be obtained from splenocytes. These plasmablasts have been variously described as $CD19^{low/-}B220^{low/-}$ and $CD22^{low}$ or $CD11c^{+}$, and also as $CD138^{+}$. Flow cytometry is performed to obtain plasmablasts. This population of plasmablasts is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT, 1st and 2nd PCR, followed by sequencing, assignment of sequences to wells and sequence assembly is performed as in example 26. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity.

[00567] Sequences for cloning can either be obtained through synthetic gene synthesis or amplified from the 1^{st} PCR products using cloning primers as described in example 26. Candidate heavy and light chain genes from particular cells of origin are then expressed using them for screening of desired properties as above. Once stably or transiently transfected, the expression of the paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing secreted antibodies are then screened for desired properties including but not limited to antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 30: Use of murine blasting B cells to generate monoclonal antibodies

[00568] A mouse is challenged with an antigen of interest, and may be given booster shots several times before sacrificing the mouse to obtain murine B cells. Murine blasting B cells may typically be obtained from splenocytes. Blasting cells are activated B cells, and therefore are cells that have responded against the antigen and are actively proliferating. These B cells consist of clonal families and their paired heavy and light chains can be used to obtain evolutionary trees. Blasting B cells may be gated as FSC^{hi}, and may also be further identified via cell surface markers such as $CD44^{hi}$ $CD69^{hi}$ and as blasting B cells are proliferating, they may also be identified as having entered the cell cycle by having increased DNA content as stained by cell

permeably DNA stains such as SYTO Blue. Flow cytometry is performed to obtain blasting B cells. This population of plasmablasts is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT, 1st and 2nd PCR , followed by sequencing, assignment of sequences to wells and sequence assembly is performed as in example 26. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity.

[00569] Sequences for cloning can either be obtained through synthetic gene synthesis or amplified from the 1st PCR products using cloning primers as described in example 26. Candidate heavy and light chain genes from particular cells of origin are then expressed using the for screening of desired properties as above. Once stably or transiently transfected, the expression of the paired heavy and light chains will result in generation of monoclonal antibodies recapitulating the specificity of the initially sorted cell. Supernatants containing secreted antibodies are then screened for desired properties including but not limited to antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees

Example 31: Obtaining monoclonal antibodies from unbiased or antigen-specific human IgA+ B cells

[00570] From a subject with or without documented or suspected prior exposure an antigen of interest, FACS is performed on peripheral blood (either whole blood or isolate peripheral blood mononuclear cells, PBMCs) or on bone marrow to isolate IgA+ B cells. These B cells may be memory B cells, plasma cells, or plasmablasts. These IgA^+ B cells may also be antigen-specific, by sorting for antigen-positive B cells using a fluorophore-conjugated antigen to stain for the IgA+ B cells. This population of cells is then sorted by FACS either as single cells or multiple cells into wells. The process described in detail in Example 21 is repeated to barcode and obtain sequences from 454 sequencing and assign sequences to wells and assemble sequences, and IgA constant region specific primers used for PCR are in Table 10. HighV-QUEST is used to identify VDJ gene usage, and a few members of each clonal family on an evolutionary tree selected for cloning and expression as in example 8. Supernatants containing expressed antibodies will be screened for antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of

obtaining human monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 32: Obtaining monoclonal antibodies from unbiased or antigen-specific human IgM+ B cells

[00571] From a subject with or without documented or suspected prior exposure an antigen of interest, FACS is performed on peripheral blood (either whole blood or isolate peripheral blood mononuclear cells, PBMCs) to isolate IgM+ B cells. These B cells may be memory B cells, plasma cells, or blasting B cells. These IgM^+ B cells may also be antigen-specific, by sorting for antigen-positive B cells using a fluorophore-conjugated antigen to stain for the IgM+ B cells. This population of cells is then sorted by FACS either as single cells or multiple cells into wells. The process described in detail in example 21 is repeated to barcode and obtain sequences from 454 sequencing and assign sequences to wells and assemble sequences, and IgM constant region specific primers used for PCR are in Table 10. HighV-QUEST is used to identify VDJ gene usage, and a few members of each clonal family on an evolutionary tree selected for cloning and expression as in example 8. Supernatants containing expressed antibodies will be screened for antigen specificity against the target antigen of interest as well as functionality by appropriate functional assays. Figs. 9 and 6 each provide one example of a general methodology for carrying out this method. Figures 26-27 provide another example of obtaining human monoclonal antibodies from functional characterization of selected cloned and expressed antibodies from evolutionary trees.

Example 33: Obtaining monoclonal antibodies from unbiased or antigen-specific murine IgA+ B cells

[00572] A mouse is challenged with an antigen of interest, and may be given booster shots several times before sacrificing the mouse to obtain murine IgA^+ B cells. These B cells may be memory B cells, plasma cells, plasmablasts or blasting B cells, and can typically be obtained from splenocytes. These IgA^+ B cells may also be antigen-specific, by sorting for antigen-positive B cells using a fluorophore-conjugated antigen to stain for the IgA^+ B cells. This population of IgA^+ B cells is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT, 1st and 2nd PCR , followed by sequencing, assignment of sequences to wells and sequence assembly is performed as in example 26, and IgA constant region specific primers used for PCR are in Table 11. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity. Figs. 9 and 6 provides one example of a general methodology for carrying out this method.

Example 34: Obtaining monoclonal antibodies from unbiased or antigen-specific murine IgM+ B cells

[00573] A mouse is challenged with an antigen of interest, and may be given booster shots several times before sacrificing the mouse to obtain murine IgM^+ B cells. These B cells may be memory B cells, plasma cells, plasmablasts or blasting B cells, and can typically be obtained from splenocytes. These IgM^+ B cells may also be antigen-specific, by sorting for antigen-positive B cells using a fluorophore-conjugated antigen to stain for the IgM+ B cells. This population of IgM^+ B cells is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT, 1st and 2nd PCR , followed by sequencing, assignment of sequences to wells and sequence assembly is performed as in example 26, and IgA constant region specific primers used for PCR are in Table 11. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity. Figs. 9 and 6 provides one example of a general methodology for carrying out this method.

Example 35: Sequencing of more than one sequence from human T cells

[00574] From a subject with a recent or current condition resulting in acute, subacute, or ongoing generation of circulating T cells, flow cytometry is performed on peripheral blood (either whole blood or peripheral blood mononuclear cells (PBMCs)) to identify the T cell population of interest. This population of T cells may be activated T cells or blasting T cells. Activated T cells may be identified using $CD44^{hi}$, $CD69^{hi}$, $CD154^+$, $CD137^+$, or blasting T cells, which are also activated T cells may be delineated by their size or FSC^{hi}, and may also be identified as being in the cell cycle using a cell permeant DNA dye such as SYTO Blue. Activated T cells should consist of clonal families which then can be clustered into an evolutionary tree, with identical family members in clonal families, which can be used to select clones for downstream functional analysis. T cells are then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT and PCR to barcode the TCR genes are detailed in "Non-touchdown PCR" in the materials and methods, sequencing prep is detailed in "Preparing for 454 XLR70 sequencing run" in the materials and methods, and assignment of sequences to wells and assembly of reads are detailed in "Assignment of sequences to wells" and "assembly of sequences" in the materials and methods, TCR gene-specific primers are found in Table 10. Evolutionary trees are then constructed and candidate genes from particular cells of origin are then chosen to be cloned and expressed for screening of desired properties. Sequences for cloning can either be gene-synthesized or amplified from the

1st PCR products with cloning primers. Specific clones can be isolated from a pool of clones by having the forward cloning primer specific for the sample-ID barcode sequence. Reverse cloning primers are complementary for the appropriate gene. Both forward and reverse primers contain flanking restriction sites to integrate the clone (coding frame aligned) into a vector. Cells are either doubly transfected with two expression vectors, each containing either a gene of interest, or singly transfected with a dual expression vector that expresses both genes of interest, e.g., T cell alpha and beta chains.

[00575] Once stably or transiently transfected, the genes of interest can be expressed and screened for function properties using the desired screens.

Example 36: Sequencing of more than one sequence from murine T cells

[00576] A mouse is challenged with an antigen of interest, and may be given booster shots several times before sacrificing the mouse to obtain murine T cells. T cells are $CD3^+$, and helper T cells are $CD4^+$ and cytotoxic T cells are $CD8^+$. This population of T cells may be memory or activated T cells or blasting T cells. Memory T cells may be identified as $CD45RO^+$. Activated T cells may be identified using $CD44^{hi}$, $CD69^{hi}$, or blasting T cells, which are also activated T cells may be delineated by their size or FSC^{hi}, and may also be identified as being in the cell cycle using a cell permeant DNA dye such as SYTO Blue. All these T cells in a mouse kept in a clean environment after repeated antigen exposure should have a large fraction of clonal families which can then be displayed as an evolutionary tree, which can then be used to select TCRs for cloning and expression and downstream functional analysis.

[00577] T cells are sorted by flow cytometry using the suggested markers above. as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT and PCR to barcode the TCR genes are detailed in "Non-touchdown PCR" in the materials and methods, sequencing prep is detailed in "Preparing for 454 XLR70 sequencing run" in the materials and methods, and assignment of sequences to wells and assembly of reads are detailed in "Assignment of sequences to wells" and "assembly of sequences" in the materials and methods, TCR gene-specific primers are found in Table 11. Evolutionary trees are then constructed and candidate genes from particular cells of origin are then chosen to be cloned and expressed for screening of desired properties. Sequences for cloning can either be gene-synthesized or amplified from the 1st PCR products with cloning primers. Specific clones can be isolated from a pool of clones by having the forward cloning primer specific for the sample-ID barcode sequence. Reverse cloning primers are complementary for the appropriate gene. Both forward and reverse primers contain flanking restriction sites to integrate the clone (coding frame aligned) into a vector.

Cells are either doubly transfected with two expression vectors, each containing either a gene of interest, or singly transfected with a dual expression vector that expresses both genes of interest, e.g., T cell alpha and beta chains.

[00578] Once stably or transiently transfected, the genes of interest can be expressed and screened for function properties using the desired screens.

Example 37: Sequencing of more than one sequence from a sample

[00579] A single sample comprising nucleic acids of interest is identified. The single sample can have a single cell or a population of cells. The sample can be sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. This population of B cells is then sorted by flow cytometry as single cells into wells containing a hypotonic buffer with an RNAse inhibitor. Sorted cells can be frozen at this time or used immediately for RT-PCR to create cDNA. RT, 1st and 2nd PCR is performed as detailed in "non-touchdown PCR" in the materials and methods. Multiple plates are then pooled according to the method detailed in "preparing for 454 XLR70 sequencing run" and subjected to high-throughput 454 DNA sequencing and individual sequences identified with their barcodes serving as identifiers of which heavy and/or light chain is obtained from each well thus providing a guide for matching individual variable heavy and light chains derived from the same initial cell, according to the methods detailed in "Assignment of sequences to wells" and "assembly of sequences" in the materials and methods. Evolutionary trees are then drawn and antibodies selected for cloning, expression and determination of functional activity.

[00580] Candidate genes from particular cells of origin are then chosen to be cloned and expressed for screening of desired properties or other needs. Sequences for cloning can either be gene-synthesized or amplified from the 1st PCR products with cloning primers. Specific clones can be isolated from a pool of clones by having the forward cloning primer specific for the sample-ID barcode sequence. Reverse cloning primers are complementary for the appropriate gene. Both forward and reverse primers contain flanking restriction sites to integrate the clone (coding frame aligned) into a vector. Cells are either transfected with two or more expression vectors, each containing either a gene of interest, or singly transfected with an expression vector that expresses the genes of interest.

[00581] Once stably or transiently transfected, the genes of interest can be expressed and screened where desired. Figs. 9 and 6 provides one example of a general methodology for carrying out this method.

Example 38: Cloning of immunoglobulin V(D)J regions by DNA synthesis.

[00582] The desired immunoglobulin light chain and heavy chain V(D)J regions can be synthetically generated by DNA synthesis for cloning into expression vectors. The sequence used for the synthesis can be derived directly from the high-throughput 454 sequences, or alternatively cDNA encoding the heavy and light chain immunoglobulins from the sample(s) of interest can be re-sequenced from the individual sample or pooled samples for further verification of the sequence, and this sequence is used to synthesize the selected light chain and heavy chain V(D)J regions. Variable regions of Ig genes may be cloned by DNA synthesis, and incorporating the synthesized DNA into the vector containing the appropriate constant region using restriction enzymes and standard molecular biology. During synthesis, the exact nucleotide sequence need not be followed as long as long as the amino acid sequence is unchanged, unless mutagenesis is desired. This allows for codon optimization that may result in higher expression levels. This also allows for adding in restriction sites for the purpose of cloning. Non-translated sequences such as 5' UTR and barcode sequences need not be synthesized, leader sequences can also be swapped for other signal peptide sequences known for higher expression levels. These result in an Ig nucleotide sequence that can be very different from the high-throughput reads but give identical amino acid sequenced when expressed.

[00583] In one embodiment, the amplified V(D)J regions are inserted into vectors that already contain either the kappa, lambda, gamma or other heavy chain isotype constant regions with the appropriate restriction sites needed for inserting the amplified clones in the open reading frame. In another embodiment, the entire variable region may be gene synthesized with constant region and cloned into an expression vector for expression and downstream functional testing of antibody properties.

Example 39: Cloning of immunoglobulin V(D)J regions by using restriction site already present in sample identifcation adaptor.

[00584] In another aspect, the desired immunoglobulin light chain and heavy chain V(D)J regions can be cloned using restriction site already incorporated in the sample-ID adaptor added on during reverse transcription. This results in an adaptor with a restriction site 3' of the well-ID barcode in the PCR amplicon pool. During cloning with cloning primers, desired amplicons are amplified from a plate-specific amplicon pool using 5' primers that are complementary to the well-ID barcode sequences, and chain specific 3' primers (for the kappa, lambda and gamma chains). 3' primers will add on 3' restriction sites. 5' primers do not need to add restriction sites as the 5' primer already contains a restriction site 3' of the well-ID barcode. Following this amplification, restriction enzymes are used to cut the amplicon for ligation into the vector

containing the constant region insert. During the restriction enzyme digest, sequences added on to the 5' end of the Ig gene sequences, such as barcodes and universal sequences are cut as they are 5' of the 5' restriction site.

Example 40: Identification of clonal families by sequencing of just one immunoglobulin chain (heavy chain or light chain), followed by cloning of paired immunoglobulin heavy and light chain V(D)J regions

[00585] Antibody heavy and light chains are reversed transcribed from mRNA, incorporating distinct sample-IDs on the cDNAs generated from each sample, and sample cDNAs pooled for amplifying PCR. The immunoglobulin cDNAs are amplified and either the immunoglobulin heavy chain or the light chain is sequenced using 454 high-throughput sequencing, and the sequences grouped according to their use of immunoglobulin heavy chain V(D)J or light chain V(D)J sequences that exhibit use of the same genome-encoded V(D)J segments. Bioinformatics is used to identify clonal families of interest, and the desired immunoglobulin light and heavy chain V(D)J regions from the same sample are then selectively amplified for sequencing and/or cloning. For PCR amplification, the forward primer includes the sample-ID and the reverse primer is specific for the light chain or heavy chain constant region. The primers can incorporate restriction sites into the amplicons. Amplicons can then be inserted into the appropriate expression vectors that already contain a heavy or light chain constant region. Antibodies can then be expressed and screened for desired properties.

Example 41: Identification of clonal families from immunoglobulin heavy and light chain V(D)J sequencing for cloning and expression of antibodies using only sample-IDs (and no plate-IDs)

[00586] Antibody heavy and light chains are reversed transcribed from mRNA in each sample, incorporating distinct sample-IDs into the cDNA generated from each sample. Each sample-ID is at least 6 nucleotides long and 1 base-pair different, resulting in 4096 distinct potential sample-IDs. A distinct sample-ID is used for each sample, and the unique sample-IDs identify cDNA derived from different samples and enables paired sequencing and cloning of 2 or more cDNAs expressed in an individual sample. Heavy and light chain amplicons are amplifed using PCR, which adds on the Titanium adaptors A and B required for 454 high-throughput sequencing and all samples are then sent for sequencing. Sequences are assigned to wells and assembled following "Assignment of sequences to wells" and "Assembly of sequences" sections in the materials and methods. V(D)J assignments are made using HighV-QUEST and grouped into clonal families based on their V(D)J usage. Selected clones are then specifically amplified with cloning primers, which also adds in restriction sites into the amplicon. Amplicons are then inserted in-frame into expression vectors which already contain

the appropriate heavy or light constant regions for expression of the antibodies for screening for desired properties.

Example 42: Cloning of paired sequences by ligating on the universal primer sequence.

[00587] Antibody heavy and light chain genes are reversed transcribed from mRNA, which adds a 3' sequence to the newly synthesized cDNA consisting of an adaptor region and a sample-ID barcode. Samples are then pooled together and a universal primer sequence added to the 3' end of the 1st strand cDNA using T4 DNA ligase and a 5' phophorylated anti-sense universal primer oligonucleotide. Alternatively, 2nd strand cDNA synthesis may be done to obtain double stranded cDNA instead of an mRNA/cDNA hybrid before ligating on the universal primer sequence. Two rounds of PCR are then performed to amplify the cDNA and to add on plate-IDs and Titanium primers A and B for 454 sequencing. Alternatively, plate-IDs and Titanium Primers may also be added by DNA ligation instead of incorporated during PCR by using T4 DNA ligase. After 454 sequencing, sequences are assembled and clonal families identified. Selected clones from clonal families may be specifically cloned using cloning primers that add restriction sites to the amplicons. Sequences are then inserted in-frame into expression vectors that already have the appropriate heavy or light chain constant regions. Antibodies are then expressed and screened for desired properties.

Example 43: Testing of gene-specific primers for reverse transcription (RT-GSPs) of immunoglobulin heavy and light chains.

[00588] RT-GSPs were used instead of oligo(dT)s as primers in reverse transcription of heavy and light chain genes. cDNA were then amplified by PCR and visualized on an agarose gel. RT-GSP primers were IgKC_v3(a), IgLC_v5, IgLC_v6, IgLC_v7 and IgLC_v8 in lanes 1-4 respectively (b), IgHGC_v10, IgHGC_v11, IgHGC_v13 and IgGC_v15 in lanes 1-4 respectively (c) and IgHGC_v16 (d). KC, LC and GC in the primer names indicate that the primer is specific for the kappa chain, lambda chain and gamma heavy chain respectively. White bands in gel photos indicate where non-relevant lanes had been cropped out. *See* Figure 10 and Table 6.

Example 44: Testing of adaptor region sequences.

[00589] RNA was reversed transcribed with oligonucleotides comprising a universal primer region and an adaptor region at the 3' terminal end. cDNA was then amplified using the universal primer region sequence as a forward primer and gene-specific sequences as reverse primers. Amplified products were visualized on an agarose gel. Adaptor region consists of G (a), GGGGG and rGrGrG in lanes 1 and 2 respectively (b). rG indicates RNA nucleotides instead of DNA nucleotides. *See* Figure 11 and Table 6.

Example 45: Testing of universal primer sequences.

[00590] RNA was reverse transcribed with oligonucleotides comprising a universal primer sequence and an adaptor region at the 3' terminal end. cDNA were then amplified by PCR using a forward primer complementary to the universal primer region and a reverse primer complementary to the gene specific sequence. Univ_seq_4 (a), univ_seq_5 (b) and univ_seq_f (c). Vertical white bands in gel photos indicate where non-relevant lanes have been cropped out. Otherwise lanes belong to the same gel photo. *See* Figure 12 and Table 6.

Example 46: Testing of gene-specific primer sequences for 1st PCR reaction.

[00591] Gene-specific reverse primers were used in amplification of sequences in the first PCR reaction. Either the 1st PCR reaction or the subsequent 2nd nested PCR products were run and visualized on an agarose gel. Reverse primers used were IgKC_v4, IgLC_v5, IgHGC_v13 on lanes 1-3 respectively (a), K_GSP1, L_GSP1, G_GSP1 on lanes 1-3 respectively (b), K_GSP1c, L_GSP1c on lanes 1-2 respectively (c), G_GSP1 (d), L_GSP1d, G_GSP1g on lanes 1-2 respectively (e), G_GSP1h, G_GSP1k, L_GSP1f, L_GSP1g on lanes 1-4 respectively (f), G_GSP1d (g) L_GSP1h-o on lanes 1-8 respectively (h), G_GSP1m-q and G_GSP1t on lanes 1-6 respectively (K, L and G in the primer names indicate that the primers are specific for the kappa, lambda and gamma immunoglobulin constant regions respectively). Each gel starts with a lane marker on the left followed by sample lanes. White bars between lanes on the same gel photo indicate where non-relevant lanes in-between have been cropped out. *See* Figure. 13. Also, more primers were tested in Figure 43. These primers were used for the 1st PCR, and then the 2nd PCR was done using the primers from Table 1 and PCR products ran on a 2% agarose gel and image was taken. Primers used for 1st PCR are Kappa GSP1, kappa GSP1e, kappa GSP1f, lambda GSP1, lambda GSP1x and lambda GSP1y respectively. Also see Table 6 for sequences used. .

Example 47: Testing of gene-specific sequences for the 2nd PCR reaction.

[00592] Gene-specific reverse primers were used in amplification of sequences in the 2nd PCR reaction. PCR products were run and visualized on an agarose gel. Reverse primers used were K_GSP2, L_GSP2, G_GSP2 in lanes 1-3 respectively (a), K_GSP2v2a, K_GSP2v2b, L_GSP2v2 in lanes 1-3 respectively (b), K_GSP2v2c, K_GSP2v2c, G_GSP2v2c1, G_GSP2v2c2 in lanes 1-4 respectively (c), K_GSP2v2d-f in lanes 1-3 respectively (d),K_GSP2v2g, L_GSP2v2d and G_GSP2b in lanes 1-3 respectively (e). K, L, G in the primer names indicates that they are specific for the kappa, lambda and gamma immunoglobulin constant regions respectively. Each gel starts with a lane marker on the left followed by sample

lanes. White bars between lanes on the same gel photo indicate that non-relevant lanes in-between have been cropped out. *See* Figure 14 and Table 6.

Example 48: Testing of gene-specific primers for other human variable region genes

[00593] 1st and 2nd PCR were done using gene-specific reverse primers and products ran on a 2% agarose gel and imaged. Lanes are from left: marker, mu, alpha constant regions, TCR alpha (a) and marker, TCR beta (b). sequences of 3' primers used are in table 10. White bars between lanes on the same gel photo indicate where non-relevant lanes in-between have been cropped out. See Figure 44.

Example 49: Testing of gene-specific primers for mouse variable region genes

[00594] 1st and 2nd PCR were done and products ran on a 2% agarose gel and imaged. Lanes are from left: marker, kappa, lambda, lambda, lambda, lambda light chains and mu heavy chain (a). The 4 lambda lanes had this combination of primers used: mouse_lambda_GSP1a with mouse_lambda GSP2a, mouse_lambda_GSP1a with mouse_lambda GSP2b, mouse_lambda_GSP1b with mouse_lambda GSP2a and mouse_lambda_GSP1b with mouse_lambda GSP2a. Marker and alpha heavy chain (b). Gamma1, 2a, 2c heavy chains with 2nd PCR using mo_g12_GSP2d and mo_g12_GSP2e respectively, marker (c). Gamma 3 heavy chain with 2nd PCR using mo_g3_GSP2d, mo_g3_GSP2e respectively followed by gamma 2b heavy chain with 2nd PCR using mo_g2b_GSP2d, mo_g2b_GSP2e respectively, followed by marker (d). Marker, TCR alpha (e). Marker, TCR beta(f). White bars between lanes on the same gel photo indicate where non-relevant lanes in-between have been cropped out. See Figure 45 and Table 11.

Example 50: Generation of linked pairs of antibody heavy and light chains with a barcode at one end.

[00595] As shown in Figure 1, individual B cells can be sorted by flow cytometry from blood, bulk peripheral blood mononuclear cells (PBMCs), bulk B cells, plasmablasts, plasma cells, memory B cells, or other B cell populations. B cells are single-cell-sorted into 96-well PCR plates, leaving one column of wells empty, as a negative control. Figure 17 describes the general methodology for a method that can be used to link two polynucleotide sequences of interest and add a barcode at one end.

[00596] Single step multiplex overlap-extension RT-PCR can be performed using a commercially available one-step RT-PCR kit (e.g., Qiagen one-step RT PCR kit) according to the manufacturer's recommendations. In this particular example, a polynucleotide synthesis reaction, such as a reverse transcription reaction, is used to generate cDNA templates from an mRNA sample. Referring to Figure 17, the forward gene specific primer for the RT-PCR

reaction contains a restriction enzyme site (RE1), a sequencing primer site, and a barcode, to add these elements to a first cDNA of interest. Two additional primers (shown as containing RE3) have complementary overlap-extension tails. Use of these primers in a PCR reaction results in the two cDNAs of interest carrying overlap extension tails, which allow the two cDNAs of interest to anneal and link during amplification. In the example shown, a product of the indicated structure would be generated in which the LC and HC chains are physically linked with a barcode at one end.

[00597] The RE1 and RE2 restriction sites can be used clone the PCR product into suitable vectors for sequencing.

Example 51: Generation of linked pairs of antibody heavy and light chains with an internal barcode.

[00598] As shown in Figure 1, individual B cells can be sorted by flow cytometry from blood, bulk peripheral blood mononuclear cells (PBMCs), bulk B cells, plasmablasts, plasma cells, memory B cells, or other B cell populations. B cells are single-cell-sorted into 96-well PCR plates, leaving one column of wells empty, as a negative control. Figure 18 describes the general methodology for a method that can be used to link two polynucleotides of interest with a barcode located in between. Primers and oligonucleotides which can be used for antibody heavy and light chains are shown in Table 30. Restriction sites AsiSI and PacI are included in the RT oligonucleotides. Sample-ID sequences are shown in Table 2.

[00599] The method shown in Figure 18 relies on the 3' tailing and template switching activities of reverse transcriptase during a cDNA synthesis reaction. The 3' C tail added to the synthesized cDNA can be used for the annealing of an adaptor molecule carrying an overlap extension sequence and a barcode. Two types of adaptor molecules are used to link two cDNAs. A first adaptor carrying an overlap extension and barcode sequence is added to a first cDNA. A second adaptor carrying the reverse complement of the overlap extension without a barcode sequence is added to a second cDNA. The template switching property of reverse transcriptase adds these sequences to the 3' ends of their respective cDNAs.

[00600] In a PCR reaction, as shown in Figure 18, the complementary overlap extension sequences anneal and corresponding strands of DNA are synthesized from the site of annealing. Subsequent rounds of PCR using external primers results in amplification of the linked cDNA molecules.

[00601] Through the addition of appropriate restriction sites and the addition of sequencing primer sites incorporated into primers for the amplification reaction or later by ligation, the PCR products can be cloned into suitable vectors for sequencing.

Example 52: Generation of linked pairs of antibody heavy and light chains with two internal barcodes using universal sequence overlap-extension primers.

[00602] As shown in Figure 1, individual B cells can be sorted by flow cytometry from blood, bulk peripheral blood mononuclear cells (PBMCs), bulk B cells, plasmablasts, plasma cells, memory B cells, or other B cell populations. B cells are single-cell-sorted into 96-well PCR plates, leaving one column of wells empty, as a negative control. Figure 19 describes the general methodology for a method that can be used to introduce two internal barcodes in between two linked polynucleotides of interest. Primers and oligonucleotides which can be used for antibody heavy and light chains are shown in Table 31. Restriction sites AsiSI and PacI are included in the RT oligonucleotide. Sample-ID sequences are shown in Table 2.

[00603] The method shown in Figure 19 relies on the 3' tailing and template switching activities of reverse transcriptase during a cDNA synthesis reaction. In this example, the 3' C tail added to oligo (dT) primed cDNA can be used for the annealing of an adaptor molecule carrying a universal sequence and a barcode to each of the cDNAs to be joined. The template switching property of reverse transcriptase adds these sequences to the 3' ends of their respective cDNAs. Subsequent overlap-extension PCR using primers to the universal sequence which carry complementary overlap-extension sequences in combination with external LC and HC specific primers results in a structure in which LC is linked to HC with two internal barcodes between them as shown in Figure 19.

[00604] Through the addition of appropriate restriction sites and the addition of sequencing primer sites incorporated into primers for the amplification reaction or later by ligation, the PCR products can be cloned into suitable vectors for sequencing.

Example 53: Generation of linked pairs of antibody heavy and light chains with two internal barcodes using overlap-extension adaptors.

[00605] As shown in Figure 1, individual B cells can be sorted by flow cytometry from blood, bulk peripheral blood mononuclear cells (PBMCs), bulk B cells, plasmablasts, plasma cells, memory B cells, or other B cell populations. B cells are single-cell-sorted into 96-well PCR plates, leaving one column of wells empty, as a negative control. Figure 20 describes the general methodology for another method that can be used to introduce two internal barcodes in between two linked polynucleotides of interest. Primers and oligonucleotides which can be used for antibody heavy and light chains are shown in Table 32. Restriction sites AsiSI and PacI are included in the RT oligonucleotides. Sample-ID sequences are shown in Table 2.

[00606] The method shown in Figure 20 also relies on the 3' tailing and template switching activities of reverse transcriptase during a cDNA synthesis reaction. In this example, the 3' C

tail added to cDNA synthesized using gene specific primers can be used for the annealing of adaptor molecules carrying self complementary or palindromic overlap-extension sequences and a barcode to each of the cDNAs to be joined. The template switching property of reverse transcriptase adds these sequences to the 3' ends of their respective cDNAs. Subsequent annealing of the overlap-extension sequences added to the LC and HC cDNAs links them together at the site of overlap. Overlap-extension PCR using external primers to LC and HC results in a structure in which LC is linked to HC with two internal barcodes between them as shown in Figure 20.

[00607] Through the addition of appropriate restriction sites and the addition of sequencing primer sites incorporated into primers for the amplification reaction or later by ligation, the PCR products can be cloned into suitable vectors for sequencing.

Example 54: Studies on different methods of adding barcodes.

[00608] We investigated a variety of methods through which barcode sequences could be added during the course of a reverse transcription or amplification reaction using an oligonucleotide comprising the barcode sequence. We tested the addition of barcodes by incorporating them into gene-specific primers (GSPs) and into oligonucleotides containing one or more Gs that can be added to the 3' end of cDNAs by template switching. Based on the literature and our scientific knowledge, our expectation was that we would be able to effectively barcode cDNA using either 5' barcoded oligonucleotides or 3' barcoded GSPs.

[00609] As demonstrated in Figure 21, RT was performed with 1 μg of total PBMC RNA and 0.5 μM of univ_seq_2 template-switching oligo and 0.1 μM of IgKC_v3 GSP (lanes 1-2) and IgLC_v5 GSP (lanes 3-4) with an additional 5' flanking sequence, of which the first part is the Fixed_PCR3 sequence, and the last 8 bp AACAATAC is the barcode. The RT reaction was cleaned up using NucleoTraPCR (Macherey-Nagel) and dissolved in a final volume of 50 μl. 2 μl of this reaction was used in each subsequent PCR reaction, with either an internal 5' V_K (lane 1) or V_L (lane 3) primer or the Univ_seq_2 (lanes 2 and 4) as the 5' primer, and Fixed_PCR3 as the 3' primer. Note that the V_K primer is specific for kappa V genes 1 and 2, and the V_L primer is specific for lambda V gene 2. Sequences are in Table 33. As can be seen, the PCR products in lanes 2 and 4 ran as a smear. In contrast, the internal 5' primers produced distinct bands (lanes 1 and 3), showing that the primer pairs do work, and the smearing shown in lanes 1 and 3 cannot be attributed to poorly designed primers. As the oligonucleotide is added to all full-length reverse transcribed cDNA sequences, when a smear is obtained during PCR amplification with the univ_seq_2 and 3' barcoded GSPs, this suggests that reverse transcription with barcoded GSPs results in non-specifically primed nucleic acid sequences in the RT reaction.

Our results suggest that use of 5' universal sequence adaptors and 3' barcoded primers is not a good strategy for the barcoding and specific amplification of immunoglobulin or other genes expressed by a B cell or other cell.

[00610] In hindsight, several biologic properties of DNA and the molecular reactions used likely contribute to our observations. Reverse transcription is usually performed at low temperatures such as 42°C or 56°C. Unlike PCR, where the annealing step is usually performed at a temperature just slightly below the Tm of the primers to promote priming specificity, this cannot be done for reverse transcription, as reverse transcriptases are inactivated at high temperatures. Therefore, gene specific primers used during RT are typically not very specific for the gene of interest because the reaction proceeds at a temperature much lower than the Tm of the primer. In such a situation, the primer can also bind to off-target mRNA sequences with some mismatches, and mispriming occurs. If the barcode is added on the GSP, the primer also has to have a fixed sequence 5' of the barcode for use in subsequent PCR. This makes the primer very long (~60 nt), resulting in a primer with even a much higher degree of mispriming. However, specific amplification during PCR usually can still be achieved by using a highly gene-specific forward primer; as long as one member of the primer pair is specific, there usually can be specific amplification, as shown in lanes 1 and 3.

[00611] If a template switching technique is used to add an adaptor, this adaptor is added to all mRNAs that are synthesized to form first-strand cDNA. As mentioned above, the barcoded GSPs will have significant amounts of mispriming, especially as the RT enzyme Superscript III loses its template-switching activity at 56°C, and reverse transcription proceeds at 42°C. Specific nested 5' or 3' primers cannot be used as one would either lose the ability to PCR amplify all immunoglobulin genes (thus having to resort to multiplex PCR with multiple degenerate 5' primers due to variable V genes) or else lose the 3' barcode.

Therefore, barcoded GSPs are not suitable for use with template-switch added adaptors, or any other 5' adaptors, as other methods to add 5' adaptors such as TdT tailing or blunt end cloning also add adaptors non-discriminately. Therefore, internal 3' primers or a nested or semi-nested PCR amplification strategy is also required, and barcoded 3' GSPs do not allow for the use of these strategies for specific amplification of genes from a B cell or other cell. Based on our results, one would also anticipate that barcoded oligo(dT)s would also perform poorly for many of the same reasons we believe barcoded GSPs perform poorly. These reasons include but are not limited to an inability to use internal 3' primers or a nested or semi-nested PCR strategy for specific amplification of genes from a B cell or other cell.

[00612] In contrast, our results (see other examples) demonstrate the superiority of barcode sequence addition during the course of a reverse transcription reaction using a primer comprising the barcode sequence and an adaptor that anneals to the 3' tail of a cDNA generated during a reverse transcription reaction. In such an embodiment, the adaptor sequence can comprise a barcode sequence and be used to label genes encoding antibody heavy and light chains. Thus, as disclosed herein, template switching, or any other methods of tailing a cDNA adds a sequence that can be used for PCR amplification without prior knowledge of the 5' sequences themselves, enabling efficient and unbiased representation of the antibody repertoire. Furthermore, this approach allows one to obtain the repertoire of other co-expressed genes encoding proteins in addition to antibodies. Further, the approach of using template-switch adaptors has clear advantages over methods disclosed in the art that use sets of degenerate forward primers to amplify multiple V genes. These methods also fail to capture the entire antibody repertoire, because the known 5' primer sets: a) cover most but not the entire repertoire set; b) are not able to cover as yet known V genes variants (polymorphisms) in the human population; and c) may not be able to effectively amplify antibody sequences that have undergone extensive somatic hypermutation (SHM). See, e.g., Scheid et al., Sciencexpress, 14 July 2011 for an example of the effect of SHM.

[00613] Accordingly, the use of template-switch adaptors for the preparation of libraries of expressed genes, e.g., antibody heavy and light chains, provides clear advantages over other methods known in the art by allowing for unbiased representation of particular gene families and other co-expressed genes. The use of template-switch adaptors or 5' adaptors added using any other methods such as but not limited to TdT tailing and blunt end ligation are also more compatible with the use of barcoded 5' adaptors rather than barcoded 3' GSPs or barcoded oligo(dT)s for the reasons discussed above.

Example 55: Sorting of plasmablasts by forward-scatter and/or side-scatter on flow cytometer and/or in conjunction with other cell surface markers

[00614] Plasmablasts are blasting B cells that are activated, have proliferated/are proliferating and have undergone affinity maturation. Plasmablasts represent the active immune response and by practicing the methods and compositions herein allow for the bioinformatic construction of evolutionary trees with clonal families of antibodies that bind to target antigens of interest, whether it is an infection, a vaccine, autoimmune or cancer antigens.

[00615] Plasmablasts are blasting B cells and are larger than resting B cells (Figure 40A-B). Therefore, they can be sorted on a flow cytometer using their forward- and side-scatter properties. As shown in Figure 40c, plasmablasts have a median FSC-A that is ~1.29-1.66x

larger than the median FSC-A of other $CD20^+$ B cells, with a median FSC-W that is 1.04-1.16x larger than resting $CD20^+$ B cells. Plasmablasts also have a median FSC-A that is 0.85-0.98x that of monocytes, and a median FSC-W that is 0.92-1.02x that of monocytes as determined by the 95% confidence interval. Here FSC-A and FSC-H could be interchangeable and equivalent as FSC-A and FSC-H are scaled on calibrated flow cytometers to give the same values. Similarly for SSC-A (and SSC-H, due to scaling) and SSC-W, plasmablasts have a median SSC-A that is 0.74-2.56x that of $CD20^+$ B cells and 0.21-0.84x that of monocytes, and a median SSC-W that is 1.01-1.20x that of $CD20^+$ B cells and 0.82-1.03x that of monocytes. The ratio of plasmablasts to B cells is representative of that to lymphocytes, as resting lymphocytes are similar in size.

[00616] An alternative approach to identify plasmablasts is to use the 20th percentile FSC or SSC of plasmablasts to the median FSC or SSC of $CD20^+$ B cells or monocytes (Figure 44D), which is 1.04-1.50x (1.02-1.11x) that of median FSC-A (FSC-W) for $CD20^+$ B cells and 0.70-0.88x (0.88-1.00x) for monocytes. Plasmablasts have a 20th percentile SSC-A (SSC-W) which is 0.67-1.89x (0.99-1.11x) that of median SSC-A (SSC-W) for $CD20^+$ B cells and 0.20-0.62x (0.77-0.99x) for monocytes. These numbers allows a gating cutoff to include 80% of plasmablasts and exclude other lymphocytes. This allows for using FSC (and/or SSC) in conjunction with single or dual color stains to gate for plasmablasts in single cell sorting plasmablasts. Such combinations may include $FSC^{hi}CD19^{lo}$ (Figure 39b), $CD19^+FSC^{hi}$ (Figure 39c) $CD19^+FSC^{hi}CD20^-$ (Figure 39d), $CD19^+FSC^{hi}CD38^{hi}$ (Figure 39e) and $CD19^+FSC^{hi}CD27^+$ (Figure 39f). Sorted cells may then undergo RT, PCR for barcoding as carried out as in "non-touchdown" PCR in the materials and methods. Downstream preparation for sequencing, cloning and expression are as follows in examples 6 and 8. Note that ratios given are that of the 95% confidence interval, or where 95% of ratios should fall within this range.

Example 56: Sorting of plasmablasts by size on any sieving device, such as a microfluidics device

[00617] Plasmablasts are blasting B cells that are activated, have proliferated/are proliferating and have undergone affinity maturation. Plasmablasts represent the active immune response and by practicing the methods and compositions herein allow for the bioinformatic construction of evolutionary trees with clonal families of antibodies that bind to target antigens of interest, whether it is an infection, a vaccine, autoimmune or cancer antigens.

[00618] Plasmablasts are blasting B cells and are larger than resting B cells. Plasmablasts and $CD20^+$ B cells were FACS sorted and stained with trypan blue to exclude dead cells and imaged at 200x magnification. 52 plasmablasts and 51 $CD20^+$ B cells were imaged and cell area

measured with ImageJ. Plasmablasts imaged were between 7.8-13uM in diameter, and between 48-121uM^2 in area, and between in 251-996uM^3 in volume. CD20$^+$ B cells are not blasting and are smaller, the majority is between 6-8uM in diameter, or smaller than 50uM^2, or smaller than 268uM^3, with only 4 cells of 51 larger than that (Figure 41). Any sieving device that is capable of separating cells larger or smaller than 8uM in diameter or 50uM^2 in area or 268uM^3 in volume is capable of separating plasmablasts from CD20$^+$ resting B cells, with 96% of the plasmablasts captured, and sieving out 92% of the resting B cells. Such a device may be a fine sieve with 8uM diameter holes, or a microfluidics device with channels that only allows or prevents cells greater than 8uM in diameter or 50uM^2 in area or 268uM^3 in volume in passing through. These cells can then be sorted by actuators/pumps in the microfluidics device into wells such that there is only 1 or a few cell(s)/well and RT, PCR for barcoding may then be carried out as in "non-touchdown" PCR in the materials and methods using the same concentrations of reagents. Downstream preparation for sequencing, cloning and expression are as follows in examples 6 and 8.

Example 57: Anti-*Staphylococcus aureus* antibodies enhance phagocytosis of *S. aureus* by a neutrophil cell line

[00619] Humans with *S. aureus* infections who mount effective immune responses against their *S. aureus* infection, for example humans who clear *S. aureus* without the need for antibiotic therapy, are used as sources for peripheral blood from which peripheral blood plasmablasts are stained and sorted. Plasmablasts are single-cell sorted and barcoded as detailed in "Non touchdown PCR" in the materials and methods, and prepared for sequencing as detailed in "Preparing for 454 XLR70 sequencing" in the materials and methods. Evolutionary trees are bioinformatically constructed and a few select representatives of each clonal family are selected and cloned for expression as recombinant antibodies as in example 8. *S. aureus* Wood strain, which is ~5% protein A positive, is plated on 5% trypticase soy agar (TCA) blood agar and a colony grown and kept at 4°C as stock. This stock is refreshed weekly by picking another colony. 1mL of this stock is used to inoculate *S. aureus* growth till OD550 = 0.5, which is approximately mid-log growth phase. *S. aureus* is lightly fixed in 4% paraformaldehyde (PFA) for 15 minutes at room temperature and washed once with Hanks balanced salt solution (HBSS), before staining with 1uM CFSE for 15minutes at room temperature. Fixed bacteria are then washed and incubated with 10ug/ml of the expressed recombinant anti-*S. aureus* antibodies, or 10ug/ml of expressed anti-influenza virus antibodies as a negative control. Bacteria are then washed twice. HL-60, a neutrophil cell line, is activated for 96hr with 25uM retinoic acid, and incubated with labeled, fixed bacteria at 1:1 to 1:100 for 45 minutes at 37°C gently shaking at

300rpm in 96-well plates. HL-60 is then washed twice and analyzed on a flow cytometer. The amount of CFSE labeling in HL-60s is indicative of the amount of *S. aureus* phagocytosed. Some expressed anti-*S. aureus* antibodies will binding to staph cell surface proteins and opsonize the bacteria, leading to increased phagocytosis.

Example 58: Anti-*Staphylococcus aureus* antibodies enhance neutrophil-mediated killing of *S. aureus*

[00620] Humans who were able to effectively control and/or clear their *S. aureus* infections were selected as in the relevant example above, and plasmablasts were isolated and single cell sorted for sequencing and cloning and expression as in the relevant example above. An *S. aureus* clinical isolate, was plated on 5% TCA blood agar and a colony grown and kept at 4°C as stock. This stock was refreshed weekly by picking another colony. 1mL of this stock was used to inoculate *S. aureus* growth till OD550 = 0.5, which is approximately mid-log growth phase. *S. aureus* was then incubated with 2ug/ml of expressed anti- *S. aureus* antibodies for 30 minutes at 4°C before washing twice. The HL-60 neutrophil cell was activated for 96hr with 25uM of retinoic acid, and incubated with baby rabbit complement and *S. aureus* in a 1:1 to 1:100 ratio for 45 minutes at 37°C shaking at 300rpm in 96-well plates. HL-60 cells were then rapidly put on ice and washed 3x to remove loosely attached *S. aureus*. Extracellular *S. aureus* was then serially diluted and plated on 5%TSA blood agar and cultured overnight at 37°C. Colonies were counted the next day to determine the number of colony forming units (CFU). A decrease in CFUs by specific anti-*S. aureus* recombinant antibodies (after incubation with *S. aureus*) demonstrate that those antibodies were effective in mediating enhanced phagocytosis and killing, or reducing growth, of *S. aureus* by HL-60 cells (Figure 46).

Example 59: Treatment of *Staphylococcus aureus*-infected mice using expressed anti-*S. aureus* antibodies in mouse model

[00621] Anti-*S. aureus* antibodies that demonstrate *in vitro* killing, reduced growth or binding activity as in example 58 may also have *in vivo* activity. Anti-*S. aureus* antibodies with killing activity are isolated from *S. aureus*-infected humans who are able to control their staph infection as in examples 55-58. Mice are given a lethal dose of *S. aureus* and are then treated with a control antibody or a recombinant anti-*S. aureus* antibody(ies) with demonstrated *in vitro* killing, reduced growth or binding activity. Mice are deemed to be protected if they have a longer survival or reduced severity of infection as determined by the Kaplan-Meier survival test. Anti-*S. aureus* antibodies derived from humans who control or reduce the severity of their *S. aureus* infections are thereby evaluated for their ability to confer passive protection against *S. aureus*.

Example 60: Use of the antigen targets of effective anti-*Staphylococcus aureus* immune responses to develop vaccines.

[00622] *S. aureus* antigens that are targeted by anti-*S. aureus* antibodies that exhibit killing, reduced growth or binding activity are good candidates for a *S. aureus* vaccine. Vaccinees who develop a strong response against those specific antigens may be protected against or exhibit reduced severity of infection with *S. aureus*. Anti-*S. aureus* antibodies with killing, reduced growth or binding activity are isolated from *S. aureus*-infected humans who are able to control their *S. aureus* infection, and their target antigens identified using mass spectrometry as in examples 55-58. Mice are either vaccinated with a mock vector or vaccinated with candidate *S. aureus* antigens and then boosted twice over a period of a two months. Mice may be immunized with candidate *S. aureus* antigens individually or in combination. Anti-*S. aureus* antigen antibody titer is confirmed by ELISA. Mice are then challenged with a lethal dose of *S. aureus*. Mice are deemed to be protected against *S. aureus* if they have a longer survival as determined by the Kaplan-Meier survival test. Immunization against these selected *S. aureus* antigens therefore confers protection or reduces the severity of infection, showing that the compositions and methods herein can aid in vaccine design.

Example 61: Treatment of *Staphylococcus aureus*-infected humans using recombinant anti-*S. aureus* antibodies with *in vivo* killing, reduced growth or binding activity in humans

[00623] Antibodies derived from humans who control their *S. aureus* infection and exhibit *in vitro* and *in vivo* killing, reduced growth or binding activity as in examples 55-59 may be used to treat *S. aureus*-infected patients. Antibodies are obtained and tested for *in vitro* and *in vivo* killing activity as in examples 55-59. Good Manufacturing Practice (GMP) manufactured anti-*S. aureus monoclonal antibodies* may be given intravenously or subcutaneously to *S. aureus*-infected humans, especially patients infected with methicillin resistant *S. aureus* (MRSA) or other strains of drug-resistant-*S. aureus*, and compared to antibiotics along for efficacy. The anti-*S. aureus* antibodies are deemed to have therapeutic utility if patients are protected against invasive *S. aureus* infections, have less severe *S. aureus* infections, and/or recover more rapidly than patients given antibiotics alone or not given antibiotics. Recombinant anti-*S. aureus* antibodies can be given therapeutically to patients with active *S. aureus* infections to reduce the severity of infection and/or to enhance clearance of the infection as well as prophylactically to high-risk patient populations, such as patients on hemodialysis for renal failure, patients admitted to the hospital, or patients with a positive-screen for *S. aureus* or MRSA.

Example 62: *Staphylococcal aureus* vaccination using identified *S. aureus* antigens to confer protective immunity in humans

[00624] *S. aureus* antigens that are targeted by anti-*S. aureus* antibodies that exhibit killing or binding activity *in vivo*, and when vaccinated against in a mouse model confer protection against *S. Aureus* challenge, may be good candidates for a prophylactic vaccine in humans. Anti-*S. aureus* antibodies are derived from humans who control their *S. aureus* infections and are cloned, expressed and tested for *in vivo* killing, reduced growth or binding activity and as vaccine candidates in mice as in examples 55-58 and 59. Humans are then given the *S. aureus* vaccine containing *S. aureus* antigens that are the targets of anti-*S. aureus* antibodies with killing, reduced growth or binding activity, with a placebo being the control. The cohorts are tracked for their incidence or severity of *S. aureus* infections. The vaccine is deemed successful if it lowers the vaccinated cohort's incidence or severity of *S. aureus* infections compared to the placebo cohort.

Example 63: Monitoring of immune responses induced by candidate *Staphylococcus aureus* vaccines as a correlate of protection.

[00625] After immunization of humans with a candidate *S. aureus* vaccine as in example 62, the vaccine response may be monitored by determining if robust clonal families against the target *S. aureus* antigens of interest is elicited. Blood is drawn between 7-14 days post-vaccination and plasmablasts are single cell sorted, barcoded and 454 sequenced as detailed in "non-touchdown PCR" and "preparing for 454 XLR70 sequencing" in the materials and methods. Evolutionary trees are drawn and 2-3 members of each clonal family are then cloned and expressed as in example 8 and tested for their binding to the staph antigens of interest in an ELISA. We expect that humans who have a strong vaccine-induced anti-*S. aureus* immune response will exhibit large clonal families against the *S. aureus* antigens targeted in effective human immune responses. Such an approach has the potential to provide a correlate of protection for a *S. aureus* vaccine and in doing so enable clinical trials and development to be streamlined. This antibody and/or TCR immune repertoire monitoring would enable rapid assessment of the likelihood that a candidate vaccine would provide efficacy.

Example 64: Treatment of mice with lung-adenocarcinoma using recombinant anti-lung adenocarcinoma antibodies.

[00626] The anti-lung adenocarcinoma antibody that binds to a cell surface protein or other lung adenocarcinoma proteins may be useful as a carrier to target toxins to lung adenocarcinoma cells or to target other molecules expressed by lung adenocarcinoma cells. Anti-lung adenocarcinoma antibodies with cell surface binding activity or other lung adenocarcinoma

antigens are isolated from a long-term non-progressor lung adenocarcinoma cancer patient(s) as in example 11. Nude mice are given a subcutaneous injection of H1650 lung adenocarcinoma cell line and the tumor allowed to grow for one week. Anti-lung adenocarcinoma antibody is then conjugated to a toxin, such as diphtheria toxin lacking the R-domain, which is the cell-binding domain and allows diphtheria toxin into the cell. Diphtheria toxin lacking R-domain therefore is lethal only to lung adenocarcinoma cells which the antibody binds to and delivers the diphtheria toxin payload. Control antibody conjugated to diphtheria toxin without R domain is used as the control. The lung-adenocarcinoma antibody is deemed to have successfully delivered its payload to kill adenocarcinoma cells if the tumor load decreases more than in the control. Alternatively, in certain cases the recombinant antibody itself may be able to mediate tumor cell killing or to prevent tumor cell growth (in the absence of a conjugated toxin).

Example 65: Treatment of lung adenocarcinoma patients using expressed anti-lung adenocarcinoma antibody.

[00627] The anti-lung adenocarcinoma antibody that binds to cell surface antigen may be useful as a carrier to target toxins to lung adenocarcinoma cells. Anti-lung adenocarcinoma antibodies with cell surface binding activity are isolated from a long-term non-progressor lung adenocarcinoma cancer patient as in example 11. GMP monoclonal antibody or other anti-lung adenocarcinoma monoclonal antibodies may be given intravenously or subcutaneously to lung adenocarcinoma patients, especially to patients whose biopsied adenocarcinoma cells expressed high levels of the cell surface antigen targeted by the monoclonal antibody. The recombinant monoclonal antibody(ies) lung adenocarcinoma antigen(s), or other members of the clonal families from which they are derived, can be used to immunohistochemically stain a biopsy specimen of an individual patient's lung adenocarcinoma to gain information on tumor antigen expression levels, and this information can be used to determine whether an individual patient is likely to respond to therapy with this monoclonal antibody. Anti-lung adenocarcinoma antibodies can be conjugated to a toxin, such as diphtheria toxin lacking the R-domain, which is the cell-binding domain and allows diphtheria toxin into the cell. Diphtheria toxin lacking R-domain therefore is lethal only to lung adenocarcinoma cells which the antibody binds to and delivers the diphtheria toxin payload. Standard of care chemotherapy is used for treatment of the comparator group. The anti-adenocarcinoma antibodies are deemed to have delivered their payload and have therapeutic utility if patients survive longer or exhibit longer times prior to relapse or progression. Alternatively, in certain cases the recombinant antibody itself, against lung adenocarcinoma antigens, may be able to mediate tumor cell killing or prevent tumor cell growth (in the absence of a conjugated toxin).

Example 66: Lung adenocarcinoma therapeutic vaccination using identified antigens in humans.

[00628] The cell surface antigens bound by the anti-lung adenocarcinoma antibodies may be used in a therapeutic vaccine to treat established, or to protect against development of, lung adenocarcinoma. Anti-lung adenocarcinoma antibodies with cell surface binding activity are isolated from a long-term non-progressor lung adenocarcinoma cancer patient(s) as in example 11, and the target antigen identified using immunoprecipitation or immunoblotting and mass spectrometry. Vaccinees are given either a vaccine containing the lung adenocarcinoma antigen(s) of interest or a control vaccine. Cohorts are then followed and their incidence or progression of lung adenocarcinoma tracked. The vaccine is deemed to be successful if humans vaccinated with the target antigen have prolonged survival or extended time to relapse compared to the standard-of-care comparator group.

Example 67: Immune monitoring of lung adenocarcinoma vaccination for efficacy of response.

[00629] After immunization of humans with lung adenocarcinoma vaccine as in example 66, the vaccine response may be monitored by determining if robust clonal families against adenocarcinoma antigen(s) of interest in the vaccine are elicited. Blood is drawn between 7-14 days post-vaccination and plasmablasts are single cell sorted and barcoded, and 454 sequencing is performed as detailed in "non-touchdown PCR" and "preparing for 454 XLR70 sequencing" in the materials and methods. Evolutionary trees are drawn and 2-3 members of each clonal family are then cloned and expressed as in example 8 and tested for their binding to the staph antigens of interest in an ELISA. We expect that humans who have a strong immune response will have large clonal families against lung adenocarcinoma antigen(s) of interest and/or many clonal families against the adenocarcinoma antigen(s) of interest. This immune monitoring allows us to rapidly predict the efficacy of a candidate vaccine.

Example 68. Use of Superscript III for template switching during reverse transcription.

[00630] In our methods, the sample-identification region and adaptor regions are added on during reverse transcriptase. This utilizes the 3' tailing activity and template switching activity of RNase H^- reverse transcriptases. Most frequently, a reverse transcriptase such as Superscript II (Invitrogen) is used at its working temperature of 42°C. MMLV H^- reverse transcriptases that have also been engineered for thermal stability, such as Superscript III, with a recommended working temperature of 50°C, have been reported not to have this 3' tailing activity and therefore no template switching ability

(http://tools.invitrogen.com/content/sfs/ProductNotes/F_Superscript%20III%20Enzyme%20RD-

MKT-TL-HL0506021.pdf?ProductNoteId=36). However, in Figure 42, we showed that Superscript III does have 3' tailing and template switching activity. This property is weak at 50°C, the recommended reverse transcription temperature for Superscript III, and may explain why the 3' tailing activity of Superscript III has not been reported before. However, 3' tailing activity and template switching increases significantly as the RT temperature was lowered from 50°C to 45.5°C to 42°C. We would expect all MMLV RNase H- reverse transcriptases that have been engineered for thermal stability to also have 3' tailing activity at lower working temperatures, i.e. between 42°C to 50°C.

Example 69. Analysis of co-expressed genes to identify antibodies associated with memory B cell and plasma cell responses as well as homing to specific tissues.

[00631] Barcoding of all the cDNA produced by B cells, T cells or other cells sorted into individual wells as described by the methods and compositions herein enables characterization of gene co-expression in plasmablasts, other B cells, T cells and other cells at the single cell level. This enables use of co-expressed genes to identify the specific antibodies and TCRs expressed by B and T cells that have been induced to differentiate into memory B cells, plasma cells, memory T cells, specific types of effector T cells (for example, Th1, Th2, Th17 or T-regulatory T cells) or induced to home to a specific tissue or site (for example, the gastrointestinal tract, skin, or brain). The barcoding of all cDNA produced by the individual cell or collection of cells in a specific sample enables use of additional 3' PCR primers for both 1st and 2nd PCR to characterize the co-expression of specific such genes. 5' primers remain the same as those used to amplify variable regions genes. Furthermore, analysis of co-expressed genes enables bioinformatic analysis of the relationships between the affinity maturation of clonal families and the co-expression of genes associated with the differentiation of B cells to memory B cells, short-lived plasmablasts, and long-lived plasma cells (Table 34), the differentiation of naïve or memory T cells to Tregs or Th1, Th2, Th17 cells (Table 35), or the homing of B or T cells to specific sites. Such analysis can further pinpoint the critical antibodies or TCRs mediating an effective immune response.

[00632] For example, PMBCs derived from individuals mounting immune response are used to single cell sort plasmablasts. The methods and compositions herein are used to analyze co-expression of genes associated with homing of plasmablasts into different tissues (see Table 36). Bioinformatic analysis of the datasets identifies antibodies associated with secretion at different bodily locations. These antibody genes are then recombinantly expressed for characterization in in vitro screening assays as in Example 8.

[00633] It will be understood by persons skilled in the relevant art that various changes in form and details can be made therein without departing from the spirit and scope of the various aspects of the invention.

[00634] It must be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

[00635] All references, issued patents and patent applications cited within the body of the instant specification are hereby incorporated by reference in their entirety, for all purposes.

TABLES

Table 1 – Primers and adapter molecules

seq id no	DESCRIPTION	sequence
	Sample-ID adaptor	CACGACCGGTGCTCGATTTAG(sample-ID)GGG
	FW long primer1	GAGAGACTGACAGCGTATCGCCTCCCTCGCGCCATCAG (plate-ID)CACGACCGGTGCTCGATTTAG
	FW short primer1	GAGAGACTGACAGCGTATCGCCTC
	kappa GSP1	CGATTGGAGGGCGTTATCCAC
	lambda GSP1	TYTGTGGGACTTCCACTGCTC
	gamma GSP1	TCTTGTCCACCTTGGTGTTGCTG
	FW primer2	CGTATCGCCTCCCTCGCG
	kappa GSP long primer2	CTATGCGCCTTGCCAGCCCGCTCAG(plate-ID)TCAGATGGC GGGAAGATGAAGAC
	lambda GSP long primer2	CTATGCGCCTTGCCAGCCCGCTCAG(plate-ID)GAGGAGGGY GGGAACAGAGTGAC
	gamma GSP long primer2	CTATGCGCCTTGCCAGCCCGCTCAG(plate-ID)GGAAGTAGT CCTTGACCAGGCAG
	kappa GSP 2s	CTATGCGCCTTGCCAGCCCGCTCAGTCAGATGGCGGGAAGATGAAGAC
	lambda GSP 2s	CTATGCGCCTTGCCAGCCCGCTCAGGAGGAGGGY GGGAACAGAGTGAC
	gamma GSP 2s	CTATGCGCCTTGCCAGCCCGCTCAGGGAAGTAGTCCTTGACCAGGCAG
	RV primer2	CTATGCGCCTTGCCAGCCC

*For sample-ID sequences , see Table 2. For plate-ID sequences, see Table 3. Kappa GSP 2s, lambda GSP 2s and gamma GSP 2s are identical to the kappa, lambda and gamma GSP long primer 2 except that they do not have a plate-ID sequence. The plate-ID sequence is not necessary when doing XL+ runs, or when only forwards reads are desired when doing XLR70 runs with the Titanium LibA chemistry.

**primers sequences were designed to be able to amplify all different constant gene variants as found in IMGT database (http://imgt.cines.fr/).

Table 2 – Sample-ID			
Well	SampleID	Sequence of Sample-ID	SEQ ID NO
A1	1	ACGTCTCATCA	
A2	2	ACTCATCTACA	
A3	3	AGAGCGTCACA	
A4	4	AGTAGTGATCA	
A5	5	ATAGATAGACA	
A6	6	ATCTACTGACA	
A7	7	CACGTGTCGCA	
A8	8	CATACTCTACA	
A9	9	CGAGACGCGCA	
A10	10	CGTCGATCTCA	
A11	11	CTACGACTGCA	
B1	12	TAGTGTAGATA	
B2	13	TCTAGCGACTA	
B3	14	TGTGAGTAGTA	
B4	15	ACAGTATATAA	
B5	16	AGCTCACGTAA	
B6	17	TCGATAGTGAA	
B7	18	TCGCTGCGTAA	
B8	19	TGAGTCAGTAA	
B9	20	TGTAGTGTGAA	
B10	21	TGTCGTCGCAA	
B11	22	ACGACAGCTCA	
C1	23	TACACGTGATTAGGGATT	
C2	24	TACAGATCGTTAGGGAAA	
C3	25	TAGTGTAGATTTGGGTTT	
C4	26	TCTAGCGACTTTGGGTTT	
C5	27	ACGCGATCGAAGGGTTT	
C6	28	AGCTCACGTATTGGGTTT	
C7	29	AGTGCTACGAAGGGAAA	
C8	30	TCTGACGTCAAAGGGAAA	
C9	31	ACGTCTCATCAAGGAAGGAA	
C10	32	TATAGACATCAACACACAAA	
C11	33	AGTGAGCTCGTTGGGTTT	
D1	34	ATCGTCTGTGTTGGGTTT	
D2	35	CACACGATAGTTGGGTTT	
D3	36	CTGCGTCACGATCTCTCTT	
D4	37	TAGCATACTGTTGGGTTT	
D5	38	TATCTGATAGTTGGGTTT	
D6	39	TGACATCTCGCAGTTCTTT	
D7	40	TGATAGAGCGTAACAACAGA	

D8	41	TCACGCGAGAAAGGGAAA	
D9	42	ACACATACGCAGGGAAA	
D10	43	ACTAGCAGTAA	
D11	44	CGCAGTACGAA	
E1	45	TGAGTCAGTAAGGGAAA	
E2	46	ACTCATCTACAGGGAAA	
E3	47	ACTCGCGCACAGAGAGAA	
E4	48	AGAGCGTCACAGAGAGAA	
E5	49	AGCGACTAGCAACACACAAA	
E6	50	ATCTACTGACAACACACAA	
E7	51	CATACTCTACAACACACAAA	
E8	52	TCGAGCTCTCAGAGAGAAA	
E9	53	AGAGAGTGTGTTGGGTTT	
E10	54	ATCGTAGCAGAACACACAAA	
E11	55	CACTCGCACGTTGGGTTT	
F1	56	CAGACGTCTGAACACACAAA	
F2	57	CTCGATATAGTTGGGTTT	
F3	58	TCTGATCGAGAAGGGAAA	
F4	59	TACACACACTTAGGGATT	
F5	60	TACGTCATCAGGGAAA	
F6	61	CTACGCTCTAAGGGAAA	
F7	62	TAGTCGCATAAAGGGAAA	
F8	63	CGATCGTATAA	
F9	64	CGCGTATACAA	
F10	65	CTACGCTCTAA	
F11	66	TCACGCGAGAA	
G1	67	AGTATACATATTGGGTTT	
G2	68	TCGATAGTGAAGGGAAA	
G3	69	TCGCTGCGTAAAGGAGAAA	
G4	70	TGTAGTGTGAAGGGAAA	
G5	71	TGTCGTCGCAAGAGAGAG	
G6	72	CTACGACTGCAAGGGAAA	
G7	73	CTCTACGCTCA	
G8	74	TAGCTCTATCA	
G9	75	TATAGACATCA	
G10	76	TCACTCATACA	
G11	77	CTAGTCACTCAAGGGAAA	
H1	78	TGTGAGTAGTTTGGGTTT	
H2	79	TGTCACACGAAGGGAAA	
H3	80	CTGTGCGTCGAAGGGAAA	
H4	81	TAGTGTAGATTCGC	
H5	82	TCGAGCTCTCTCGC	

H6	83	ATCACGTGCGTCGC	
H7	84	CAGACGTCTGTCGC	
H8	85	TATCACTCAGTCGC	
H9	86	TGCTATAGACTTGGGTTT	
H10	87	CAGTACTGCGTTGGGTTT	
H11	88	CGACAGCGAGAACACACAAA	
A12-H12	89	TATGCTAGTAA (negative control)	

Table 3 – Plate-ID		
Plate	Sequence of Plate-ID	SEQ ID NO
1	ACGAGTGCGT	
2	ACGCTCGACA	
3	AGACGCACTC	
4	AGCACTGTAG	
5	ATCAGACACG	
6	ATATCGCGAG	
7	CGTGTCTCTA	
8	CTCGCGTGTC	
9	TGATACGTCT	
10	CATAGTAGTG	
11	CGAGAGATAC	
12	ATACGACGTA	
13	TCACGTACTA	
14	CGTCTAGTAC	
15	TCTACGTAGC	
16	TGTACTACTC	
17	CGTAGACTAG	
18	TACGAGTATG	
19	TACTCTCGTG	
20	TAGAGACGAG	
21	TCGTCGCTCG	
22	ACATACGCGT	
23	ACGCGAGTAT	
24	ACTACTATGT	
25	ACTGTACAGT	
26	AGACTATACT	
27	AGCGTCGTCT	
28	AGTACGCTAT	
29	ATAGAGTACT	
30	CACGCTACGT	
31	CAGTAGACGT	

32	CGACGTGACT	
33	TACACACACT	
34	TACACGTGAT	
35	TACAGATCGT	
36	TACGCTGTCT	
37	TAGTGTAGAT	
38	TCGATCACGT	
39	TCGCACTAGT	
40	TCTAGCGACT	
41	TCTATACTAT	
42	TGACGTATGT	
43	TGTGAGTAGT	
44	ACAGTATATA	
45	ACGCGATCGA	
46	ACTAGCAGTA	
47	AGCTCACGTA	
48	AGTATACATA	
49	AGTCGAGAGA	
50	AGTGCTACGA	
51	CGATCGTATA	
52	CGCAGTACGA	
53	CGCGTATACA	
54	CGTACAGTCA	
55	CGTACTCAGA	
56	CTACGCTCTA	
57	CTATAGCGTA	
58	TACGTCATCA	
59	TAGTCGCATA	
60	TATATATACA	

Table 4 – Cloning primers

seq id no	DESCRIPTION	sequence
	Clon_PacI	ACTGTTAATTAA(sample-ID)
	Clon_AscI	ATTAGGCGCGCC(sample-ID)
	Clon_FseI	ATTAGGCCGGCC(sample-ID)
	Clon_AsiSI	ATTAGCGATCGC(sample-ID)
	K_NheIa_DHFR	ACGTGCTAGCAGTTCCAGATTTCAACTGCTCATCAGA
	K_XhoId_DHFR	ACGTCTCGAGGATAGAAGTTATTCAGCAGGCACACAACA

	L_XhoI_PspXI_DHFR	ACTTGCTCGAGTCTGCYTTCCARGCMACTGT
	L_NheI_DHFR	AGTCGCTAGCCGCRTACTTGTTGTTGCTYTGTTTG
	G_EcoRI_DHFR	AGTCGAATTCCACGACACCGTCACCGGTT
	G_SacII_DHFR	ATTACCGCGGGGAAGGTGTGCACGCCG
	G_XhoI_PspXI_Lonza	ACGTCTCGAGGGTGCCAGGGGGAAGACCGATG
	G_AgeI_Lonza	ACTGACCGGTTCGGGGAAGTAGTCCTTGACCAGGCA
	G_EcoRI_Lonza	TGCAGAATTCCACGACACCGTCACCG
	G_ApaI_Lonza	TGTAGGGCCCTGAGTTCCACGACACCGTC
	L_XmaI_Lonza	TGATCCCGGGATAGAAGTCACTKATSAGRCACACYAGTGTGG
	L_BstEII_Lonza	TGCAGGTCACCGCTCCCGGGTAGAAGTCACTKATSAGR
	L_XhoI_PspXI_v2_Lonza	TGATGCTCGAGTCTGCYTTCCARGCMACTGTC
	K_XmaI_Lonza	TAGTCCCGGGGATAGAAGTTATTCAGCAGGCACAC

*Cloning forward primers start with a 5' flanking restriction site and end with sample-ID sequences on the 3' end. This enables cloning primers to discriminate between sequences with different well origins and selectively amplify amplicons with specific sample-ID sequences. Therefore, there are multiple cloning forward primers, each specific for particular sample-ID(s). The 3' sequences of the cloning forward primer are complementary to the well-ID and are provided in Table 5. Primers with names starting with "Clon" are the forward primers. Primers with names starting with "K", "L" or "G" are the reverse primers that are constant region specific for kappa, lambda and gamma chains respectively. The name of the reverse primers also denote the restriction site that the primer will incorporate. Finally, "DHFR" or "Lonza" denotes whether the constant region primers are for the vector set pcDNA3.3 and pOptivec or Lonza vectors pEE12.4 and pEE6.4 respectively, with constant region inserts added in.

Table 5 - Cloning Primers Well-Specific Sequence

Well	Sequence	SEQ ID NO
A1	GGTGCTCGATTTAGACGTCTCATCAG	
A2	CGGTGCTCGATTTAGACTCATCTACAG	
A3	GTGCTCGATTTAGAGAGCGTCACAG	
A4	CGGTGCTCGATTTAGAGTAGTGATCA	
A5	ACCGGTGCTCGATTTAGATAGATAGACA	
A6	CGGTGCTCGATTTAGATCTACTGACAG	
A7	CTCGATTTAGCACGTGTCGCA	
A8	CGGTGCTCGATTTAGCATACTCTACA	
A9	CGATTTAGCGAGACGCGCA	
A10	TGCTCGATTTAGCGTCGATCTCA	
A11	GTGCTCGATTTAGCTACGACTGCA	

B1	GACCGGTGCTCGATTTAGTAGTGTAGATAG	
B2	CGGTGCTCGATTTAGTCTAGCGACTAG	
B3	ACCGGTGCTCGATTTAGTGTGAGTAGTAG	
B4	CGACCGGTGCTCGATTTAGACAGTATATAA	
B5	GGTGCTCGATTTAGAGCTCACGTAAG	
B6	CGGTGCTCGATTTAGTCGATAGTGAA	
B7	TGCTCGATTTAGTCGCTGCGTAAG	
B8	CGGTGCTCGATTTAGTGAGTCAGTAA	
B9	CGGTGCTCGATTTAGTGTAGTGTGAA	
B10	GCTCGATTTAGTGTCGTCGCAA	
B11	GTGCTCGATTTAGACGACAGCTCA	
C1	CGGTGCTCGATTTAGTACACGTGATT	
C2	CGGTGCTCGATTTAGTACAGATCGTT	
C3	GACCGGTGCTCGATTTAGTAGTGTAGATTT	
C4	CGGTGCTCGATTTAGTCTAGCGACTTT	
C5	GCTCGATTTAGACGCGATCGAA	
C6	GGTGCTCGATTTAGAGCTCACGTATT	
C7	GGTGCTCGATTTAGAGTGCTACGAA	
C8	GGTGCTCGATTTAGTCTGACGTCAA	
C9	GGTGCTCGATTTAGACGTCTCATCAA	
C10	ACCGGTGCTCGATTTAGTATAGACATCAA	
C11	GGTGCTCGATTTAGAGTGAGCTCGT	
D1	GGTGCTCGATTTAGATCGTCTGTGT	
D2	GGTGCTCGATTTAGCACACGATAGT	
D3	GCTCGATTTAGCTGCGTCACGA	
D4	CCGGTGCTCGATTTAGTAGCATACTGT	
D5	GACCGGTGCTCGATTTAGTATCTGATAGT	
D6	GTGCTCGATTTAGTGACATCTCGC	
D7	CGGTGCTCGATTTAGTGATAGAGCGT	
D8	GCTCGATTTAGTCACGCGAGAAA	
D9	GGTGCTCGATTTAGACACATACGCA	
D10	CCGGTGCTCGATTTAGACTAGCAGTAA	
D11	TGCTCGATTTAGCGCAGTACGAA	
E1	CGGTGCTCGATTTAGTGAGTCAGTAA	
E2	CGGTGCTCGATTTAGACTCATCTACAG	
E3	GCTCGATTTAGACTCGCGCACA	
E4	GTGCTCGATTTAGAGAGCGTCACAG	
E5	GGTGCTCGATTTAGAGCGACTAGCA	
E6	CGGTGCTCGATTTAGATCTACTGACAA	
E7	CGGTGCTCGATTTAGCATACTCTACA	
E8	GTGCTCGATTTAGTCGAGCTCTCAG	
E9	CGGTGCTCGATTTAGAGAGAGTGTGT	

E10	GGTGCTCGATTTAGATCGTAGCAGA	
E11	GCTCGATTTAGCACTCGCACGT	
F1	TGCTCGATTTAGCAGACGTCTGAA	
F2	CGGTGCTCGATTTAGCTCGATATAGT	
F3	GGTGCTCGATTTAGTCTGATCGAGA	
F4	CGGTGCTCGATTTAGTACACACACTT	
F5	CGGTGCTCGATTTAGTACGTCATCA	
F6	CGGTGCTCGATTTAGCTACGCTCTAA	
F7	CGGTGCTCGATTTAGTAGTCGCATAA	
F8	GGTGCTCGATTTAGCGATCGTATAA	
F9	GGTGCTCGATTTAGCGCGTATACAA	
F10	CGGTGCTCGATTTAGCTACGCTCTAA	
F11	GCTCGATTTAGTCACGCGAGAAG	
G1	ACGACCGGTGCTCGATTTAGAGTATACATAT	
G2	CGGTGCTCGATTTAGTCGATAGTGAA	
G3	GCTCGATTTAGTCGCTGCGTAAA	
G4	CGGTGCTCGATTTAGTGTAGTGTGAA	
G5	GCTCGATTTAGTGTCGTCGCAA	
G6	GTGCTCGATTTAGCTACGACTGCA	
G7	GGTGCTCGATTTAGCTCTACGCTCA	
G8	CCGGTGCTCGATTTAGTAGCTCTATCA	
G9	ACCGGTGCTCGATTTAGTATAGACATCAG	
G10	CGGTGCTCGATTTAGTCACTCATACA	
G11	CGGTGCTCGATTTAGCTAGTCACTCA	
H1	CGGTGCTCGATTTAGTGTGAGTAGTTT	
H2	GGTGCTCGATTTAGTGTCACACGAA	
H3	GCTCGATTTAGCTGTGCGTCGA	
H4	GACCGGTGCTCGATTTAGTAGTGTAGATTC	
H5	GGTGCTCGATTTAGTCGAGCTCTCTC	
H6	TGCTCGATTTAGATCACGTGCGT	
H7	GTGCTCGATTTAGCAGACGTCTGTC	
H8	CCGGTGCTCGATTTAGTATCACTCAGT	
H9	ACCGGTGCTCGATTTAGTGCTATAGACT	
H10	GGTGCTCGATTTAGCAGTACTGCGT	
H11	GCTCGATTTAGCGACAGCGAGA	

TABLE 6		
SEQ ID NO	DESCRIPTION	SEQUENCE
	For kappa	
	IgKC v3	CAGATGGCGGGAAGATGAAGAC
	For Lambda	
	IgLC v5	CTCCCGGGTAGAAGTCAC
	IgLC v6	TCCCGGGTAGAAGTGAC

	IgLC v7	GCTCCCGGGTAGAAGTC
	IgLC v8	AGTGTGGCCTTGTTGG
	For gamma	
	IgHGC v10	GCCAGGGGGAAGACCGATG
	IgHGC v11	CAGGGGGAAGACCGATG
	IgHGC v13	AAGTAGTCCTTGACCAGGC
	IgHGC v15	GAAGACCGATGGGCCCTTGG
	IgHGC v16	AAGACCGATGGGCCCTTG
	Adaptor_v1	G
	Adaptor_v2	GGGGG
	Adaptor_v3	rGrGrG
	Univ seq 2	AACGCGTGACGAGAGACTGACAG
	Univ seq 4	TTGTTGCGTTCCTAGCCGCTATAG
	Univ seq 5	CTCTACGACCGGTGCTCGATTTAG
	Univ seq e	CCGTCGGTCGGCAGTG
	Kappa light chain specific	
	IgKC v4	ATGGCGGGAAGATGAAGAC
	K GSP1	GTGCTGTCCTTGCTGTCCTGCT
	K GSP1c	AGCAGGCACACAACAGAG
	K GSP1e	TTGTGTTTCTCGTAGTCTGCTTTGC
	K GSP1f	TCTCCCCTGTTGAAGCTCTTTGTG
	IgLC v5	CTCCCGGGTAGAAGTCAC
	L GSP1	ATCTGCCTTCCAGGCCACTGTC
	L GSP1c	CTCCCGGGTAGAAGTCAC
	L GSP1d	ACRGCTCCCGGGTAGAAGTCAC
	L GSP1f	TCCACGGTGCTCCCTTCAT
	L GSP1g	GGCCGCRTACTTGTTGTTGC
	L GSP1h	GCCTTCCAGGCCACTGTCAC
	L GSP1i	CTGCCTTCCAGGCCACTGTC
	L GSP1j	CTCCACGGTGCTCCCTTCA
	L GSP1k	GCTCCCTTCATGCGTGACC
	L GSP1l	TCTGTGGGACTTCCACTGCTC
	L GSP1m	GGGGCCACTGTCTTCTCCA
	L GSP1n	CTTCTGTGGGACTTCCACTGCT
	L GSP1o	ATCTGCCTTCCAGGCCACTGT
	L GSP1x	CTTYTGTGGGACTTCCACTGCTC
	L GSP1y	GCTTYTGTGGGACTTCCACTGCTC
	IgHGC v13	AAGTAGTCCTTGACCAGGC
	G GSP1c	TTCCACGACACCGTCAC
	G GSP1d	CACGCCGCTGGTCAG
	G GSP1g	GCTGCTGAGGGAGTAGAGTCCTGA
	G GSP1h	TCTTGTCCACCTTGGTGTTGCT
	G GSP1k	GCTGGAGGGCACGGTCAC
	G GSP1	TCTTGTCCACCTTGGTGTTGCTG
	G GSP1m	TCTTGTCCACCTTGGTGTTGCT
	G GSP1n	GACTGTAGGACAGCCGGGAAGG
	G GSP1o	ACCACGCTGCTGAGGGAGTAG
	G GSP1p	TTGTCCACCTTGGTGTTGCTG
	G GSP1q	TGAGTTCCACGACACCGTCAC
	G GSP1t	GAGTTCCACGACACCGTCACC

	Kappa specific	
	K GSP2	ATGGCGGGAAGATGAAGAC
	K GSP2v2a	ATGGCGGGAAGATGAAGAC
	K GSP2v2b	TGGCGGGAAGATGAAGAC
	K GSP2v2d	CGGAAGATGAAGACAGATGGT
	K GSP2v2e	GCAGTTCCAGATTTCAACTG
	K GSP2v2f	ATGGTGCAGCCACAGTT
	K GSP2v2c	CAGATTTCAACTGCTCATCAGAT
	K GSP2v2g	TCAGATGGCGGGAAGATGAAGAC
	Lambda specific	
	L GSP2	CTCCCGGGTAGAAGTCAC
	L GSP2v2c	AGGGYGGGAACAGAGTGAC
	L GSP2v2	CTCCCGGGTAGAAGTCAC
	L GSP2v2d	GAGGAGGGYGGGAACAGAGTGAC
	Gamma specific	
	G GSP2v2c1	GCCAGGGGGAAGACCG
	G GSP2v2c2	GGAAGTAGTCCTTGACCAGG
	G GSP2b	GGAAGTAGTCCTTGACCAGGCAG
	G GSP2	AAGTAGTCCTTGACCAGGC

Table 7	
Plate-ID	SEQ ID NO
TATGCTAGTA	
TCACGCGAGA	
TCGATAGTGA	
TCGCTGCGTA	
TCTGACGTCA	
TGAGTCAGTA	
TGTAGTGTGA	
TGTCACACGA	
TGTCGTCGCA	
ACACATACGC	
ACAGTCGTGC	
ACATGACGAC	
ACGACAGCTC	
ACGTCTCATC	
ACTCATCTAC	
ACTCGCGCAC	
AGAGCGTCAC	
AGCGACTAGC	
AGTAGTGATC	
AGTGACACAC	
AGTGTATGTC	

ATAGATAGAC	
ATATAGTCGC	
ATCTACTGAC	
CACGTAGATC	
CACGTGTCGC	
CATACTCTAC	
CGACACTATC	
CGAGACGCGC	
CGTATGCGAC	
CGTCGATCTC	
CTACGACTGC	
CTAGTCACTC	
CTCTACGCTC	
CTGTACATAC	
TAGACTGCAC	
TAGCGCGCGC	
TAGCTCTATC	
TATAGACATC	
TATGATACGC	
TCACTCATAC	
TCATCGAGTC	
TCGAGCTCTC	
TCGCAGACAC	
TCTGTCTCGC	
TGAGTGACGC	
TGATGTGTAC	
TGCTATAGAC	
TGCTCGCTAC	
ACGTGCAGCG	
ACTCACAGAG	
AGACTCAGCG	
AGAGAGTGTG	
AGCTATCGCG	
AGTCTGACTG	
AGTGAGCTCG	
ATAGCTCTCG	
ATCACGTGCG	
ATCGTAGCAG	
ATCGTCTGTG	
ATGTACGATG	
ATGTGTCTAG	
CACACGATAG	
CACTCGCACG	
CAGACGTCTG	

CAGTACTGCG	
CGACAGCGAG	
CGATCTGTCG	
CGCGTGCTAG	
CGCTCGAGTG	
CGTGATGACG	
CTATGTACAG	
CTCGATATAG	
CTCGCACGCG	
CTGCGTCACG	
CTGTGCGTCG	
TAGCATACTG	
TATACATGTG	
TATCACTCAG	
TATCTGATAG	
TCGTGACATG	
TCTGATCGAG	
TGACATCTCG	
TGAGCTAGAG	
TGATAGAGCG	
TGCGTGTGCG	
TGCTAGTCAG	
TGTATCACAG	
TGTGCGCGTG	

Table 8	
Sample-ID	SEQ ID NO
ACGAGTGCGT	
TAGACTGCAC	
TAGCGCGCGC	
TCATCGAGTC	
TCGCAGACAC	
TCTGTCTCGC	
TGATACGTCT	
TGAGTGACGC	
TGCTCGCTAC	
ACGTGCAGCG	
ACTCACAGAG	
AGACTCAGCG	
AGCTATCGCG	
AGTCTGACTG	
ATAGCTCTCG	
CATAGTAGTG	
CGATCTGTCG	

CGCGTGCTAG	
CGCTCGAGTG	
CGAGAGATAC	
TGAGCTAGAG	
ATACGACGTA	
TGCGTGTGCG	
TGCTAGTCAG	
TGTATCACAG	
TGTGCGCGTG	
TCACGTACTA	
CGTCTAGTAC	
TCTACGTAGC	
TGTACTACTC	
ACGCTCGACA	
ACGACTACAG	
CGTAGACTAG	
TACTCTCGTG	
TAGAGACGAG	
TCGTCGCTCG	
ACATACGCGT	
ACGCGAGTAT	
ACTGTACAGT	
AGACGCACTC	
AGACTATACT	
AGCGTCGTCT	
AGTACGCTAT	
ATAGAGTACT	
CACGCTACGT	
CAGTAGACGT	
CGACGTGACT	
AGCACTGTAG	
TACGCTGTCT	
TCGATCACGT	
TCGCACTAGT	
TCTATACTAT	
ATCAGACACG	
AGTCGAGAGA	
CGTACAGTCA	
CGTACTCAGA	
ATATCGCGAG	
CTATAGCGTA	
TATATATACA	
CGTGTCTCTA	
ACAGTCGTGC	

CTCGCGTGTC	
AGTGACACAC	
ATATAGTCGC	
CACGTAGATC	
CGACACTATC	
CTGTACATAC	

Table 9 - Primers to ligate on adaptors for XL+ sequencing		
seq id no	DESCRIPTION	sequence
	5LIB-LA	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTATCGCCTCCCTCGCGCCAT
	5LIB-LB	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCGTATCGCCTCCCTCGCGCCAT
	3LIB-LA	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTATGCGCCTTGCCAGCCCGCTCA
	3LIB-LB	CCTATCCCCTGTGTGCCTTGGCAGTCTCAGCTATGCGCCTTGCCAGCCCGCTCA

Table 10 - 3' primers for other human genes		
seq id no	DESCRIPTION	sequence
	mu constant region specific	
	mu GSP1	CTCTCAGGACTGATGGGAAGCC
	mu GSP2	CTATGCGCCTTGCCAGCCCGCTCAGGGGAATTCTCACAGGAGACGAGG
	alpha constant region specific	
	alpha GSP1	ATTCGTGTAGTGCTTCACGTGGC
	alpha GSP2	CTATGCGCCTTGCCAGCCCGCTCAGCTCAGCGGGAAGACCTTGGG
	TCR alpha constant region specific	
	TR alpha GSP1a	cgtttgcacatgcaaagtcagatt
	TR alpha GSP2b	CTATGCGCCTTGCCAGCCCGCTCAGtcggtgaataggcagacagacttg
	TRC beta constant region specific	
	TR beta GSP1	CCTATCCTGGGTCCACTCGTCA
	TR beta GSP2	CTATGCGCCTTGCCAGCCCGCTCAGCTGCTTCTGATGGCTCAAACACA

Table 11 - 3' primers for mouse genes		
seq id no	DESCRIPTION	sequence
	mu constant region specific	
	mouse_mu_GSP1	CTGAACCTTCAAGGATGCTCTTGG
	mouse_mu_GSP2	CTATGCGCCTTGCCAGCCCGCTCAGGGAAGACATTTGGGAAGGACTGAC
	alpha constant region specific	
	mouse_alpha_GSP1	TCTCCTTCTGGGCACTCGACAG
	mouse_alpha_GSP2	CTATGCGCCTTGCCAGCCCGCTCAGGGGAGTGTCAGTGGGTAGATGGTG
	gamma constant region specific	
	mo_g12b_GSP1d	AGGGGACAGTCACTGAGCTGCT
	mo_g2ac_GSP1d	TCGAGGTTACAGTCACTGAGCTGCT
	mo_g3_GSP1d	TGGAGGGTACAGTCACCAAGCTGCT
	mo_g12_GSP2d	CTATGCGCCTTGCCAGCCCGCTCAGGGGCCAGTGGATAGACHGATGG
	mo_g3_GSP2d	CTATGCGCCTTGCCAGCCCGCTCAGGGGACCAAGGGATAGACAGATGG
	mo_g12_GSP2e	CTATGCGCCTTGCCAGCCCGCTCAGCTGGACAGGGATCCAGAGTTCC
	mo_g3_GSP2e	CTATGCGCCTTGCCAGCCCGCTCAGCTGGACAGGGCTCCATAGTTCC
	kappa constant region specific	
	mouse_kappa_GSP1	GAAGTTGATGTCTTGTGAGTGGCCT
	mouse_kappa_GSP2	CTATGCGCCTTGCCAGCCCGCTCAGTGCTCACTGGATGGTGGGAA
	lambda constant region specific	
	mouse_lambda_GSP1a	ACTCTTCTCCACAGTGTCCCCTTCATG
	mouse_lambda_GSP1b	ACTCTTCTCCACAGTGTGACCTTCATG
	mouse_lambda_GSP2a	CTATGCGCCTTGCCAGCCCGCTCAGAGAGGAAGGTGGAAACASGGTGA
	mouse_lambda_GSP2b	CTATGCGCCTTGCCAGCCCGCTCAGAGGGGAAGGTGGAAACATGGTGA
	TCR alpha constant region specific	
	mo_TRA_GSP1b	TTGAAGATATCTTGGCAGGTGAAGCTT
	mouse_TRA_GSP2	CTATGCGCCTTGCCAGCCCGCTCAGCACAGCAGGTTCTGGGTTCTGG
	TRC beta constant region specific	
	mouse_TRB_GSP1	GAAAGCCCATGGAACTGCACTTG
	mouse_TRB_GSP2	CTATGCGCCTTGCCAGCCCGCTCAGGGGTGGAGTCACATTTCTCAGATC

*lambda GSP1a and GSP1b are to be mixed 50:50, lambda GSP2a and lambda GSP2b ar also to be mixed 50:50 to amplify all lambda constant region alleles that are found in IMGT. All the gamma GSP1ds are also to be mixed equally to amplify al gamma 1, 2a, 2b, 2c and 3 constant region alleles. gamma GSP2ds are also to be mixed 50:50, and gamma GSP2es are also to be mixed 50:50 to amplify all gamma 1 2a, 2b, 2c and 3 constant regions alleles that are found in IMGT database.

Table 12 - Plate to patient referencing

PlateID	Patient
2	B Staph
3	B Staph
4	B Staph
5	B Staph
7	Lung Adeno
8	Lung Adeno
9	Lung Adeno
10	Lung Adeno
11	360 CCP+RA
12	360 CCP+RA
13	360 CCP+RA
14	361 Staph
15	361 Staph
16	361 Staph
17	361 Staph
18	368 CCP+RF+RA
19	368 CCP+RF+RA
21	368 CCP+RF+RA
22	368 CCP+RF+RA
26	372 CCP+RF+RA
27	372 CCP+RF+RA
40	375 CCP+RF+
41	375 CCP+RF+
43	375 CCP+RF+
44	369 CCP+RF+RA
46	372 CCP+RF+RA
47	372 CCP+RF+RA
48	375 CCP+RF+
49	Flu
51	Flu
52	Flu
53	Flu

Table 13

Plate ID	Plate Identification Region Regular Expression
1	ACGAGTGCGT
2	ACGCTCGACA
3	AGACGCACTC

4	AGCACTGTAG
5	ATCAGACACG
6	ATATCGCGAG
7	CGTGTCTCTA
8	CTCGCGTGTC
9	TGATACGTCT
10	CATAGTAGTG
11	CGAGAGATAC
12	ATACGACGTA
13	TCACGTACTA
14	CGTCTAGTAC
15	TCTACGTAGC
16	TGTACTACTC
17	CGTAGACTAG
18	TACGAGTATG
19	TACTCTCGTG
20	TAGAGACGAG
21	TCGTCGCTCG
22	ACATACGCGT
23	ACGCGAGTAT
24	ACTACTATGT
25	ACTGTACAGT
26	AGACTATACT
27	AGCGTCGTCT
28	AGTACGCTAT
29	ATAGAGTACT
30	CACGCTACGT
31	CAGTAGACGT
32	CGACGTGACT
33	TACACACACT
34	TACACGTGAT

35	TACAGATCGT
36	TACGCTGTCT
37	TAGTGTAGAT
38	TCGATCACGT
39	TCGCACTAGT
40	TCTAGCGACT
41	TCTATACTAT
42	TGACGTATGT
43	TGTGAGTAGT
44	ACAGTATATA
45	ACGCGATCGA
46	ACTAGCAGTA
47	AGCTCACGTA
48	AGTATACATA
49	AGTCGAGAGA
50	AGTGCTACGA
51	CGATCGTATA
52	CGCAGTACGA
53	CGCGTATACA
54	CGTACAGTCA
55	CGTACTCAGA
56	CTACGCTCTA
57	CTATAGCGTA
58	TACGTCATCA
59	TAGTCGCATA
60	TATATATACA

Table 14	
Sample ID	Sample Identification Region Regular Expression
1,31	ACGTCTCATC

2,46	ACTCATCTAC
3,48	AGAGCGTCAC
4	AGTAGTGATC
5	ATAGATAGAC
6,50	ATCTACTGAC
7	CACGTGTCGC
8,51	CATACTCTAC
9	CGAGACGCGC
10	CGTCGATCTC
11,72	CTACGACTGC
12,81,25	TAGTGTAGAT
13,26	TCTAGCGACT
14,78	TGTGAGTAGT
15	ACAGTATATA
16,28	AGCTCACGTA
17,68	TCGATAGTGA
18,69	TCGCTGCGTA
19,45	TGAGTCAGTA
20,70	TGTAGTGTGA
21,71	TGTCGTCGCA
22	ACGACAGCTC
23	TACACGTGAT
24	TACAGATCGT
27	ACGCGATCGA
29	AGTGCTACGA
30	TCTGACGTCA
75,32	TATAGACATC
33	AGTGAGCTCG
34	ATCGTCTGTG
35	CACACGATAG
36	CTGCGTCACG

37	TAGCATACTG
38	TATCTGATAG
39	TGACATCTCG
40	TGATAGAGCG
66,41	TCACGCGAGA
42	ACACATACGC
43	ACTAGCAGTA
44	CGCAGTACGA
47	ACTCGCGCAC
49	AGCGACTAGC
52,82	TCGAGCTCTC
53	AGAGAGTGTG
54	ATCGTAGCAG
55	CACTCGCACG
56,84	CAGACGTCTG
57	CTCGATATAG
58	TCTGATCGAG
59	TACACACACT
60	TACGTCATCA
65,61	CTACGCTCTA
62	TAGTCGCATA
63	CGATCGTATA
64	CGCGTATACA
67	AGTATACATA
73	CTCTACGCTC
74	TAGCTCTATC
76	TCACTCATAC
77	CTAGTCACTC
79	TGTCACACGA
80	CTGTGCGTCG
83	ATCACGTGCG

85	TATCACTCAG
86	TGCTATAGAC
87	CAGTACTGCG
88	CGACAGCGAG
89	TATGCTAGTA

Table 15	
Sample ID	Sample Identification Region Regular Expression
1,31	ACGTCTCATCAAGGAAGGAAGG+
2,46	ACTCATCTACAGG+AA+GG+
3,48	AGAGCGTCACAGAGAGAAGG+
6,50	ATCTACTGACAACACACAAGG+
8,51	CATACTCTACAACACACAA+GG+
11,72	CTACGACTGCAAGG+AA+GG+
12,81,25	TAGTGTAGATT+GG+TT+GG+
12,81	TAGTGTAGATTCGCGG+
13,26	TCTAGCGACTT+GG+TT+GG+
14,78	TGTGAGTAGTT+GG+TT+GG+
16,28	AGCTCACGTATTGG+TT+GG+
17,68	TCGATAGTGAAGG+AA+GG+
18,69	TCGCTGCGTAA+GGAGAA+GG+
19,45	TGAGTCAGTAAGG+AA+GG+
20,70	TGTAGTGTGAAGG+AA+GG+
21,71	TGTCGTCGCAAGAGAGAGG+
75,32	TATAGACATCAACACACAA+GG+
66,41	TCACGCGAGAAAGG+AA+GG+
52,82	TCGAGCTCTCTCGCGG+
56,84	CAGACGTCTGTCGCGG+
65,61	CTACGCTCTAAGG+AA+GG+

Table 16. Constant region insert sequences for DHFR vectors pcDNA3.3 and pOptivec		
Seq ID no	Description	Sequence
	IGHG1-G1m3 constant region insert sequence (restriction sites BstBI, EcoRI, SacII, BbvCI for splicing in variable region introduced via silent mutations)	TAAGCTTACTGATAAGGCGCGCCGCGATCGCGCCTCCACCAAGGACCCCTC TGTCTTCCCCCTGGCACCCTCTTCGAAGAGCACCTCTGGGGACACAGCAGC CCTGGGCTGCCTGGTCAAGGACTACTTCCCCGAACCAGTGACAGTGAGTT GGAATTCAGGCGCCCTGACCAGCGGCGTGCACACCTTCCCcGcGGTCCTTC AGTCTTCAGGACTCTACTCCCTCAGCAGCGTGGTGACCGTGCCCTCCAGCA GCTTGGACACCCAGACCTACATCTGCAACGTGAATCACAAGCCCAGCAACA CCAAGGTGGACAAGAGAGTTGAGCCCAAATCTTGTGACAAAACTCACACA TGCCCACCGTGCCCAGCACCTGAACTCCTGGAGGGACCGTCAGTCTTCCTC TTCCTCCCAAAACCCAAGGACACCCTCATGATCTCCCGGACCCTAGAGGCC ACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCCGAGGTCAAGTTCAA CTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCTCGG GAGGAGCAGTACAACAGCACTTACCGTGTGGTCAGCGTCCTCACCGTCCT GCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACA AAGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAG CCCCGAGAACCACAGGTCTACACCCTGCCCCCATCCCGGGAGGAGATGAC CAAGAACCAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGA CATCGCCGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAG ACCACGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTATAGCAAGC TCACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTTTCATGCTCC GTGATGCATGAGGCTCTGCACAACCACTACACACAGAAGAGCCTCTCCCTG TCCCCAGGTAAATGATAATGTACAACTGACTGAGGATCCT
	IGKC-Km3 constant region insert sequence (restriction sites NheI, XhoI for splicing in variable region introduced via silent mutations)	TAAGCTTACTGACTAGGCCGGCCTTAATTAACGTACGGTGGCTGCACCATC TGTCTTCATCTTCCCTCCATCTGATGAGCAGTTGAAATCTGGAACTGCTAGC GTCGTGTGCCTGCTGAATAACTTTTATCCTCGAGAGGCCAAAGTGCAGTGG AAGGTGGATAACGCCCTCCAATCCGGTAACTCCCAGGAGTCCGTCACAGA GCAGGACAGCAAGGACAGCACCTACAGCCTGAGCAGCACCCTGACACTGA GCAAAGCAGACTACGAGAAACACAAAGTCTACGCCTGCGAAGTCACCCAT CAGGGCCTGAGCTCCCCCGTCACAAAGAGCTTCAACAGGGGAGAGTGTTG ATAAATCGATACTGACTGAGGATCCT
	IGLC-Mcg Ke Oz constant region insert sequence (restriction sites Bsu36I, XhoI, PspXI, NheI for splicing in variable region introduced via silent mutations)	TAAGCTTACTGACTAGGCCGGCCTTAATTAAGGTCAGCCTAAGGCTGCCCC CAGCGTCACTCTGTTCCCTCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGC CACACTGGTGTGTCTCATCAGTGACTTCTACCCCGGAGCCGTGACAGTGGC TTGGAAAGCAGACTCGAGCCCCGTCAAGGCTGGAGTGGAGACCACCACAC CTTCCAAACAAAGCAACAACAAGTACGCAGCTAGCAGCTACCGCAGCCTG ACCCCTGAGCAGTGGAAGTCCCACAGAAGCTACTCCTGCCAGGTCACACAT GAAGGGAGCACCGTGGAGAAGACAGTGGCCCCCACAGAATGTTCATGAT AAATCGATACTGACTGAGGATCCT

Table17 . Constant region insert sequences for Lonza vectors		
seq id no	DESCRIPTION	sequence
	IGHG1-G1m3 constant region insert sequence (restriction sites EcoRI, ApaI, AgeI,	AAGCTTGGCGCGCCTTAATTAAGCCAGCACAAAAGGCCCCAGTG TGTTTCCCTTGGCACCCTCGAGCAAGAGTACATCTGGAGGTACA GCTGCCTTGGGCTGTTTGGTGAAAGACTATTTCCCCGAACCGGT TACTGTCTCTTGGAATTCCGGGGCCCTCACCAGTGGTGTCCATA

	KpnI, XhoI for splicing in variable region introduced via silent mutations)	CCTTTCCCGCGGTGCTTCAGAGTTCCGGTTTGTATTCCCTGTCA AGTGTCGTGACGGTACCAAGTTCAAGTCTAGGCACCCAGACATA TATCTGTAACGTCAACCACAAGCCAAGCAACACCAAGGTTGACA AGCGGGTTGAACCTAAGTCCTGTGACAAGACCCATACCTGCCCC CCATGCCCCGCACCCGAGCTCCTCGGAGGGCCTTCCGTCTTTCT TTTCCCTCCCAAACCCAAGGACACTTTGATGATCTCAAGAACAC CAGAAGTCACTTGCGTCGTGGTTGACGTGTCTCACGAAGATCCC GAAGTGAAGTTCAACTGGTACGTGGATGGGGTAGAGGTTCATAA CGCCAAGACCAAACCCCGAGAGGAACAGTATAACTCCACCTATA GGGTAGTGTCCGTGCTCACCGTGCTCCACCAAGACTGGCTGAAT GGCAAGGAATACAAGTGCAAGGTGAGTAATAAGGCACTGCCTGC ACCCATTGAGAAGACAATATCTAAAGCAAAGGGACAGCCCAGAG AGCCCCAGGTTTATACTCTGCCACCTAGCAGAGAGGAAATGACT AAAAACCAGGTCAGCCTTACTTGTCTCGTAAAAGGCTTTTATCC AAGCGACATCGCTGTGGAGTGGGAATCAAATGGCCAACCTGAGA ATAATTATAAGACTACACCTCCCGTCCTTGACTCAGACGGTTCC TTCTTCCTGTATAGCAAGCTCACCGTCGATAAAAGTCGGTGGCA ACAGGGAAACGTGTTCTCATGCAGCGTCATGCACGAGGCCTTGC ACAATCATTACACCCAGAAGTCTCTGTCCCTGAGCCCTGGAAAG TGATCA
	IGKC-Km3 constant region insert sequence (restriction sites XmaI, EcoRI, BstEII, DraIII for splicing in variable region introduced via silent mutations)	AAGCTTAATTAAGGCGCGCCGAACAGTGGCTGCTCCTTCCGTGT TCATATTCCCCCCATCCGACGAGCAGCTTAAATCTGGGACTGCT AGCGTCGTGTGCCTGTTGAATAATTTTTATCCCCGGGAGGCTAA GGTACAGTGGAAGGTGGACAACGCCCTCCAATCAGGGAATTCCC AGGAGTCGGTCACCGAACAGGACAGCAAGGACTCAACCTACTCT CTGTCATCCACTCTCACACTCAGCAAAGCCGACTATGAAAAACA CAAAGTGTATGCTTGCGAGGTGACTCATCAAGGGCTCTCCAGTC CTGTGACTAAATCCTTCAACCGAGGCGAATGCTGATCA
	IGLC-Mcg^{-}Ke^{-}Oz^{-} constant region insert sequence (restriction sites DraIII, XmaI, BstEII, PspXI for splicing in variable region introduced via silent mutations)	AAGCTTGGCGCGCCTTAATTAAGGCCAGCCTAAAGCCGCACCCA GTGTGACCCTGTTTCCTCCCTCCTCTGAAGAGCTCCAGGCAAAC AAAGCTACTCTGGTGTGTCTTATTAGCGATTTCTATCCCGGGGC GGTGACCGTGGCTTGGAAGGCCGACTCGAGCCCAGTGAAGGCCG GAGTGGAAACTACAACCCCTTCCAAACAGTCAAACAATAAATAC GCCGCTAGCAGCTATCTCTCTCTCACCCCAGAACAGTGGAAATC CCACAGGTCCTATTCTTGCCAGGTCACACACGAGGGGTCAACCG TTGAGAAGACTGTTGCCCCAACAGAGTGCAGCTGATCA

Table 18. All expressed antibodies

SEQ ID NO	Antibody	Chain Type	Clone	V-GENE and allele	J-GENE and allele	D-GENE and allele
	LC1	Light chain	K8B8	IGKV3-11*01 F	IGKJ3*01 F	
	LC1	Heavy chain	G8B8	IGHV1-46*01 F, or IGHV1-46*03 F	IGHJ3*01 F, or IGHJ3*02 F	IGHD5-24*01 ORF
	LC2	Light chain	K8C11	IGKV3-11*01 F	IGKJ3*01 F, or IGKJ4*01 F	
	LC2	Heavy chain	G8C11	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ4*02 F	IGHD6-25*01 F
	LC3	Light chain	K8D6	IGKV3-20*01 F	IGKJ1*01 F	
	LC3	Heavy chain	G8D6	IGHV1-46*01 F, or IGHV1-46*03 F	IGHJ3*01 F, or IGHJ3*02 F	IGHD5-24*01 ORF
	LC5	Light chain	K10G5	IGKV3-11*01 F	IGKJ4*01 F	
	LC5	Heavy	G10G5	IGHV3-33*05 F	IGHJ4*02 F	IGHD6-13*01 F

		chain				
	LC6	Light chain	K8D6	IGKV3-20*01 F	IGKJ1*01 F	
	LC6	Heavy chain	G10H2	IGHV3-11*01 F	IGHJ4*02 F	IGHD3-9*01 F
	LC7	Light chain	L8D9	IGLV2-8*01 F	IGLJ2*01 F, or IGLJ3*01 F or IGLJ3*02 F	
	LC7	Heavy chain	G8D9	IGHV3-53*02 F	IGHJ4*02 F	IGHD4-17*01 F
	LC9	Light chain	L10A1	IGLV10-54*01 F	IGLJ3*02 F	
	LC9	Heavy chain	G10A1	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ3*02 F	IGHD2-2*01 F
	LC10	Light chain	K9C11	IGKV3-11*01 F	IGKJ4*01 F	
	LC10	Heavy chain	G9C11	IGHV3-33*01 F, or IGHV3-33*06 F	IGHJ6*02 F	IGHD5-18*01 F
	LC11	Light chain	L10A6	IGLV2-14*01 F	IGLJ3*02 F	
	LC11	Heavy chain	G10A6	IGHV3-21*01 F, or IGHV3-21*04 F	IGHJ4*02 F	IGHD1-20*01 F
	LC12	Light chain	L9C9	IGLV2-14*01 F	IGLJ3*02 F	
	LC12	Heavy chain	G9C9	IGHV3-15*01 F	IGHJ4*02 F	IGHD1-26*01 F
	LC13	Light chain	L9B1	IGLV2-8*01 F	IGLJ1*01 F	
	LC13	Heavy chain	G9B1	IGHV3-66*01 F, or IGHV3-66*04 F	IGHJ3*02 F	IGHD2-8*01 F
	LC14	Light chain	L9A1	IGLV2-14*01 F	IGLJ3*02 F	
	LC14	Heavy chain	G9A1	IGHV3-15*01 F	IGHJ4*02 F	IGHD1-26*01 F
	LC15	Light chain	K10A9	IGKV3-11*01 F	IGKJ3*01 F	
	LC15	Heavy chain	G10A9	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ4*02 F	IGHD6-25*01 F
	LC16	Light chain	K10D2	IGKV3-20*01 F	IGKJ2*02 F	
	LC16	Heavy chain	G10D2	IGHV3-15*01 F	IGHJ4*02 F	IGHD4-23*01 ORF
	LC17	Light chain	K8D5	IGKV3-11*01 F	IGKJ3*01 F	
	LC17	Heavy chain	G8D5	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ4*02 F	IGHD4-23*01 ORF
	LC18	Light chain	L10D5	IGLV2-8*01 F	IGLJ3*02 F	
	LC18	Heavy chain	G10D5	IGHV3-53*02 F	IGHJ4*02 F	IGHD4-17*01 F
	Flu14	Light	L51A6	IGLV3-25*03 F	IGLJ1*01 F	

		chain				
	Flu14	Heavy chain	G51A6	IGHV3-30-3*01 F	IGHJ3*01 F	IGHD1-14*01 ORF
	Flu15	Light chain	L51C4	IGLV3-1*01 F	IGLJ2*01 F, or IGLJ3*01 F or IGLJ3*02 F	
	Flu15	Heavy chain	G51C4	IGHV3-30*04 F	IGHJ6*02 F	IGHD3-10*01 F
	Flu16	Light chain	K51G11	IGKV1-33*01 F, or IGKV1D-33*01 F	IGKJ4*01 F	
	Flu16	Heavy chain	G51G11	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ6*02 F	IGHD4-17*01 F
	Flu17	Light chain	K49F7	IGKV1-33*01 F, or IGKV1D-33*01 F	IGKJ4*01 F	
	Flu17	Heavy chain	G49F7	IGHV1-69*02 F, or IGHV1-69*04 F	IGHJ3*02 F	IGHD3-22*01 F
	Flu18	Light chain	K51D7	IGKV1-33*01 F, or IGKV1D-33*01 F	IGKJ4*01 F	
	Flu18	Heavy chain	G51D7	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ6*02 F	IGHD4-11*01 ORF
	Flu19	Light chain	K51D8	IGKV1-39*01 F, or IGKV1D-39*01 F	IGKJ4*01 F	
	Flu19	Heavy chain	G51D8	IGHV3-49*04 F	IGHJ6*02 F	IGHD3-22*01 F
	Flu20	Light chain	K51G10	IGKV3-15*01 F	IGKJ2*01 F	
	Flu20	Heavy chain	G51G10	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ6*01 F	IGHD3-16*02 F
	Flu21	Light chain	L49A9	IGLV3-21*02 F	IGLJ3*02 F	
	Flu21	Heavy chain	G49A9	IGHV3-74*01 F	IGHJ2*01 F	IGHD2-21*01 F
	Flu22	Light chain	L52A6	IGLV2-14*01 F	IGLJ3*02 F	
	Flu22	Heavy chain	G52A6	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ6*02 F	IGHD4-23*01 ORF
	Flu23	Light chain	K49F11	IGKV3-11*01 F	IGKJ3*01 F	
	Flu23	Heavy chain	G49F11	IGHV3-30*04 F, or IGHV3-30*08 F or IGHV3-30-3*01 F	IGHJ4*02 F	IGHD3-16*01 F
	Flu24	Light chain	K51C8	IGKV3-15*01 F	IGKJ2*01 F	
	Flu24	Heavy chain	G51C8	IGHV3-30*14 F	IGHJ5*02 F	IGHD3-22*01 F
	Flu25	Light chain	K51H1	IGKV1-39*01 F, or IGKV1D-39*01 F	IGKJ4*01 F	
	Flu25	Heavy chain	G51H1	IGHV3-9*01 F	IGHJ4*02 F	IGHD6-13*01 F
	Flu26	Light chain	K52A2	IGKV1-33*01 F, or IGKV1D-33*01 F	IGKJ4*01 F	

	Flu26	Heavy chain	G52A2	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ6*02 F	IGHD4-23*01 ORF
	Flu27	Light chain	L49A11	IGLV2-14*01 F	IGLJ3*02 F	
	Flu27	Heavy chain	G49A11	IGHV3-7*03 F	IGHJ5*02 F	IGHD3-3*01 F
	Flu28	Light chain	L49C4	IGLV3-10*01 F	IGLJ2*01 F, or IGLJ3*01 F	
	Flu28	Heavy chain	G49C4	IGHV3-43*01 F	IGHJ4*02 F	IGHD4-17*01 F
	Flu29	Light chain	L51H5	IGLV3-27*01 F	IGLJ2*01 F, or IGLJ3*01 F	
	Flu29	Heavy chain	G51H5	IGHV1-2*04 F	IGHJ6*02 F	IGHD2-21*02 F
	Flu30	Light chain	L52B8	IGLV2-14*01 F	IGLJ3*02 F	
	Flu30	Heavy chain	G52B8	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ6*02 F	IGHD4-17*01 F
	Flu31	Light chain	K49A5	IGKV1-33*01 F, or IGKV1D-33*01 F	IGKJ4*01 F	
	Flu31	Heavy chain	G49A5	IGHV3-30-3*01 F	IGHJ4*02 F	IGHD5-12*01 F
	Flu32	Light chain	K49C11	IGKV1-33*01 F, or IGKV1D-33*01 F	IGKJ4*01 F	
	Flu32	Heavy chain	G49C11	IGHV3-66*01 F, or IGHV3-66*04 F	IGHJ4*02 F	IGHD6-19*01 F
	Flu33	Light chain	L51E9	IGLV2-14*01 F	IGLJ3*02 F	
	Flu33	Heavy chain	G51E9	IGHV3-23*01 F	IGHJ4*02 F	IGHD3-16*01 F
	Flu34	Light chain	L52G10	IGLV2-14*01 F	IGLJ3*02 F	
	Flu34	Heavy chain	G52G10	IGHV3-30-3*01 F	IGHJ3*01 F, or IGHJ3*02 F	IGHD6-19*01 F
	Flu35	Light chain	L53F10	IGLV3-21*02 F	IGLJ3*02 F	
	Flu35	Heavy chain	G53F10	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ6*02 F	IGHD3-3*01 F
	Flu36	Light chain	L52G7	IGLV2-8*01 F	IGLJ2*01 F, or IGLJ3*01 F or IGLJ3*02 F	
	Flu36	Heavy chain	G52G7	IGHV3-66*02 F	IGHJ4*02 F	IGHD2-15*01 F
	Flu37	Light chain	K51E8	IGKV3-15*01 F	IGKJ1*01 F	
	Flu37	Heavy chain	G51E8	IGHV1-2*04 F	IGHJ6*02 F	IGHD2-21*02 F
	Flu39	Light chain	K53G7	IGKV1-5*03 F	IGKJ1*01 F	
	Flu39	Heavy chain	G53G7	IGHV1-46*01 F, or IGHV1-46*03 F	IGHJ4*02 F	IGHD3-9*01 F

	Flu40	Light chain	L51A5	IGLV3-10*01 F	IGLJ2*01 F, or IGLJ3*01 F or IGLJ3*02 F	
	Flu40	Heavy chain	G51A5	IGHV4-34*01 F	IGHJ6*02 F	IGHD5-24*01 ORF
	Flu41	Light chain	L51B1	IGLV3-25*03 F	IGLJ1*01 F	
	Flu41	Heavy chain	G51B1	IGHV1-2*02 F	IGHJ4*02 F	IGHD3-10*01 F
	Flu43	Light chain	L51D3	IGLV3-22*01 F	IGLJ3*02 F	
	Flu43	Heavy chain	G51D3	IGHV4-39*01 F	IGHJ3*02 F	IGHD1-26*01 F
	Flu44	Light chain	L51D4	IGLV2-14*01 F	IGLJ3*02 F	
	Flu44	Heavy chain	G51D4	IGHV3-11*01 F	IGHJ6*02 F	IGHD4-17*01 F
	Flu45	Light chain	L52D4	IGLV2-14*01 F	IGLJ3*02 F	
	Flu45	Heavy chain	G52D4	IGHV3-30-3*01 F	IGHJ3*01 F, or IGHJ3*02 F	IGHD6-19*01 F
	Flu46	Light chain	L52H4	IGLV1-51*01 F	IGLJ3*02 F	
	Flu46	Heavy chain	G52H4	IGHV3-23*01 F, or IGHV3-23*04 F	IGHJ4*02 F	IGHD6-13*01 F
	S1	Light chain	K3G4	IGKV3-11*01 F	IGKJ3*01 F	
	S1	Heavy chain	G3G4	IGHV3-53*02 F	IGHJ4*02 F	IGHD2-21*01 F
	S2	Light chain	K4C4	IGKV2-28*01 F, or IGKV2D-28*01 F	IGKJ4*01 F	
	S2	Heavy chain	G4C4	IGHV3-23*04 F	IGHJ5*02 F	IGHD6-19*01 F
	S3	Light chain	K15C6	IGKV1-5*01 F	IGKJ4*01 F	
	S3	Heavy chain	G15C6	IGHV3-7*01 F	IGHJ1*01 F	IGHD2-2*01 F
	S4	Light chain	K15G1	IGKV1-6*01 F	IGKJ1*01 F	
	S4	Heavy chain	G15G1	IGHV3-7*03 F	IGHJ3*02 F	IGHD6-13*01 F
	S5	Light chain	K17C3	IGKV1-39*01 F, or IGKV1D-39*01 F	IGKJ2*02 F	
	S5	Heavy chain	G17C3	IGHV1-8*02 F	IGHJ4*02 F	IGHD3-16*01 F
	S6	Light chain	K3E11	IGKV3-15*01 F	IGKJ1*01 F	
	S6	Heavy chain	G3E11	IGHV3-30*04 F, or IGHV3-30-3*01 F	IGHJ4*02 F	IGHD4-23*01 ORF
	S7	Light chain	L4B8	IGLV2-8*01 F	IGLJ2*01 F, or IGLJ3*01 F	

	S7	Heavy chain	G4B8	IGHV3-30*04 F, or IGHV3-30*10 F	IGHJ4*01 F, or IGHJ4*02 F	IGHD5-18*01 F
	S8	Light chain	L4D2	IGLV2-23*01 F, or IGLV2-23*03 F	IGLJ3*02 F	
	S8	Heavy chain	G4D2	IGHV3-33*03 F	IGHJ6*02 F	IGHD3-10*01 F
	S9	Light chain	L4D6	IGLV2-8*01 F	IGLJ2*01 F, or IGLJ3*01 F	
	S9	Heavy chain	G4D6	IGHV3-20*01 F	IGHJ4*02 F	IGHD2-2*01 F
	S10	Light chain	L4F4	IGLV3-1*01 F	IGLJ3*02 F	
	S10	Heavy chain	G4F4	IGHV4-59*01 F, or IGHV4-59*08 F	IGHJ4*02 F	IGHD3-3*01 F
	S11	Light chain	L15D1	IGLV8-61*01 F	IGLJ3*02 F	
	S11	Heavy chain	G15D1	IGHV3-7*01 F	IGHJ4*02 F	IGHD3-10*01 F
	S12	Light chain	L17C6	IGLV1-47*01 F, or IGLV1-47*02 F	IGLJ3*02 F	
	S12	Heavy chain	G17C6	IGHV3-7*03 F	IGHJ4*02 F	IGHD5-18*01 F
	S13	Light chain	L17C9	IGLV7-46*01 F	IGLJ3*02 F	
	S13	Heavy chain	G17C9	IGHV5-a*03 F	IGHJ6*02 F	IGHD6-13*01 F
	RA1	Light chain	K11G5	IGKV3-11*01 F	IGKJ5*01 F	
	RA1	Heavy chain	G11G5	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ4*02 F	IGHD4-23*01 ORF
	RA2	Light chain	K22C7	IGKV3-15*01 F	IGKJ1*01 F	
	RA2	Heavy chain	G22C7	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ4*02 F	IGHD6-25*01 F
	RA3	Light chain	K26B1	IGKV3-15*01 F	IGKJ4*01 F	
	RA3	Heavy chain	G26B1	IGHV4-39*07 F	IGHJ4*02 F	IGHD4-23*01 ORF
	RA4	Light chain	K26F5	IGKV3-15*01 F	IGKJ5*01 F	
	RA4	Heavy chain	G26F5	IGHV3-23*01 F	IGHJ4*02 F	IGHD6-19*01 F
	RA5	Light chain	K26H1	IGKV3-11*01 F	IGKJ4*01 F	
	RA5	Heavy chain	G26H1	IGHV3-9*01 F	IGHJ4*02 F	IGHD6-13*01 F
	RA6	Light chain	K40C5	IGKV1-39*01 F, or IGKV1D-39*01 F	IGKJ3*01 F	
	RA6	Heavy chain	G40C5	IGHV3-30*04 F, or IGHV3-30*08 F or IGHV3-30-3*01 F	IGHJ4*02 F	IGHD3-16*01 F

	RA7	Light chain	K40G1	IGKV3-11*01 F	IGKJ1*01 F	
	RA7	Heavy chain	G40G1	IGHV4-39*01 F	IGHJ6*02 F	IGHD2-15*01 F
	RA8	Light chain	K40H4	IGKV1-33*01 F, or IGKV1D-33*01 F	IGKJ4*01 F	
	RA8	Heavy chain	G40H4	IGHV1-2*02 F	IGHJ5*02 F	IGHD2-21*02 F
	RA9	Light chain	K41A2	IGKV1-33*01 F, or IGKV1D-33*01 F	IGKJ4*01 F	
	RA9	Heavy chain	G41A2	IGHV1-2*02 F	IGHJ6*02 F	IGHD1-1*01 F
	RA10	Light chain	K47A2	IGKV3-15*01 F	IGKJ5*01 F	
	RA10	Heavy chain	G47A2	IGHV3-23*01 F	IGHJ4*02 F	IGHD6-19*01 F
	RA11	Light chain	K47E2	IGKV3-15*01 F	IGKJ5*01 F	
	RA11	Heavy chain	G47E2	IGHV3-23*01 F	IGHJ4*02 F	IGHD6-19*01 F
	RA12	Light chain	K47F9	IGKV3-11*01 F	IGKJ4*01 F	
	RA12	Heavy chain	G47F9	IGHV4-39*02 F	IGHJ3*02 F	IGHD3-3*02 F
	RA13	Light chain	L13B10	IGLV3-1*01 F	IGLJ1*01 F	
	RA13	Heavy chain	G13B10	IGHV5-51*01 F	IGHJ6*02 F	IGHD6-25*01 F
	RA14	Light chain	L13G5	IGLV3-1*01 F	IGLJ1*01 F	
	RA14	Heavy chain	G13G5	IGHV5-51*01 F	IGHJ6*02 F	IGHD6-25*01 F
	RA15	Light chain	K40D6	IGKV3-20*01 F	IGKJ5*01 F	
	RA15	Heavy chain	G40D6	IGHV4-39*07 F	IGHJ4*02 F	IGHD4-23*01 ORF
	RA16	Light chain	K25D6	IGKV3-11*01 F	IGKJ3*01 F	
	RA16	Heavy chain	G26D6	IGHV3-72*01 F	IGHJ6*03 F	IGHD4-17*01 F
	RA17	Light chain	K25E9	IGKV3-11*01 F	IGKJ3*01 F	
	RA17	Heavy chain	G25E9	IGHV3-30*03 F, or IGHV3-30*18 F or IGHV3-33*05 F	IGHJ5*02 F	IGHD3-22*01 F
	RA18	Light chain	K25G4	IGKV1-27*01 F	IGKJ2*03 F	
	RA18	Heavy chain	G25G4	IGHV3-30*09 F	IGHJ4*02 F	IGHD1-1*01 F
	RA19	Light chain	K45D9	IGKV3-15*01 F	IGKJ1*01 F	

	RA19	Heavy chain	G45D9	IGHV3-30*03 F, or IGHV3-30*18 F	IGHJ4*02 F	IGHD3-22*01 F
	RA21	Light chain	L13E11	IGLV2-23*02 F	IGLJ1*01 F	
	RA21	Heavy chain	G13E11	IGHV3-15*01 F	IGHJ4*02 F	IGHD6-13*01 F
	RA22	Light chain	L13G5	IGLV3-1*01 F	IGLJ1*01 F	
	RA22	Heavy chain	G13G5	IGHV3-7*01 F	IGHJ5*02 F	IGHD5-12*01 F
	RA23	Light chain	L44C5	IGLV2-23*01 F, or IGLV2-23*02 F or IGLV2-23*03 F	IGLJ1*01 F	
	RA23	Heavy chain	G44C5	IGHV3-30*14 F	IGHJ5*02 F	IGHD7-27*01 F
	RA24	Light chain	L44D6	IGLV2-23*01 F, or IGLV2-23*02 F or IGLV2-23*03 F	IGLJ1*01 F	
	RA24	Heavy chain	G44D6	IGHV3-30*04 F	IGHJ6*03 F	IGHD3-10*01 F
*VDJ identity as given by V-QUEST.						

Table19 - antibodies used in Fluzone ELISA

Antibody	Chain Type	Clone
Flu14	Light chain	L51A6
Flu14	Heavy chain	G51A6
Flu15	Light chain	L51C4
Flu15	Heavy chain	G51C4
Flu16	Light chain	K51G11
Flu16	Heavy chain	G51G11
Flu17	Light chain	K49F7
Flu17	Heavy chain	G49F7
Flu18	Light chain	K51D7
Flu18	Heavy chain	G51D7
Flu19	Light chain	K51D8
Flu19	Heavy chain	G51D8
Flu20	Light chain	K51G10
Flu20	Heavy chain	G51G10
Flu21	Light chain	L49A9
Flu21	Heavy chain	G49A9
Flu22	Light chain	L52A6
Flu22	Heavy chain	G52A6
Flu23	Light chain	K49F11
Flu23	Heavy chain	G49F11
Flu25	Light chain	K51H1
Flu25	Heavy chain	G51H1

Flu26	Light chain	K52A2
Flu26	Heavy chain	G52A2
Flu27	Light chain	L49A11
Flu27	Heavy chain	G49A11
Flu29	Light chain	L51H5
Flu29	Heavy chain	G51H5
Flu30	Light chain	L52B8
Flu30	Heavy chain	G52B8
Flu33	Heavy chain	G51E9
Flu34	Light chain	L52G10
Flu34	Heavy chain	G52G10
Flu35	Light chain	L53F10
Flu35	Heavy chain	G53F10
Flu37	Light chain	K51E8
Flu37	Heavy chain	G51E8
Flu39	Light chain	K53G7
Flu39	Heavy chain	G53G7
Flu40	Light chain	L51A5
Flu40	Heavy chain	G51A5
Flu41	Light chain	L51B1
Flu41	Heavy chain	G51B1
Flu43	Light chain	L51D3
Flu43	Heavy chain	G51D3
Flu44	Light chain	L51D4
Flu44	Heavy chain	G51D4
Flu45	Light chain	L52D4
Flu45	Heavy chain	G52D4
Flu46	Light chain	L52H4
Flu46	Heavy chain	G52H4
S1	Light chain	K3G4
S1	Heavy chain	G3G4
S2	Light chain	K4C4
S2	Heavy chain	G4C4

Table 20- Antibodies used in surface plasmon resonance		
Antibody	Chain Type	Clone
Flu14	Light chain	L51A6
Flu14	Heavy chain	G51A6
Flu15	Light chain	L51C4
Flu15	Heavy chain	G51C4
Flu16	Light chain	K51G11
Flu16	Heavy chain	G51G11
Flu17	Light chain	K49F7
Flu17	Heavy chain	G49F7

Flu18	Light chain	K51D7
Flu18	Heavy chain	G51D7
Flu19	Light chain	K51D8
Flu19	Heavy chain	G51D8
Flu20	Light chain	K51G10
Flu20	Heavy chain	G51G10
Flu21	Light chain	L49A9
Flu21	Heavy chain	G49A9
Flu22	Light chain	L52A6
Flu22	Heavy chain	G52A6
Flu26	Light chain	K52A2
Flu26	Heavy chain	G52A2
Flu29	Light chain	L51H5
Flu29	Heavy chain	G51H5
Flu34	Light chain	L52G10
Flu34	Heavy chain	G52G10
Flu35	Light chain	L53F10
Flu35	Heavy chain	G53F10
Flu46	Light chain	L52H4
Flu46	Heavy chain	G52H4

Table 21 -Antibodies used in RA antigen array

Antibody	Chain Type	Clone
RA1	Light chain	K11G5
RA1	Heavy chain	G11G5
RA2	Light chain	K22C7
RA2	Heavy chain	G22C7
RA4	Light chain	K26F5
RA4	Heavy chain	G26F5
RA5	Light chain	K26H1
RA5	Heavy chain	G26H1
RA8	Light chain	K40H4
RA8	Heavy chain	G40H4
RA9	Light chain	K41A2
RA9	Heavy chain	G41A2
RA10	Light chain	K47A2
RA10	Heavy chain	G47A2
RA11	Light chain	K47E2
RA11	Heavy chain	G47E2
RA12	Light chain	K47F9
RA12	Heavy chain	G47F9
RA13	Light chain	L13B10
RA13	Heavy chain	G13B10
RA16	Light chain	K25D6

RA16	Heavy chain	G26D6
RA19	Light chain	K45D9
RA19	Heavy chain	G45D9
RA22	Light chain	L13G5
RA22	Heavy chain	G13G5
RA23	Light chain	L44C5
RA23	Heavy chain	G44C5
Flu14	Light chain	L51A6
Flu14	Heavy chain	G51A6
Flu26	Light chain	K52A2
Flu26	Heavy chain	G52A2

Table 22 -Antibodies used in Histone 2A ELISA and CCP ELISA		
Antibody	Chain Type	Clone
RA1	Light chain	K11G5
RA1	Heavy chain	G11G5
RA2	Light chain	K22C7
RA2	Heavy chain	G22C7
RA4	Light chain	K26F5
RA4	Heavy chain	G26F5
RA5	Light chain	K26H1
RA5	Heavy chain	G26H1
RA6	Light chain	K40C5
RA6	Heavy chain	G40C5
RA7	Light chain	K40G1
RA7	Heavy chain	G40G1
RA8	Light chain	K40H4
RA8	Heavy chain	G40H4
RA9	Light chain	K41A2
RA9	Heavy chain	G41A2
RA10	Light chain	K47A2
RA10	Heavy chain	G47A2
RA11	Light chain	K47E2
RA11	Heavy chain	G47E2
RA12	Light chain	K47F9
RA12	Heavy chain	G47F9
RA13	Light chain	L13B10
RA13	Heavy chain	G13B10
RA16	Light chain	K25D6
RA16	Heavy chain	G26D6
RA17	Light chain	K25E9
RA17	Heavy chain	G25E9
RA18	Light chain	K25G4
RA18	Heavy chain	G25G4

RA19	Light chain	K45D9
RA19	Heavy chain	G45D9
RA22	Light chain	L13G5
RA22	Heavy chain	G13G5
RA23	Light chain	L44C5
RA23	Heavy chain	G44C5
RA24	Light chain	L44D6
RA24	Heavy chain	G44D6

Table 23 -Antibodies used in RF ELISA		
Antibody	Chain Type	Clone
RA1	Light chain	K11G5
RA1	Heavy chain	G11G5
RA2	Light chain	K22C7
RA2	Heavy chain	G22C7
RA4	Light chain	K26F5
RA4	Heavy chain	G26F5
RA5	Light chain	K26H1
RA5	Heavy chain	G26H1
RA6	Light chain	K40C5
RA6	Heavy chain	G40C5
RA8	Light chain	K40H4
RA8	Heavy chain	G40H4
RA9	Light chain	K41A2
RA9	Heavy chain	G41A2
RA10	Light chain	K47A2
RA10	Heavy chain	G47A2
RA11	Light chain	K47E2
RA11	Heavy chain	G47E2
RA12	Light chain	K47F9
RA12	Heavy chain	G47F9
RA14	Light chain	L13G5
RA14	Heavy chain	G13G5

Table 24 - Antibodies used in lung cancer Tissue IHC and flow cytometry of lung cancer cell lines		
Antibody	Chain Type	Clone
LC1	Light chain	K8B8
LC1	Heavy chain	G8B8
LC5	Light chain	K10G5
LC5	Heavy chain	G10G5
LC6	Light chain	K8D6
LC6	Heavy chain	G10H2
LC7	Light chain	L8D9

LC7	Heavy chain	G8D9
LC9	Light chain	L10A1
LC9	Heavy chain	G10A1
LC10	Light chain	K9C11
LC10	Heavy chain	G9C11
LC11	Light chain	L10A6
LC11	Heavy chain	G10A6
LC12	Light chain	L9C9
LC12	Heavy chain	G9C9
LC13	Light chain	L9B1
LC13	Heavy chain	G9B1
LC14	Light chain	L9A1
LC14	Heavy chain	G9A1
LC15	Light chain	K10A9
LC15	Heavy chain	G10A9
LC16	Light chain	K10D2
LC16	Heavy chain	G10D2
LC17	Light chain	K8D5
LC17	Heavy chain	G8D5
LC18	Light chain	L10D5
LC18	Heavy chain	G10D5
Flu14	Light chain	L51A6
Flu14	Heavy chain	G51A6

Table 25 - Antibodies used in *S. aureus* surface staining		
Antibody	Chain Type	Clone
S1	Light chain	K3G4
S1	Heavy chain	G3G4
S2	Light chain	K4C4
S2	Heavy chain	G4C4
S3	Light chain	K15C6
S3	Heavy chain	G15C6
S4	Light chain	K15G1
S4	Heavy chain	G15G1
S6	Light chain	K3E11
S6	Heavy chain	G3E11
S7	Light chain	L4B8
S7	Heavy chain	G4B8
S8	Light chain	L4D2
S8	Heavy chain	G4D2
S9	Light chain	L4D6
S9	Heavy chain	G4D6
S10	Light chain	L4F4
S10	Heavy chain	G4F4

S11	Light chain	L15D1
S11	Heavy chain	G15D1
S12	Light chain	L17C6
S12	Heavy chain	G17C6
S13	Light chain	L17C9
S13	Heavy chain	G17C9
Flu14	Light chain	L51A6
Flu14	Heavy chain	G51A6
Flu26	Light chain	K52A2
Flu26	Heavy chain	G52A2

Table 26 - Antibodies used in microneutralization assay

Antibody	Chain Type	Clone
Flu15	Light chain	L51C4
Flu15	Heavy chain	G51C4
Flu16	Light chain	K51G11
Flu16	Heavy chain	G51G11
Flu18	Light chain	K51D7
Flu18	Heavy chain	G51D7
Flu19	Light chain	K51D8
Flu19	Heavy chain	G51D8
Flu20	Light chain	K51G10
Flu20	Heavy chain	G51G10
Flu21	Light chain	L49A9
Flu21	Heavy chain	G49A9

Table 27 - antibodies used in staph inhibition assay

Antibody	Chain Type	Clone
S6	Light chain	K3E11
S6	Heavy chain	G3E11
S9	Light chain	L4D6
S9	Heavy chain	G4D6
LC1	Light chain	K8B8
LC1	Heavy chain	G8B8

Table 28 - antibodies used in staph IP

Antibody	Chain Type	Clone
S1	Light chain	K3G4
S1	Heavy chain	G3G4
S2	Light chain	K4C4
S2	Heavy chain	G4C4

S3	Light chain	K15C6
S3	Heavy chain	G15C6
S4	Light chain	K15G1
S4	Heavy chain	G15G1
S5	Light chain	K17C3
S5	Heavy chain	G17C3
S6	Light chain	K3E11
S6	Heavy chain	G3E11
S7	Light chain	L4B8
S7	Heavy chain	G4B8
S8	Light chain	L4D2
S8	Heavy chain	G4D2
S9	Light chain	L4D6
S9	Heavy chain	G4D6
S10	Light chain	L4F4
S10	Heavy chain	G4F4
S11	Light chain	L15D1
S11	Heavy chain	G15D1
S12	Light chain	L17C6
S12	Heavy chain	G17C6
S13	Light chain	L17C9
S13	Heavy chain	G17C9
Flu14	Light chain	L51A6
Flu14	Heavy chain	G51A6

Table 29 - Antibody used in staph mass spec		
Antibody	Chain Type	Clone
S4	Light chain	K15G1
S4	Heavy chain	G15G1

Table 30.	
Name	Sequence
RT oligo	CACGACCGGTGCTCGATTTAGTTAATTAA[sampleID]AGCGATCGCTGGG
RT oligo'	CTAAATCGAGCACCGGTCGTGTGGG
Fwd Primer (forkappa chain)	CGATTGGAGGGCGTTATCCAC
Fwd Primer (for lambda chain)	TYTGTGGGACTTCCACTGCTC

Table 31.	
Name	Sequence
RT oligo	CACGACCGGTGCTCGATTTAGTTAATTAA[sample-ID]AGCGATCGCTGGG
Overlap-extension	CGTATCGCTCCTAGGAGCGATACGCACGACCGGTGCTCGATTTAG

primer	
LC Primer (for kappa chain)	CGATTGGAGGGCGTTATCCAC
LC Primer (for lambda chain)	TYTGTGGGACTTCCACTGCTC
HC Primer	TCTTGTCCACCTTGGTGTTGCTG

Table 32.	
Name	Sequence
RT oligo	CGTATCGCTCCTAGGAGCGATACGTTAATTAA[sample-ID]AGCGATCGCTGGG
LC Primer (for kappa chain)	CGATTGGAGGGCGTTATCCAC
LC Primer (for lambda chain)	TYTGTGGGACTTCCACTGCTC
HC Primer	TCTTGTCCACCTTGGTGTTGCTG

Table 33.	
Name	Sequence
Univ_seq_2	AACGCGTGACGAGAGACTGACAG
VK	ATGAGGSTCCCYGCTCAGCTGCTG G
VL	GGTCCTGGGCCCAGTCTGCCCTG
IgKC_v3_barcoded	AGGCCCTTACGACTGCGTCTTG AACAATAC CAGATGGCGGGAAGATGAAGAC
IgLC_v5_barcoded	AGGCCCTTACGACTGCGTCTTGAACAATAC CTCCCGGGTAGAAGTCAC
Fixed_PCR3	AGGCCCTTACGACTGCGTCTTG

Table 34. Co-expressed genes associated with B cell differentiation into memory B cells, short-live plasma cell, long-lived plasma cells and antibody secreting cells.

Generation of memory B cells	Generation of short-lived plasma cells	Generation of long-lived plasma cells	Antibody-secreting cells	
PAX-5	Blimp-1	Blimp-1	SLC7A7	CD36
Microphthalmia-assoc. transcription factor (MITF)	X-box binding protein 1 (XBP-1)	X-box binding protein 1 (XBP-1)	IL6R	BCL2L1
			RPN2	IL21R
	IRF-4	IRF-4	PDIA4	IKZF1
		BCMA	IG HD	BAC 2

Table 35. Co-expressed genes associated with T cell differentiation into Treg, Th1, Th2, Th17 cells				
Generation of Th1	Generation of Th2	Generation of Th17	Generation of Tregs	
T-bet	Gata-3	RORγt	FoxP3	GITR

Table 36. Co-expressed genes associated with plasmablast homing to specific tissues.		
Homing of plasmablasts to the small intesting	Homing of plasmablasts to mucosal tissues	Homing of plasmablasts to ski
CCR9	CCR10	cutaneous lymphocyte-associated antigen (CLA)
α4β7		

GENERAL REFERENCES

Burbelo, P. D., S. K. Browne, et al. (2010). "Anti-cytokine autoantibodies are associated with opportunistic infection in patients with thymic neoplasia." Blood **116**(23): 4848-4858.

Hua, J., K. Kirou, et al. (2006). "Functional assay of type I interferon in systemic lupus erythematosus plasma and association with anti-RNA binding protein autoantibodies." Arthritis Rheum **54**(6): 1906-1916.

May, L. T., R. Neta, et al. (1993). "Antibodies chaperone circulating IL-6. Paradoxical effects of anti-IL-6 "neutralizing" antibodies in vivo." J Immunol **151**(6): 3225-3236.

Mostbock, S. (2009). "Cytokine/Antibody complexes: an emerging class of immunostimulants." Curr Pharm Des **15**(7): 809-825.

Robinson, W. H., C. DiGennaro, et al. (2002). "Autoantigen microarrays for multiplex characterization of autoantibody responses." Nat Med **8**(3): 295-301.

Watanabe, M., K. Uchida, et al. (2007). "Anti-cytokine autoantibodies are ubiquitous in healthy individuals." FEBS Lett **581**(10): 2017-2021.

Wildbaum, G., M. A. Nahir, et al. (2003). "Beneficial autoimmunity to proinflammatory mediators restrains the consequences of self-destructive immunity." Immunity **19**(5): 679-688.

Wrammert, J., K. Smith, et al. (2008). "Rapid cloning of high-affinity human monoclonal antibodies against influenza virus." Nature **453**(7195): 667-671.

CLAIMS

1. A composition comprising a polynucleotide, wherein the polynucleotide comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises at least one identification region, and wherein the first region is coupled to the second region.

2. The composition of claim 1, wherein the variable immunoglobulin region comprises a VDJ region of an IgG immunoglobulin nucleotide sequence isolated from an activated human B cell greater than or equal to 8 μm in diameter, and wherein the 5' end of the immunoglobulin region is coupled to the 3' end of the identification region.

3. The composition of any one of claims 1-2, wherein the composition is comprised in a clonal family.

4. The composition of any one of claims 1-3, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is an activated B cell.

5. The composition of any one of claims 1-4, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasmablast.

6. The composition of any one of claims 1-5, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single B cell.

7. The composition of any one of claims 1-6, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell.

8. The composition of any one of claims 1-7, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell located in the blood of a subject.

9. The composition of any one of claims 1-8, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a human activated B cell.

10. The composition of any one of claims 1-9, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a memory B cell.

11. The composition of any one of claims 1-10, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasma cell.

12. The composition of any one of claims 1-11, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is an antigen–specific B cell.

13. The composition of any one of claims 1-12, wherein the immunoglobulin region is isolated from a mammalian B cell.

14. The composition of any one of claims 1-13, wherein the immunoglobulin region is isolated from a human B cell.

15. The composition of any one of claims 1-14, wherein the immunoglobulin region is isolated from a mouse B cell.

16. The composition of any one of claims 1-15, wherein the immunoglobulin region is isolated from a B cell from a subject with a disease or condition of interest.

17. The composition of any one of claims 1-16, wherein the immunoglobulin region is isolated from a B cell from a subject recovering from or recovered from a disease or condition of interest.

18. The composition of any one of claims 1-17, wherein the immunoglobulin region is isolated from a B cell from a subject administered with at least one antigen of interest.

19. The composition of any one of claims 1-18, wherein the immunoglobulin region is isolated from a B cell from a subject administered with at least one antigen of interest and an adjuvant.

20. The composition of any one of claims 1-19, wherein the immunoglobulin region is isolated from a B cell located in the blood of a subject.

21. The composition of any one of claims 1-20, wherein the immunoglobulin region is isolated from a B cell located in the bone marrow of a subject.

22. The composition of any one of claims 1-21, wherein the immunoglobulin region is isolated from a B cell located in the spleen of a subject.

23. The composition of any one of claims 1-22, wherein the immunoglobulin region is isolated from a B cell located in at least one lymph node of a subject.

24. The composition of any one of claims 1-23, wherein the immunoglobulin region is isolated from a B cell located in lymphoid tissue of a subject.

25. The composition of any one of claims 1-24, wherein the immunoglobulin region is isolated from a B cell located in the gut of a subject.

26. The composition of any one of claims 1-25, wherein the immunoglobulin region is isolated from an activated B cell that is about 8-20 μm in diameter.

27. The composition of any one of claims 1-26, wherein the immunoglobulin region is isolated from an activated B cell that is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or greater than 20 μm in diameter.

28. The composition of any one of claims 1-27, wherein the immunoglobulin region is isolated from an activated B cell that is about 60, 70, 80, 90, 100, 120, 130, 140, 150, 200, 250, 300, 350, or greater than 350 μm^2 in area.

29. The composition of any one of claims 1-28, wherein the immunoglobulin region is isolated from an activated B cell that is about 250, 268, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, or greater than 4000 μm^3 in volume.

30. The composition of any one of claims 1-29, wherein the immunoglobulin region is isolated from an activated B cell that has a diameter of 10% or greater in size than the median diameter of a control resting B cell.

31. The composition of any one of claims 1-30, wherein the immunoglobulin region is isolated from an activated B cell that has a diameter of 15% or greater in size than the median diameter of a control resting B cell.

32. The composition of any one of claims 1-31, wherein the immunoglobulin region is isolated from an activated B cell that has a diameter of 20% or greater in size than the median diameter of a control resting B cell.

33. The composition of any one of claims 1-32, wherein the immunoglobulin region is isolated from an activated B cell capable of secreting immunoglobulin.

34. The composition of any one of claims 1-33, wherein the immunoglobulin region is isolated from a B cell in the gap 1 (G1), synthesis (S), gap 2 (G2), or mitosis (M) phase of the cell cycle.

35. The composition of any one of claims 1-34, wherein the immunoglobulin region is isolated from a B cell is not in the gap 0 (G0) phase of the cell cycle.

36. The composition of any one of claims 1-35, wherein the immunoglobulin region is isolated from a B cell characterized as having an FSC greater than 1.2x of the FSC mean of resting B lymphocytes by flow cytometry.

37. The composition of any one of claims 1-36, wherein the immunoglobulin region is isolated from a B cell characterized as having an FSC mean between 0.7 - 1.15x of the FSC mean of human monocytes by flow cytometry.

38. The composition of any one of claims 1-37, wherein the immunoglobulin region is isolated from a single CD19 positive B cell.

39. The composition of any one of claims 1-38, wherein the immunoglobulin region is isolated from a single CD38 positive B cell.

40. The composition of any one of claims 1-39, wherein the immunoglobulin region is isolated from a single CD27 positive B cell.

41. The composition of any one of claims 1-40, wherein the immunoglobulin region is isolated from a single CD20 negative B cell.

42. The composition of any one of claims 1-41, wherein the immunoglobulin region is isolated from a single $CD19^{+}CD20^{-}CD27^{+}CD38^{hi}$ B cell.

43. The composition of any one of claims 1-42, wherein the 5' end of the immunoglobulin region is coupled to the 3' end of the identification region.

44. The composition of any one of claims 1-43, wherein the variable immunoglobulin region comprises a VDJ region of an immunoglobulin nucleotide sequence.

45. The composition of any one of claims 1-44, wherein the variable immunoglobulin region comprises a VJ region of an immunoglobulin nucleotide sequence.

46. The composition of any one of claims 1-45, wherein the variable immunoglobulin region comprises a V, D, and/or J region of an immunoglobulin nucleotide sequence.

47. The composition of any one of claims 1-46, wherein the variable immunoglobulin region comprises a heavy and/or light chain of an immunoglobulin nucleotide sequence.

48. The composition of any one of claims 1-47, wherein the variable immunoglobulin region comprises an IgG, IgM, IgD, IgE, or IgA immunoglobulin sequence.

49. The composition of any one of claims 1-48, wherein the variable immunoglobulin region comprises a human IgG1, IgG2, IgG3, or IgG4 immunoglobulin sequence.

50. The composition of any one of claims 1-49, wherein the variable immunoglobulin region comprises a mouse IgG1, IgG2a, IgG2b, or IgG3 immunoglobulin sequence.

51. The composition of any one of claims 1-50, wherein the immunoglobulin region is about 200-2000 nucleotides in length.

52. The composition of any one of claims 1-51, wherein the immunoglobulin region is less than 200, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, or greater than 2000 nucleotides in length.

53. The composition of any one of claims 1-52, wherein the identification region comprises a plurality of identification regions.

54. The composition of any one of claims 1-53, wherein the identification region comprises a plurality of identification regions, and wherein each identification region in the plurality has a distinct sequence.

55. The composition of any one of claims 1-54, wherein the identification region comprises at least one sample identification region and at least one plate identification region.

56. The composition of any one of claims 1-55, wherein the identification region comprises a sequence distinct from the sequence of the immunoglobulin region.

57. The composition of any one of claims 1-56, wherein the identification region is about 2 - 100 nucleotides in length.

58. The composition of any one of claims 1-57, wherein the identification region is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, or greater than 100 nucleotides in length.

59. The composition of any one of claims 1-58, wherein the identification region is 2 – 1,000 nucleotides in length.

60. The composition of any one of claims 1-59, wherein the identification region is equal to or greater than 100 nucleotides in length.

61. The composition of any one of claims 1-60, wherein the identification region comprises a contiguous non-coding nucleotide sequence.

62. The composition of any one of claims 1-61, wherein the identification region comprises a non-coding nucleotide sequence.

63. The composition of any one of claims 1-62, wherein the identification region comprises a non-contiguous, non-coding nucleotide sequence.

64. The composition of any one of claims 1-63, wherein the length of the sequence of the identification region is less than the length of the sequence of the immunoglobulin region.

65. The composition of any one of claims 1-64, further comprising a third region, wherein the third region comprises an adapter region.

66. The composition of any one of claims 1-65, further comprising a third region, wherein the third region comprises an adapter region, and wherein the third region is located between the first region and the second region.

67. The composition of any one of claims 1-66, further comprising a third region, wherein the third region comprises an adapter region, and wherein the adapter region comprises at least one G nucleotide located at its 3' end.

68. The composition of any one of claims 1-67, wherein the identification region is 2-100 nucleotides long and has a sequence distinct from the immunoglobulin region sequence, and wherein the adaptor region comprises at least one G nucleotide at its 3' end and is located 3' of the sample identification region and 5' of the immunoglobulin region, and wherein the immunoglobulin variable region has undergone hypermutation and differs from the germline sequence of a naïve B cell.

69. The composition of any one of claims 1-68, wherein the composition is present in a container.

70. The composition of any one of claims 1-69, wherein a plurality of the compositions are present in a container.

71. The composition of any one of claims 1-70, wherein a plurality of the compositions are present in a single well of a single plate comprising a plurality of wells.

72. The composition of any one of claims 1-71, wherein the composition is in a library of compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide comprising a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, wherein the nucleotide sequence of each identification region o is distinct from the nucleotide sequence of the other identification regions present in the library, and wherein the last nucleotide sequences of a plurality of variable immunoglobulin regions in the library share at least 80-99% sequence identity.

73. The composition of any one of claims 1-72, wherein the composition is comprised in a library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide,

wherein the polynucleotide comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, and wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in each separate container in the library

74. A polynucleotide composition library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide, wherein the polynucleotide comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, and wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in each separate container in the library.

75. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide of the plurality is present in a separate container, wherein each polynucleotide of the plurality comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in the library, and wherein at least two variable immunoglobulin regions in the plurality share at least 80-99% sequence identity.

76. A polynucleotide library comprising a clonal family of polynucleotides, wherein each polynucleotide in the family comprises a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in the family, and wherein each of the variable immunoglobulin regions in the family exhibit at least 80-99% sequence identity.

77. The library of claim 76, wherein the library comprises a plurality of clonal families.

78. A clonal family of immunoglobulin sequences wherein each sequence in the family is coupled to an identification region.

79. The clonal family of claim 78, wherein each identification region is distinct from the other identification regions.

80. The clonal family of any one of claims 78-79, wherein the immunoglobulin sequences comprise heavy chain immunoglobulin sequences.

81. The clonal family of any one of claims 78-80, wherein the immunoglobulin sequences comprise light chain immunoglobulin sequences.

82. The clonal family of any one of claims 78-81, wherein the immunoglobulin sequences comprise heavy chain and light chain immunoglobulin sequences.

83. The clonal family of any one of claims 78-82, wherein one or more of the identification regions comprise a light chain immunoglobulin sequence.

84. The clonal family of any one of claims 78-83, wherein one or more of the identification regions comprise a heavy chain immunoglobulin sequence.

85. A set of two or more of the clonal family of any one of claims 78-84.

86. A set of 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more of the clonal family of any one of claims 78-85.

87. A clonal family of immunoglobulin sequences wherein each sequence in the family is operatively coupled to at least one contiguous nucleotide sequence.

88. The clonal family of claim 87, wherein the immunoglobulin sequences comprise heavy chain immunoglobulin sequences and the at least one contiguous nucleotide sequence comprises a light chain immunoglobulin sequence.

89. The clonal family of claim 87, wherein the immunoglobulin sequences comprise light chain immunoglobulin sequences and the at least one contiguous nucleotide sequence comprises a heavy chain immunoglobulin sequence.

90. A method of producing a clonal family of immunoglobulin sequences comprising obtaining a plurality of immunoglobulin sequences each having V, D, and/or J regions and each coupled to an identification region; and grouping two or more sequences from the plurality to produce the clonal family wherein each sequence in the clonal family is a mutated version of the same germline immunoglobulin sequence having a V, D, and/or J region or the germline immunoglobulin sequence having the V, D, and/or J region.

91. The method of claim 90, wherein each identification region is distinct from the other identification regions

92. A method of producing a clonal family of immunoglobulin sequences comprising obtaining a plurality of immunoglobulin sequences each having V, D, and/or J regions and each coupled to an identification region, and wherein each identification region is distinct from the other identification regions; removing one or more identification regions; and grouping two or more sequences from the plurality to produce the clonal family wherein each sequence in the clonal family is a mutated version of the same germline immunoglobulin sequence having a V, D, and/or J region or the germline immunoglobulin sequence having the V, D, and/or J region.

93. A method of identifying a second cDNA coupled to a first identification region comprising selecting a first cDNA coupled to the first identification region and identifying the second cDNA based on the shared identity of the identification region coupled to each cDNA.

94. A method of producing a 3' tail on a second nucleotide sequence comprising obtaining a first nucleotide sequence and contacting the first nucleotide sequence with a thermal stable RNase H^- reverse transcriptase having template switching activity at less than 50°C, wherein the contacting produces the 3' tail and the second nucleotide sequence.

95. The method of claim 94, wherein the first nucleotide sequence is contacted at about less than 50, 49, 48, 47, 46, 45 ,44, 43, 42, or less than 42°C.

96. The method of any one of claims 94-95, wherein the first nucleotide sequence is contacted at 42°C.

97. The method of any one of claims 94-96, wherein the first nucleotide sequence is contacted at 45.5°C.

98. The method of any one of claims 94-97, wherein the transcriptase is a Moloney Murine Leukemia Virus (MMLV) RNase H^- reverse transcriptase.

99. The method of any one of claims 94-98, wherein the transcriptase is SuperScript III.

100. A method for determining the naturally occurring sequence of a first sequence of interest comprising obtaining a plurality of sequences related to the first sequence and each coupled to a first identification region, wherein each first identification region is identical, and wherein one or more of the sequences in the plurality is distinct from the naturally

occurring sequence; and comparing the sequences in the plurality to determine the naturally occurring sequence of the first sequence of interest.

101. The method of claim 100, wherein the plurality of sequences comprise immunoglobulin sequences.

102. The method of any one of claims 100-101, wherein plurality of sequences comprise immunoglobulin sequences.

103. The method of any one of claims 100-102, wherein plurality of sequences comprise immunoglobulin sequences.

104. The method of any one of claims 100-103, wherein plurality of sequences are each coupled to a second identification region and each second identification region is identical.

105. The method of any one of claims 100-104, wherein the first sequence of interest is an immunoglobulin sequence.

106. The method of any one of claims 100-105, wherein the plurality of sequences are immunoglobulin sequences.

107. A composition comprising a polynucleotide comprising a first region and a second region, wherein the first region comprises a B cell-derived variable immunoglobulin region and the second region comprises an identification region, and wherein the first region is coupled to the second region.

108. A polynucleotide composition library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide comprising a B cell-derived variable immunoglobulin region and an identification region, wherein the variable immunoglobulin region is coupled to the identification region, wherein the nucleotide sequence of each identification region is distinct from the nucleotide sequence of the other identification regions present in each separate container in the library.

109. A method for producing a polynucleotide composition, comprising:

obtaining a polynucleotide comprising a first region, wherein the first region comprises an expressed B cell variable immunoglobulin region associated with a subject; and

generating the polynucleotide composition comprising the first region and a second region by coupling the first region to the second region, wherein the second region comprises an identification region.

110. The method of claim 109, wherein obtaining the polynucleotide comprises obtaining a B cell associated with the subject and processing the cell to prepare the polynucleotide.

111. The method of any one of claims 109-110, wherein obtaining the polynucleotide comprises receiving the polynucleotide directly or indirectly from a third party that has processed a B cell associated with the subject to prepare the polynucleotide.

112. The method of any one of claims 109-111, wherein obtaining the polynucleotide comprises receiving the polynucleotide directly or indirectly from a third party that has solubilized a B cell associated with the subject to prepare the polynucleotide.

113. The method of any one of claims 109-112, wherein obtaining the polynucleotide comprises obtaining a B cell using a flow cytometer.

114. The method of any one of claims 109-113, wherein obtaining the polynucleotide comprises obtaining a B cell using a microfluidic device.

115. The method of any one of claims 109-114, wherein the variable immunoglobulin region comprises a VDJ region of an IgG immunoglobulin nucleotide sequence is isolated from an activated human B cell greater than or equal to 8 μm in diameter, and wherein the 5' end of the immunoglobulin region is coupled to the 3' end of the identification region.

116. The method of any one of claims 109-115, wherein the composition is comprised in a clonal family.

117. The method of any one of claims 109-116, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is an activated B cell.

118. The method of any one of claims 109-117, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasmablast.

119. The method of any one of claims 109-118, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single B cell.

120. The method of any one of claims 109-119, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell.

121. The method of any one of claims 109-120, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell located in the blood of a subject.

122. The method of any one of claims 109-121, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a human activated B cell.

123. The method of any one of claims 109-122, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a memory B cell.

124. The method of any one of claims 109-123, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasma cell.

125. The method of any one of claims 109-124, wherein the immunoglobulin region is isolated from a B cell, and wherein the B cell is an antigen–specific B cell.

126. The method of any one of claims 109-125, wherein the immunoglobulin region is isolated from a mammalian B cell.

127. The method of any one of claims 109-126, wherein the immunoglobulin region is isolated from a human B cell.

128. The method of any one of claims 109-127, wherein the immunoglobulin region is isolated from a mouse B cell.

129. The method of any one of claims 109-128, wherein the immunoglobulin region is isolated from a B cell from a subject with a disease or condition of interest.

130. The method of any one of claims 109-129, wherein the immunoglobulin region is isolated from a B cell from a subject recovering from or recovered from a disease or condition of interest.

131. The method of any one of claims 109-130, wherein the immunoglobulin region is isolated from a B cell from a subject administered with an antigen of interest.

132. The method of any one of claims 109-131, wherein the immunoglobulin region is isolated from a B cell from a subject administered with an antigen of interest and an adjuvant.

133. The method of any one of claims 109-132, wherein the immunoglobulin region is isolated from a B cell located in the blood of a subject.

134. The method of any one of claims 109-133, wherein the immunoglobulin region is isolated from a B cell located in the bone marrow of a subject.

135. The method of any one of claims 109-134, wherein the immunoglobulin region is isolated from a B cell located in the spleen of a subject.

136. The method of any one of claims 109-135, wherein the immunoglobulin region is isolated from a B cell located in at least one lymph node of a subject.

137. The method of any one of claims 109-136, wherein the immunoglobulin region is isolated from a B cell located in lymphoid tissue of a subject.

138. The method of any one of claims 109-137, wherein the immunoglobulin region is isolated from a B cell located in the gut of a subject.

139. The method of any one of claims 109-138, wherein the immunoglobulin region is isolated from an activated B cell that is about 8-20 μm in diameter.

140. The method of any one of claims 109-139, wherein the immunoglobulin region is isolated from an activated B cell that is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or greater than 20 μm in diameter.

141. The method of any one of claims 109-140, wherein the immunoglobulin region is isolated from an activated B cell that is about 60, 70, 80, 90, 100, 120, 130, 140, 150, 200, 250, 300, 350, or greater than 350 μm^2 in area.

142. The method of any one of claims 109-141, wherein the immunoglobulin region is isolated from an activated B cell that is about 250, 268, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, or greater than 4000 μm^3 in volume.

143. The method of any one of claims 109-142, wherein the immunoglobulin region is isolated from an activated B cell that has a diameter of 10% or greater in size than the median diameter of a control resting B cell.

144. The method of any one of claims 109-143, wherein the immunoglobulin region is isolated from an activated B cell that has a diameter of 15% or greater in size than the median diameter of a control resting B cell.

145. The method of any one of claims 109-144, wherein the immunoglobulin region is isolated from an activated B cell that has a diameter of 20% or greater in size than the median diameter of a control resting B cell.

146. The method of any one of claims 109-145, wherein the immunoglobulin region is isolated from an activated B cell capable of secreting immunoglobulin.

147. The method of any one of claims 109-146, wherein the immunoglobulin region is isolated from a B cell in the gap 1 (G1), synthesis (S), gap 2 (G2), or mitosis (M) phase of the cell cycle.

148. The method of any one of claims 109-147, wherein the immunoglobulin region is isolated from a B cell is not in the gap 0 (G0) phase of the cell cycle.

149. The method of any one of claims 109-148, wherein the immunoglobulin region is isolated from a B cell characterized as having an FSC mean greater than 1.2x of the FSC mean of resting B lymphocytes by flow cytometry.

150. The method of any one of claims 109-149, wherein the immunoglobulin region is isolated from a B cell characterized as having an FSC between 0.7 - 1.15x of the FSC mean of human monocytes by flow cytometry.

151. The method of any one of claims 109-150, wherein the immunoglobulin region is isolated from a single CD19 positive B cell.

152. The method of any one of claims 109-151, wherein the immunoglobulin region is isolated from a single CD38 positive B cell.

153. The method of any one of claims 109-152, wherein the immunoglobulin region is isolated from a single CD27 positive B cell.

154. The method of any one of claims 109-153, wherein the immunoglobulin region is isolated from a single CD20 negative B cell.

155. The method of any one of claims 109-154, wherein the immunoglobulin region is isolated from a single $CD19^{+}CD20^{-}CD27^{+}CD38^{hi}$ B cell.

156. The method of any one of claims 109-155, wherein the 5' end of the immunoglobulin region is coupled to the 3' end of the identification region.

157. The method of any one of claims 109-156, wherein the variable immunoglobulin region comprises a VDJ region of an immunoglobulin nucleotide sequence.

158. The method of any one of claims 109-157, wherein the variable immunoglobulin region comprises a VJ region of an immunoglobulin nucleotide sequence.

159. The method of any one of claims 109-158, wherein the variable immunoglobulin region comprises a V, D, and/or J region of an immunoglobulin nucleotide sequence.

160. The method of any one of claims 109-159, wherein the variable immunoglobulin region comprises a heavy and/or light chain of an immunoglobulin nucleotide sequence.

161. The method of any one of claims 109-160, wherein the variable immunoglobulin region comprises an IgG, IgM, IgD, IgE, or IgA immunoglobulin sequence.

162. The method of any one of claims 109-161, wherein the variable immunoglobulin region comprises a human IgG1, IgG2, IgG3, or IgG4 immunoglobulin sequence.

163. The method of any one of claims 109-162, wherein the variable immunoglobulin region comprises a mouse IgG1, IgG2a, IgG2b, or IgG3 immunoglobulin sequence.

164. The method of any one of claims 109-163, wherein the immunoglobulin region is about 200-2000 nucleotides in length.

165. The method of any one of claims 109-164, wherein the immunoglobulin region is less than 200, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, or greater than 2000 nucleotides in length.

166. The method of any one of claims 109-165, wherein the identification region comprises a plurality of identification regions.

167. The method of any one of claims 109-166, wherein the identification region comprises a plurality of identification regions, and wherein each identification region in the plurality has a distinct sequence.

168. The method of any one of claims 109-167, wherein the identification region comprises at least one sample identification region and at least one plate identification region.

169. The method of any one of claims 109-168, wherein the identification region comprises a sequence distinct from the sequence of the immunoglobulin region.

170. The method of any one of claims 109-169, wherein the identification region is about 2 - 100 nucleotides in length.

171. The method of any one of claims 109-170, wherein the identification region is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, 100, or greater than 100 nucleotides in length.

172. The method of any one of claims 109-171, wherein the identification region is 2 – 1,000 nucleotides in length.

173. The method of any one of claims 109-172, wherein the identification region is equal to or greater than 100 nucleotides in length.

174. The method of any one of claims 109-173, wherein the identification region comprises a contiguous non-coding nucleotide sequence.

175. The method of any one of claims 109-174, wherein the identification region comprises a non-coding nucleotide sequence.

176. The method of any one of claims 109-175, wherein the identification region comprises a non-contiguous, non-coding nucleotide sequence.

177. The method of any one of claims 109-176, wherein the length of the sequence of the identification region is less than the length of the sequence of the immunoglobulin region.

178. The method of any one of claims 109-177, further comprising a third region, wherein the third region comprises an adapter region.

179. The method of any one of claims 109-178, further comprising a third region, wherein the third region comprises an adapter region, and wherein the third region is located between the first region and the second region.

180. The method of any one of claims 109-179, further comprising a third region, wherein the third region comprises an adapter region, and wherein the adapter region comprises at least one G nucleotide located at its 3' end.

181. The method of any one of claims 109-180, wherein the identification region is 2-100 nucleotides long and has a sequence distinct from the immunoglobulin region sequence, and wherein the adaptor region comprises at least one G nucleotide at its 3' end and is located 3' of the sample identification region and 5' of the immunoglobulin region, and wherein the immunoglobulin variable region has undergone hypermutation and differs from the germline sequence of a naïve B cell.

182. The method of any one of claims 109-181, wherein the composition is present in a container.

183. The method of any one of claims 109-182, wherein a plurality of the compositions are present in a container.

184. The method of any one of claims 109-183, wherein a plurality of the compositions are present in a single well of a single plate comprising a plurality of wells.

185. The method of any one of claims 109-184, wherein the composition is in a library of compositions, wherein each composition is present in a separate container, wherein each composition comprises a polynucleotide comprising a first region and a second region, wherein the first region comprises an expressed B cell variable immunoglobulin region and the second region comprises an identification region, wherein the first region is coupled to the second region, wherein the nucleotide sequence of each identification region o is distinct from the nucleotide sequence of the other identification regions present in the library, and wherein the nucleotide sequence of a plurality of variable immunoglobulin regions in the library share at least 80-99% sequence identity.

186. A method for producing a polynucleotide composition, comprising:

obtaining a B cell associated with a subject;

isolating polynucleotides from the cell comprising an expressed B cell variable immunoglobulin region; and

generating the polynucleotide composition comprising the variable immunoglobulin region and an identification region by coupling the variable immunoglobulin region to the identification region.

187. A method for producing a polynucleotide composition, comprising:

obtaining a polynucleotide comprising a B cell-derived variable immunoglobulin region associated with a subject; and

generating the polynucleotide composition comprising the variable immunoglobulin region and an identification region by coupling the variable immunoglobulin region to the identification region.

188. The method of claim 187, wherein obtaining the polynucleotide comprises obtaining a B cell and processing the cell to prepare the polynucleotide.

189. The method of claim 187, wherein obtaining the polynucleotide comprises receiving the polynucleotide directly or indirectly from a third party that has processed a B cell to prepare the polynucleotide.

190. A method for producing two or more polynucleotide compositions, comprising:

obtaining a polynucleotide library comprising a plurality of polynucleotides associated with a plurality of samples obtained from one or more subjects, wherein one or more polynucleotides comprises an expressed B cell variable immunoglobulin region, wherein each sample is associated with a B cell, and wherein each polynucleotide associated with each sample is present in a separate container; and

generating two or more polynucleotide compositions each comprising a polynucleotide from the plurality of polynucleotides and an identification region by coupling the polynucleotide to the identification region, wherein the sequence of each identification region is distinct from the sequence of the identification regions coupled to the other polynucleotides in the library.

191. The method of claim 190, wherein obtaining the polynucleotide library comprises obtaining a plurality of B cells and processing the cells to prepare the polynucleotide library.

192. The method of claim 190, wherein obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has processed a plurality of B cells to prepare the polynucleotide library.

193. A method for producing two or more polynucleotide compositions, comprising:

obtaining a polynucleotide library comprising a plurality of polynucleotides associated with a plurality of samples obtained from one or more subjects, wherein one or more polynucleotides comprises a B cell-derived variable immunoglobulin region, and wherein each polynucleotide associated with each sample is present in a separate container; and

generating two or more polynucleotide compositions each comprising a polynucleotide from the plurality of polynucleotides and an identification region by coupling the polynucleotide to the identification region, wherein the sequence of each identification region is distinct from the sequence of the identification regions coupled to the other polynucleotides in the library.

194. The method of claim 193, wherein obtaining the polynucleotide library comprises obtaining a plurality of B cells and processing the cells to prepare the polynucleotide library.

195. The method of claim 193, wherein obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has processed a plurality of B cells to prepare the polynucleotide library.

196. A polynucleotide composition library comprising a plurality of polynucleotide compositions, wherein each composition is present in a separate container, wherein each composition comprises a single sample-derived cDNA region comprising the nucleotide C at the 3' end of the cDNA region and a sample identification-adapter region comprising a sample identification region coupled to an adapter region, wherein the nucleotide sequence of the sample identification region of each sample identification-adapter region is distinct from the nucleotide sequence of the sample identification region of the other sample identification-adapter regions present in each separate container in the library, wherein the adapter region comprises the nucleotide G at the 3' end of the adapter region, and wherein the sample identification-adapter region is attached to the cDNA region by binding between the C and G.

197. The library of claim 196, wherein the cDNA region comprises an RNA polynucleotide hybridized to a DNA polynucleotide.

198. The library of claim 196 or claim 197, wherein the cDNA region comprises an mRNA polynucleotide hybridized to a cDNA polynucleotide.

199. The library of any one of claims 196-198, wherein the cDNA region comprises at least one C at the 3' end and wherein the adapter region comprises at least one G at the 3' end.

200. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a sample identification region, an adapter region, and a single sample-derived cDNA region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the cDNA region is coupled to the 3' end of the adapter region, wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample, and wherein the sample identification region is double-stranded.

201. The library of claim 200, wherein each polynucleotide comprises a plurality of sample identifications regions.

202. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the amplicon region is operatively coupled to the adapter region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

203. The library of claim 202, wherein the 5' end of the amplicon region is coupled to the 3' end of the adapter region, wherein the universal primer region comprises the sequence CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

204. The library of any one of claims 202-203, wherein the sequence of the universal primer region is not fully complementary to any human gene exon, and wherein the universal primer region has minimal secondary structure that does not interfere with the adapter region.

205. The library of any one of claims 202-204, wherein the universal primer region is the sequence CACGACCGGTGCTCGATTTAG.

206. The library of any one of claims 202-205, wherein the amplicon region comprises a cDNA region comprising a cDNA nucleotide sequence.

207. The library of any one of claims 202-206, wherein the sequence of the sample identification region of each polynucleotide from a first single sample differs by at least 1 nucleotide from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

208. The library of any one of claims 202-207, wherein the sequence of the sample identification region of each polynucleotide from a first single sample differs by at least 2 nucleotides from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

209. The library of any one of claims 202-208, wherein the sample identification region is at least 10 nucleotides in length.

210. The library of any one of claims 202-209, wherein the sample identification region is at least 1 nucleotide in length.

211. The library of any one of claims 202-210, wherein the sequence of each sample identification region is selected from Tables 2 and 7.

212. The library of any one of claims 202-211, wherein the sequence of the adapter region comprises at least one G nucleotide at its 3' end.

213. The library of any one of claims 202-212, wherein the amplicon region comprises an immunoglobulin heavy chain amplicon sequence, an immunoglobulin light chain amplicon sequence, a T cell receptor alpha amplicon sequence, or a T cell receptor beta amplicon sequence.

214. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-3', wherein A is a universal primer region, wherein B is a sample identification region, wherein C is an adapter region, wherein D is an amplicon region from a single sample, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

215. A polynucleotide comprising a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the amplicon region is operatively coupled to the adapter region.

216. The polynucleotide of claim 215, wherein the 5' end of the amplicon region is coupled to the 3' end of the adapter region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

217. A polynucleotide comprising the sequence 5'-A-B-C-D-3', wherein A is a universal primer region, wherein B is a sample identification region, wherein C is an adapter region, and wherein D is an amplicon region from a single sample.

218. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a first plate identification region, a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first plate identification region is operatively coupled to the universal primer region, wherein the amplicon region is operatively coupled to the adapter region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

219. The library of claim 218, wherein the sequence of the first plate identification region of each polynucleotide from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples.

220. The library of any one of claims 218-219, wherein the sequence of the first plate identification region of each polynucleotide from the first set of single samples differs by at least 1 nucleotide from the sequence of the first plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples.

221. The library of any one of claims 218-220, wherein the sequence of the first plate identification region of each polynucleotide from the first set of single samples differs by at least 2 nucleotides from the sequence of the first plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples.

222. The library of any one of claims 218-221, wherein the first plate identification region is at least 10 nucleotides in length.

223. The library of any one of claims 218-222, wherein the sequence of the first plate identification region is selected from Tables 3 and 6.

224. The library of any one of claims 218-223, wherein the 3' end of the first plate identification region is coupled to the 5' end of the universal primer region, wherein the 5' end of the amplicon region is coupled to the 3' end of the adapter region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, wherein the adapter region comprises at least one G, wherein each polynucleotide further comprises a second plate identification region, a first sequencing region, and a second sequencing region, wherein the 5' end of the second plate identification region is coupled to the 3' end of the amplicon region, wherein the 3' end of the first sequencing region is coupled to the 5' end of the first plate identification region, and wherein the 5' end of the second sequencing region is coupled to the 3' end of the second plate identification region.

225. The library of any one of claims 218-224, wherein the sequence of the second plate identification region is identical to the sequence of the first plate identification region on each polynucleotide.

226. The library of any one of claims 218-225, wherein the sequence of the second plate identification region of each polynucleotide from a first set of single samples is distinct from the sequence of the second plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples.

227. The library of any one of claims 218-226, wherein the sequence of the second plate identification region of each polynucleotide from the first set of single samples differs by at least 1 nucleotide from the sequence of the second plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples.

228. The library of any one of claims 218-227, wherein the sequence of the second plate identification region of each polynucleotide from the first set of single samples differs by at least 2 nucleotides from the sequence of the second plate identification region of the other polynucleotides in the library from one or more single sample sets distinct from the first set of single samples.

229. The library of any one of claims 218-228, wherein the second plate identification region is at least 10 nucleotides in length.

230. The library of any one of claims 218-229, wherein the sequence of the second plate identification region is selected from Tables 3 and 6.

231. The library of any one of claims 218-230, wherein the first sequencing region comprises GAGAGACTGACAGCGTATCGCCTCCCTCGCGCCATCAG.

232. The library of any one of claims 218-231, wherein the second sequencing region comprises CTATGCGCCTTGCCAGCCCGCTCAG.

233. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-E-3', wherein A is a plate identification region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, wherein E is an amplicon region from a single sample, and wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

234. A polynucleotide comprising a first plate identification region, a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first plate identification region is operatively coupled to the universal primer region, and wherein the amplicon region is operatively coupled to the adapter region.

235. The polynucleotide of claim 234, wherein the 3' end of the first plate identification region is coupled to the 5' end of the universal primer region, wherein the 5' end of the amplicon region is coupled to the 3' end of the adapter region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, wherein the adapter region comprises at least one G, wherein each polynucleotide further comprises a second plate identification region, a first sequencing region, and a second sequencing region, wherein the 5' end of the second plate identification region is coupled to the 3' end of the amplicon region, wherein the 3' end of the first sequencing region is coupled to the 5' end of the first plate identification region, and wherein the 5' end of the second sequecing region is coupled to the 3' end of the second plate identification region.

236. A polynucleotide comprising the sequence 5'-A-B-C-D-E-3', wherein A is a plate identification region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, and wherein E is an amplicon region from a single sample.

237. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a first restriction site region, a universal primer region, a sample identification region, an adapter region, an amplicon region from a single sample, and a second restriction site region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first restriction site region is operatively coupled to the universal primer region, wherein the amplicon region is operatively coupled to the adapter region, wherein the second restriction site region is operatively coupled to the amplicon region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

238. The library of claim 237, wherein the first restriction site region comprises one or more restriction sites.

239. The library of any one of claims 237-238, wherein the first restriction site region comprises one or more restriction sites selected from the group consisting of: NheI, XhoI, BstBI, EcoRI, SacII, BbvCI, PspXI, AgeI, ApaI, KpnI, Acc65I, XmaI, BstEII, DraIII, PacI, FseI, AsiSI and AscI.

240. The library of any one of claims 237-239, wherein the second restriction site region comprises one or more restriction sites.

241. The library of any one of claims 237-240, wherein the second restriction site region comprises one or more restriction sites selected from the group consisting of: NheI, XhoI, BstBI, EcoRI, SacII, BbvCI, PspXI, AgeI, ApaI, KpnI, Acc65I, XmaI, BstEII, DraIII, PacI, FseI, AsiSI and AscI.

242. The library of any one of claims 237-241, wherein the 3' end of the first restriction site region is coupled to the 5' end of the universal primer region, wherein the 3'

end of the adapter region is coupled to the 5' end of the amplicon region, wherein the 3' end of the amplicon region is coupled to the 5' end of the second restriction site region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

243. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-E-F-3', wherein A is a first restriction site region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, wherein E is an amplicon region from a single sample, wherein F is a second restriction site region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

244. A polynucleotide for insertion into a vector, comprising a first restriction site region, a universal primer region, a sample identification region, an adapter region, an amplicon region from a single sample, and a second restriction site region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first restriction site region is operatively coupled to the universal primer region, wherein the amplicon region is operatively coupled to the adapter region, and wherein the second restriction site region is operatively coupled to the amplicon region.

245. The polynucleotide of claim 244, wherein the 3' end of the first restriction site region is coupled to the 5' end of the universal primer region, wherein the 3' end of the adapter region is coupled to the 5' end of the amplicon region, wherein the 3' end of the amplicon region is coupled to the 5' end of the second restriction site region, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

246. A polynucleotide for insertion in a vector, comprising the sequence 5'-A-B-C-D-E-F-3', wherein A is a first restriction site region, wherein B is a universal primer region, wherein C is a sample identification region, wherein D is an adapter region, wherein E is an amplicon region from a single sample, and wherein F is a second restriction site region.

247. A polynucleotide adapter molecule, comprising a universal primer region, a sample identification region, and an adapter region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, and wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region.

248. The molecule of claim 247, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, and wherein the adapter region comprises at least one G.

249. A polynucleotide primer, comprising a universal primer region and a plate identification region, and wherein the 3' end of the plate identification region is coupled to the 5' end of the universal primer region.

250. The primer of claim 247, wherein the universal primer region comprises CACGACCGGTGCTCGATTTAG, wherein the primer further comprises a sequencing region, and wherein the 3' end of the sequencing region is coupled to the 5' end of the plate identification region.

251. A vector comprising the polynucleotide of any one of claims 1-250.

252. The vector of claim 251, wherein the vector comprises a plurality of polynucleotides.

253. The vector of any one of claims 251-252, wherein the vector is selected from the group consisting of: pEE6.4 and pEE12.4.

254. An isolated host cell comprising the vector of any one of claims claim 251-253 or the polynucleotide of any one of claims 1-250.

255. The host cell of claim 254, wherein the host cell is selected from the group consisting of: CHO cells, CHO-K1 cells, CHO-S cells, NS0 cells, CHO cells that are dhfr-, CHO-dhfr-, DUKX-B11 CHO cells, and DG44 CHO cells.

256. A method for producing one or more polynucleotides of interest, comprising:

obtaining a cDNA library comprising a plurality of cDNAs associated with a plurality of samples obtained from one or more subjects, wherein each cDNA is associated with a single sample in the plurality of samples, and wherein each cDNA associated with each sample is present in a separate container; and

adding an adapter molecule to the cDNA associated with each sample to produce the one or more polynucleotides of interest, wherein the adapter molecule comprises a sample

identification region and an adapter region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the sequence of the sample identification region of each adapter molecule is distinct from the sequence of the sample identification region of the other adapter molecules added to each cDNA in the library.

257. The method of claim 256, further comprising allowing the 3' end of the adapter region to attach to the 3' end of each cDNA in the library to produce the one or more polynucleotides of interest.

258. The method of any one of claims 256-257, wherein obtaining the cDNA library comprises obtaining the plurality of samples and processing the samples to prepare the cDNA library.

259. The method of any one of claims 256-258, wherein the adapter molecule further comprises a universal primer region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region.

260. The method of any one of claims 256-259, wherein each cDNA region comprises an mRNA polynucleotide hybridized to a cDNA polynucleotide.

261. The method of any one of claims 256-260, wherein each sample comprises a cell.

262. The method of any one of claims 256-261, wherein the cell is a B cell.

263. The method of any one of claims 256-262, wherein the B cell is a plasmablast, memory B cell, or a plasma cell.

264. The method of any one of claims 256-263, wherein each sample comprises a plurality of cells.

265. The method of of any one of claims 256-264, wherein obtaining the cDNA library comprises receiving the cDNA library directly or indirectly from a third party that has processed the plurality of samples to prepare the cDNA library.

266. The method of any one of claims 256-265, wherein the adaptor is added by annealing the adaptor to the '3 tail of a cDNA generated during a reverse transcription reaction.

267. The method of any one of claims 256-266, wherein each cDNA comprises at least one C nucleotide, wherein C is located at the 3' end of each cDNA, wherein the adapter

region comprises at least one G nucleotide, wherein G is located at the 3' end of the adapter region, and wherein the adapter region is attached to each cDNA via binding between the G and C.

268. The method of any one of claims 256-267, wherein the adapter molecule is single-stranded, and further comprising incorporating the adapter molecule into each cDNA by allowing an enzyme to make the adapter molecule double-stranded.

269. The method of any one of claims 256-268, wherein the adapter molecule is incorporated into each cDNA to produce the polynucleotide of interest by an MMLV H^- reverse transcriptase.

270. A method of producing one or more polynucleotides of interest for sequencing, comprising:

obtaining a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the amplicon region is operatively coupled to the adapter region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample; and

amplifying the polynucleotide library with a set of primers to produce the one or more polynucleotides of interest for sequencing, wherein the one or more polynucleotides of interest for sequencing comprises a first sequencing region, a first plate identification region, a universal primer region, a sample identification region, an adapter region, an amplicon region from a single sample, and a second sequencing region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the first plate identification region is operatively coupled to the universal primer region, wherein the amplicon region is operatively coupled to the adapter region, wherein the first sequencing region is located at the 5' end of the

polynucleotide of interest, and wherein the second sequencing region is located at the 3' end of the polynucleotide of interest.

271. The method of claim 270, further comprising sequencing the one or more polynucleotides of interest.

272. The method of claim 270, further comprising sequencing the one or more polynucleotides of interest with 454 sequencing.

273. The method of claim 270, further comprising sequencing the one or more polynucleotides of interest with SMRT sequencing.

274. The method of claim 270, further comprising sequencing the one or more polynucleotides of interest with SOLiD sequencing.

275. The method of claim 270, further comprising sequencing the one or more polynucleotides of interest with SOLEXA sequencing.

276. The method of claim 270, further comprising sequencing the one or more polynucleotides of interest with tSMS sequencing.

277. The method of claim 270, wherein the set of primers is selected from the primers shown in Tables 1 and 5.

278. The method of claim 270, wherein obtaining the polynucleotide library comprises preparing the polynucleotide library in a laboratory.

279. The method of claim 270, wherein obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has prepared the polynucleotide library.

280. A method for analyzing sequencing data, comprising:

obtaining a dataset associated with a plurality of polynucleotides, wherein the dataset comprises sequencing data for the plurality of polynucleotides, wherein each polynucleotide in the plurality of polynucleotides comprises a sample identification region, and wherein each sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the plurality of polynucleotides from one or more samples distinct from the first single sample; and

analyzing the dataset to match together polynucleotides with identical sample identification regions, wherein a match indicates that the polynucleotides originated from the same sample.

281. The method of claim 280, wherein each polynucleotide in the plurality of polynucleotides further comprises a first plate identification region, wherein each combination of each first plate identification region and sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the first plate identification region of each polynucleotide from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the plurality of polynucleotides from one or more single sample sets distinct from the first set of single samples, and further comprising analyzing the dataset to match together polynucleotides with identical first plate identification regions and identical sample identification regions, wherein a match between both regions indicates that the polynucleotides originated from the same sample.

282. The method of claim 280, wherein obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of polynucleotides to experimentally determine the dataset.

283. The method of claim 280, wherein obtaining the dataset comprises receiving the dataset directly or indirectly from a third party that has sequenced the plurality of polynucleotides to experimentally determine the dataset.

284. The method of claim 280, wherein the dataset is stored on an electronic storage medium.

285. The method of claim 280, wherein the single sample is a single cell.

286. The method of claim 280, wherein the single sample comprises a single cell.

287. The method of claim 280, wherein the single sample comprises a single B cell.

288. The method of claim 280, wherein the single sample comprises a plurality of B cells.

289. The method of claim 280, further comprising selecting one or more polynucleotides for cloning.

290. A method for identifying a second polynucleotide of interest based on selection of a first polynucleotide of interest, comprising:

obtaining a dataset associated with a plurality of polynucleotides, wherein the dataset comprises sequencing data for the plurality of polynucleotides, wherein each polynucleotide in the plurality of polynucleotides comprises a sample identification region, and wherein each sample identification region on each polynucleotide is unique to a single sample thereby associating each polynucleotide in the plurality of polynucleotides with a distinct single sample, wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the plurality of polynucleotides from one or more samples distinct from the first single sample; and

selecting a first polynucleotide of interest associated with a first single sample from the dataset and identifying a second polynucleotide of interest in the first single sample based on the sample identification region of the first polynucleotide of interest.

291. The method of claim 290, wherein each polynucleotide in the plurality of polynucleotides further comprises a first plate identification region, wherein each combination of each first plate identification region and sample identification region on each polynucleotide is unique to a single sample, wherein the sequence of the first plate identification region of each polynucleotide from a first set of single samples is distinct from the sequence of the first plate identification region of the other polynucleotides in the plurality of polynucleotides from one or more single sample sets distinct from the first set of single samples, and further comprising identifying a second polynucleotide of interest in the first single sample based on the sample identification region and first plate identification region of the first polynucleotide of interest.

292. The method of claim 291, wherein the first single sample comprises a B cell.

293. The method of claim 291, wherein the first single sample comprises a plurality of B cells.

294. The method of claim 291, wherein the first single sample comprises a B cell, wherein the first polynucleotide of interest comprises an antibody heavy chain nucleotide sequence, and wherein the second polynucleotide of interest comprises an antibody light chain nucleotide sequence.

295. The method of claim 291, wherein the first single sample comprises a B cell, wherein the first polynucleotide of interest comprises an antibody light chain nucleotide

sequence, and wherein the second polynucleotide of interest comprises an antibody heavy chain nucleotide sequence.

296. The method of claim 291, wherein obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of polynucleotides to experimentally determine the dataset.

297. The method of claim 291, wherein obtaining the dataset comprises receiving the dataset directly or indirectly from a third party that has sequenced the plurality of polynucleotides to experimentally determine the dataset.

298. The method of claim 291, wherein the dataset is stored on an electronic storage medium.

299. A method of producing one or more polynucleotides of interest for cloning, comprising:

obtaining a polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises a universal primer region, a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the amplicon region is operatively coupled to the adapter region, wherein the sequence of the universal primer region is substantially identical on each polynucleotide in the plurality of polynucleotides, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample; and

amplifying the polynucleotide library with a set of primers to produce the one or more polynucleotides of interest for cloning, wherein the one or more polynucleotides of interest for cloning comprises a first restriction site region, a universal primer region, a sample identification region, an adapter region, an amplicon region from a single sample, and a second restriction site region, wherein the 3' end of the universal primer region is coupled to the 5' end of the sample identification region, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, wherein the amplicon region is operatively coupled to the adapter region, wherein the first restriction

site region is located at the 5' end of the polynucleotide of interest, and wherein the second restriction site region is located at the 3' end of the polynucleotide of interest.

300. The method of claim 299, wherein obtaining the polynucleotide library comprises preparing the polynucleotide library in a laboratory.

301. The method of claim 299, wherein obtaining the polynucleotide library comprises receiving the polynucleotide library directly or indirectly from a third party that has prepared the polynucleotide library.

302. A method of producing a molecule of interest, comprising:

obtaining a host cell comprising a polynucleotide comprising a sample identification region, an adapter region, and an amplicon region from a single sample, wherein the 3' end of the sample identification region is coupled to the 5' end of the adapter region, and wherein the amplicon region is operatively coupled to the adapter region; and

culturing the host cell under conditions sufficient to produce the molecule of interest.

303. The method of claim 302, wherein obtaining the host cell comprises preparing the host cell comprising the polynucleotide in a laboratory.

304. The method of claim 302, wherein obtaining the host cell comprises receiving the host cell comprising the polynucleotide directly or indirectly from a third party that has prepared the host cell.

305. The method of claim 302, wherein the molecule of interest is a protein.

306. The method of claim 302, wherein the molecule of interest is an antibody.

307. The method of claim 302, wherein the molecule of interest is a human monoclonal antibody.

308. The method of claim 302, further comprising collecting the molecule of interest.

309. A kit, comprising a polynucleotide, a polynucleotide library, a vector, or a host cell of any one of claims 1-308 and instructions for use.

310. A method of linking and barcoding a plurality of non-contiguous polynucleotide sequences of interest, said method comprising:

(a) providing a plurality of cDNA molecules;

(b) physically linking cDNA molecules of interest; and

(c) adding a barcode sequence to the cDNAs of interest prior to, during, or after physical linkage.

311. The method of claim 310, wherein the physical linking is by ligation.

312. The method of claim 310, wherein the physical linking is by recombination.

313. The method of claim 310, wherein the physical linking comprises using an overlap-extension sequence.

314. The method of claim 310, wherein the barcode sequence is located at one or both of the ends of the physically linked cDNAs.

315. The method of claim 310, wherein the barcode sequence is located in between the physically linked cDNAs.

316. The method of claim 311, wherein the ligation is performed by annealing and ligation of compatible ends.

317. The method of claim 310, wherein the compatible ends are a restriction site.

318. The method of claim 310, wherein the ligation is performed by blunt end ligation.

319. The method of claim 310, wherein the overlap-extension sequence is added during the course of amplification using a primer comprising the overlap-extension tail.

320. The method of claim 310, wherein the overlap-extension sequence is added during the course of reverse transcription using a primer comprising the overlap-extension tail.

321. The method of claim 310, wherein the overlap-extension sequence is added by annealing an adaptor to the 3' tail of a cDNA generated during a reverse transcription reaction.

322. The method of claim 310, wherein the barcode sequence is added by ligation.

323. The method of claim 310, wherein the ligation is performed by annealing and ligation of compatible ends.

324. The method of claim 310, wherein the compatible ends are a restriction site.

325. The method of claim 310, wherein the ligation is performed by blunt end ligation of an adaptor comprising the barcode sequence.

326. The method of claim 310, wherein the barcode sequence is added during the course of an amplification reaction using a primer comprising the barcode sequence.

327. The method of claim 310, wherein the barcode sequence is added during the course of a reverse transcription reaction using a primer comprising the barcode sequence.

328. The method of claim 310, wherein the barcode sequence is added by annealing an adaptor to the 3' tail of a cDNA generated during a reverse transcription reaction.

329. The method of claim 310, wherein the '3 end of the cDNA comprises at least one C nucleotide, and wherein the 3' end of the adaptor comprises at least one G nucleotide, and wherein the adaptor is attached to each cDNA via binding between the C and G.

330. The method of claim 310, wherein the adaptor is single-stranded, and further comprising incorporating the adaptor into each cDNA by allowing an enzyme to make the adaptor double-stranded.

331. The method of claim 310, wherein the adaptor is incorporated into each cDNA by an MMLV H$^-$ reverse transcriptase.

332. The method of claim 310, wherein the overlap-extension sequence comprises a barcode sequence.

333. The method of claim 310, wherein the polynucleotide sequences of interest comprise antibody heavy and light chains.

334. The method of claim 310, further comprising (d) adding a sequencing region to the cDNAs of interest prior to, during, or after physical linkage.

335. The method of claim 310, wherein the sequencing region is added with an adaptor.

336. The method of claim 310, further comprising (e) sequencing of the physically linked cDNA molecules of interest using a NextGen sequencing platform.

337. The method of claim 310, wherein the NextGen sequencing platform is 454 sequencing.

338. The method of claim 310, wherein the NextGen sequencing platform is SMRT sequencing.

339. The method of claim 310, wherein the NextGen sequencing platform is SOLiD sequencing.

340. The method of claim 310, wherein the NextGen sequencing platform is SOLEXA sequencing.

341. The method of claim 310, wherein the NextGen sequencing platform is tSMS sequencing.

342. The method of claim 310, wherein the plurality of cDNA molecules is from single samples contained in a plate with at least 6 wells, at least 12 wells, at least 24 wells, at least 48 wells, at least 96 wells, at least 384 wells, at least 1536 wells, or more wells.

343. The method of claim 310, wherein the plurality of cDNA molecules is from single samples contained in at least one, two, three, four, five, six, seven, eight, nine, ten, twenty, thirty, forty, fifty, seventy five, one hundred, or more plates with at least 96 wells each.

344. A method of linking and barcoding a plurality of samples containing polynucleotide sequences of interest, said method comprising: (a) distributing the samples into a plurality of containers; (b) synthesizing polynucleotide sequences of interest using templates from the sample, wherein said synthesis results in the addition of a barcode sequence; and (c) effecting linkage of the polynucleotide sequences of interest synthesized in step (b).

345. The method of claim 344, wherein each sample comprises a cell.

346. The method of claim 344, wherein the cell is a B cell.

347. The method of claim 344, wherein the B cell is a plasmablast, memory B cell, or a plasma cell.

348. The method of claim 344, wherein each sample comprises a plurality of cells.

349. The method of claim 344, wherein the polynucleotide sequences of interest comprise antibody heavy and light chains.

350. The method of claim 344, wherein said synthesis comprises an RT-PCR amplification.

351. The method according to claim 344, wherein said RT-PCR amplification is performed in a single step.

352. The method of claim 344, wherein said linkage of the polynucleotide of interest is performed during the course of an RT-PCR amplification using an overlap-extension primer.

353. The method of claim 344, further comprising (d) adding a sequencing region to the polynucleotide sequences of interest prior to, during, or after barcode sequence addition or linkage.

354. The method of claim 344, wherein the sequencing region is added with an adaptor.

355. The method of claim 344, further comprising (e) sequencing of the linked polynucleotide sequences of interest using a NextGen sequencing platform.

356. The method of claim 344, wherein the NextGen sequencing platform is 454 sequencing.

357. The method of claim 344, wherein the NextGen sequencing platform is SMRT sequencing.

358. The method of claim 344, wherein the NextGen sequencing platform is SOLiD sequencing.

359. The method of claim 344, wherein the NextGen sequencing platform is SOLEXA sequencing.

360. The method of claim 344, wherein the NextGen sequencing platform is tSMS sequencing.

361. The method of claim 344, wherein the plurality of samples are single samples contained in a plate with at least 6 wells, at least 12 wells, at least 24 wells, at least 48 wells, at least 96 wells, at least 384 wells, at least 1536 wells, or more wells.

362. The method of claim 344, wherein the plurality of samples are single samples contained in at least one, two, three, four, five, six, seven, eight, nine, ten, twenty, thirty, forty, fifty, seventy five, one hundred, two hundred, five hundred or more plates with at least 96 wells each.

363. A method of linking and barcoding a plurality of non-contiguous polynucleotide sequences of interest, said method comprising: (a) distributing cells into a plurality of containers to obtain isolated one or more cells; (b) amplifying polynucleotide sequences of interest using templates from said isolated one or more cells, wherein said amplification

results in the addition of a barcode sequence; and (c) effecting linkage of the polynucleotide sequences of interest amplified in step (b).

364. The method of claim 363, wherein the nucleotide sequences of interest comprise antibody heavy and light chains.

365. The method of claim 363, wherein said amplification comprises an RT-PCR amplification.

366. The method according to claim 363, wherein said RT-PCR amplification is performed in a single step.

367. The method of claim 363, wherein said linkage of the nucleotide of interest is performed during the course of an RT-PCR amplification using an overlap-extension primer.

368. The method of claim 363, further comprising (d) adding a sequencing region to the polynucleotide sequences of interest prior to, during, or after barcode sequence addition or linkage.

369. The method of claim 363, wherein the sequencing region is added with an adaptor.

370. The method of claim 363, further comprising (e) sequencing of the linked polynucleotide sequences of interest using a NextGen sequencing platform.

371. The method of claim 363, wherein the NextGen sequencing platform is 454 sequencing.

372. The method of claim 363, wherein the NextGen sequencing platform is SMRT sequencing.

373. The method of claim 363, wherein the NextGen sequencing platform is SOLiD sequencing.

374. The method of claim 363, wherein the NextGen sequencing platform is SOLEXA sequencing.

375. The method of claim 363, wherein the NextGen sequencing platform is tSMS sequencing.

376. The method of claim 363, wherein the one or more cells are contained in a plate with at least 6 wells, at least 12 wells, at least 24 wells, at least 48 wells, at least 96 wells, at least 384 wells, at least 1536 wells, or more wells.

377. The method of claim 363, wherein the one or more cells are contained in at least one, two, three, four, five, six, seven, eight, nine, ten, twenty, thirty, forty, fifty, seventy five, one hundred, or more plates with at least 96 wells each.

378. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-3', wherein A is a sample identification region (barcode sequence), wherein B is a first cDNA region from a single sample, wherein C is a linker region, wherein D is a second cDNA region from the same single sample, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

379. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-D-3', wherein A is a first cDNA region from a single sample, wherein B is a linker region, wherein C is a second cDNA region from the same single sample, wherein D is a sample identification region (barcode sequence), and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

380. A polynucleotide library comprising a plurality of polynucleotides, wherein each polynucleotide comprises the sequence 5'-A-B-C-3', wherein A is a first cDNA region from a single sample, wherein B is a linker region comprising a sample identification region (barcode sequence), wherein C is a second cDNA region from the same single sample, and wherein the sequence of the sample identification region of each polynucleotide from a first single sample is distinct from the sequence of the sample identification region of the other polynucleotides in the library from one or more samples distinct from the first single sample.

381. The polynucleotide library of claim 378, wherein said first cDNA region comprises an antibody heavy chain and said second cDNA region comprises an antibody light chain.

382. The polynucleotide library of claim 378, wherein the library comprises at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 3000, at least 10,000, at least 30,000, at least 100,000, at least 300, 000, at least 1,000,000, at least 3,000,000, at least 10,000,000, at least 30,000,000, or more members.

383. The polynucleotide library of claim 378, wherein the library comprises at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 3000, at least 10,000, at least 30,000, or more genes of a cell sample's whole transcriptome.

384. The polynucleotide library of claim 378, wherein the library comprises at least 1, at least 2, at least 3, at least 10, at least 30, at least 100, at least 300, at least 1000, at least 10,000, at least 100,000, at least 1,000,000, at least 10,000,000, at least 100,000,000 or more of the different antibody species present in the blood of an individual.

385. The antibody species of claim 378, wherein the antibodies are expressed by plasmablasts, plasma cells, memory B cells, long-lived plasma cells, other B lineage cells or combinations thereof.