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 embodiments, 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 embodiments, the composition is comprised in a clonal family.

[0011] In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is an activated B cell. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasmablast. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single B cell. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a single activated B cell. In some embodiments, 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 embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a human activated B cell. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a memory B cell. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasma cell. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is an antigen–specific B cell. In some embodiments, the immunoglobulin region is isolated from a mammalian B cell. In some embodiments, the immunoglobulin region is isolated from a human B cell. In some embodiments, the immunoglobulin region is isolated from a mouse B cell. In some embodiments, the immunoglobulin region is isolated from a B cell from a subject with a disease or condition of interest. In some embodiments, 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 embodiments, the immunoglobulin region is isolated from a B cell from a subject administered with at least one antigen of interest. In some embodiments, 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 embodiments, the immunoglobulin region is isolated from a B cell located in the blood of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in the bone marrow of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in the spleen of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in at least one lymph node of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in lymphoid tissue of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in the gut of a subject. In some embodiments, the immunoglobulin region is isolated from an activated B cell that is about 8-20 μm in diameter. In some embodiments, 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 embodiments, 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

embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, the immunoglobulin region is isolated from an activated B cell capable of secreting immunoglobulin. In some embodiments, 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 embodiments, the immunoglobulin region is isolated from a B cell is not in the gap 0 (G0) phase of the cell cycle. In some embodiments, 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 embodiments, 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 embodiments, the immunoglobulin region is isolated from a single CD19 positive B cell. In some embodiments, the immunoglobulin region is isolated from a single CD38 positive B cell. In some embodiments, the immunoglobulin region is isolated from a single CD27 positive B cell. In some embodiments, the immunoglobulin region is isolated from a single CD20 negative B cell. In some embodiments, the immunoglobulin region is isolated from a single $CD19^{+}CD20^{-}CD27^{+}CD38^{hi}$ B cell.

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

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

the immunoglobulin region is about 200-2000 nucleotides in length. In some embodiments, 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 embodiments, the identification region comprises a plurality of identification regions. In some embodiments, the identification region comprises a plurality of identification regions, and wherein each identification region in the plurality has a distinct sequence. In some embodiments, the identification region comprises at least one sample identification region and at least one plate identification region. In some embodiments, the identification region comprises a sequence distinct from the sequence of the immunoglobulin region. In some embodiments, the identification region is about 2 - 100 nucleotides in length. In some embodiments, 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 embodiments, the identification region is 2 – 1,000 nucleotides in length. In some embodiments, the identification region is equal to or greater than 100 nucleotides in length. In some embodiments, the identification region comprises a contiguous non-coding nucleotide sequence. In some embodiments, the identification region comprises a non-coding nucleotide sequence. In some embodiments, the identification region comprises a non-contiguous, non-coding nucleotide sequence. In some embodiments, the length of the sequence of the identification region is less than the length of the sequence of the immunoglobulin region.

[0015] In some embodiments, a composition described herein can include a third region, wherein the third region comprises an adapter region. In some embodiments, the third region comprises an adapter region, and wherein the third region is located between the first region and the second region. In some embodiments, 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 embodiments, 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 embodiments, the composition is present in a container. In some embodiments, a plurality of the compositions are present in a container. In some embodiments, a plurality of the compositions are present in a single well of a single plate comprising a plurality of wells.

[0018] In some embodiments, 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 embodiments, 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.

[0020a] In one aspect, there is provided a polynucleotide library comprising a plurality of compositions, wherein the library comprises cDNAs encoding cognate pairs of immunoglobulin heavy and light chain variable regions and wherein:
each composition is present in a separate container, each composition comprises
(i) polynucleotides comprising cDNA molecules derived from a single plasmablast, that comprise
cDNA molecules that encode a cognate pair of immunoglobulin heavy chain variable region and immunoglobulin light chain variable region from the plasmablast, and
a sample identification region attached to the cDNA molecules, wherein
(ii) the nucleotide sequence of the sample identification region is distinct from the nucleotide sequence of the sample identification region of the other compositions present in other separate containers 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 embodiments, 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 embodiments, each identification region is distinct from the other identification regions. In some embodiments, the immunoglobulin sequences comprise heavy chain immunoglobulin sequences. In some embodiments, the immunoglobulin sequences comprise light chain immunoglobulin sequences. In some embodiments, the immunoglobulin sequences comprise heavy chain and light chain immunoglobulin sequences. In some embodiments, one or more of the identification regions comprise a light chain immunoglobulin sequence. In some embodiments, 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 embodiments, the immunoglobulin sequences comprise heavy chain immunoglobulin sequences and the at least one contiguous nucleotide sequence comprises a light chain immunoglobulin sequence. In some embodiments, 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 embodiments, 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 embodiments, 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 embodiments, the first nucleotide sequence is contacted at 42°C. In some embodiments, the first nucleotide sequence is contacted at 45.5°C. In some embodiments, the transcriptase is a Moloney Murine Leukemia Virus (MMLV) RNase H^- reverse transcriptase. In some embodiments, 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 embodiments, the plurality of sequences comprise immunoglobulin sequences. In some embodiments, the plurality of sequences comprise immunoglobulin sequences. In some embodiments, the plurality of sequences comprise immunoglobulin sequences. In some embodiments, the plurality of sequences are each coupled to a second identification region and each second identification region is identical. In some embodiments, the first sequence of interest is an immunoglobulin sequence. In some embodiments, 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 embodiments, obtaining the polynucleotide comprises obtaining a B cell associated with the subject and processing the cell to prepare the polynucleotide. In some embodiments, 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 embodiments, 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 embodiments, obtaining the polynucleotide comprises obtaining a B cell using a flow cytometer. In some embodiments, obtaining the polynucleotide comprises obtaining a B cell using a microfluidic device.

[0037] In some embodiments, 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 embodiments, the composition is comprised in a clonal family.

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

embodiments, 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 embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a human activated B cell. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a memory B cell. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is a plasma cell. In some embodiments, the immunoglobulin region is isolated from a B cell, and wherein the B cell is an antigen–specific B cell. In some embodiments, the immunoglobulin region is isolated from a mammalian B cell. In some embodiments, the immunoglobulin region is isolated from a human B cell. In some embodiments, the immunoglobulin region is isolated from a mouse B cell. In some embodiments, the immunoglobulin region is isolated from a B cell from a subject with a disease or condition of interest. In some embodiments, 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 embodiments, the immunoglobulin region is isolated from a B cell from a subject administered with at least one antigen of interest. In some embodiments, 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 embodiments, the immunoglobulin region is isolated from a B cell located in the blood of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in the bone marrow of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in the spleen of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in at least one lymph node of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in lymphoid tissue of a subject. In some embodiments, the immunoglobulin region is isolated from a B cell located in the gut of a subject. In some embodiments, the immunoglobulin region is isolated from an activated B cell that is about 8-20 μm in diameter. In some embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, the immunoglobulin region is isolated from an activated B cell capable of secreting immunoglobulin. In some embodiments, 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 embodiments, the immunoglobulin region is isolated from a B cell is not in the gap 0 (G0) phase of the cell cycle. In some embodiments, 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 embodiments, 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 embodiments, the immunoglobulin region is isolated from a single CD19 positive B cell. In some embodiments, the immunoglobulin region is isolated from a single CD38 positive B cell. In some embodiments, the immunoglobulin region is isolated from a single CD27 positive B cell. In some embodiments, the immunoglobulin region is isolated from a single CD20 negative B cell. In some embodiments, the immunoglobulin region is isolated from a single $CD19^{+}CD20^{-}CD27^{+}CD38^{hi}$ B cell.

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

[0040] In some embodiments, the variable immunoglobulin region comprises a VDJ region of an immunoglobulin nucleotide sequence. In some embodiments, the variable immunoglobulin region comprises a VJ region of an immunoglobulin nucleotide sequence. In some embodiments, the variable immunoglobulin region comprises a V, D, and/or J region of an immunoglobulin nucleotide sequence. In some embodiments, the variable immunoglobulin region comprises a heavy and/or light chain of an immunoglobulin nucleotide sequence. In some embodiments, the variable immunoglobulin region comprises an IgG, IgM, IgD, IgE, or IgA immunoglobulin sequence. In some embodiments, the variable immunoglobulin region comprises a human IgG1, IgG2, IgG3, or IgG4 immunoglobulin sequence. In some embodiments, the variable immunoglobulin region comprises a mouse IgG1, IgG2a, IgG2b, or IgG3 immunoglobulin sequence. In some embodiments, the immunoglobulin region is about 200-2000 nucleotides in length. In some embodiments, 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 embodiments, the identification region comprises a plurality of identification regions. In some embodiments, the identification region comprises a plurality of identification regions, and wherein each identification region in the plurality has a distinct sequence. In some embodiments, the

identification region comprises at least one sample identification region and at least one plate identification region. In some embodiments, the identification region comprises a sequence distinct from the sequence of the immunoglobulin region. In some embodiments, the identification region is about 2 - 100 nucleotides in length. In some embodiments, 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 embodiments, the identification region is 2 – 1,000 nucleotides in length. In some embodiments, the identification region is equal to or greater than 100 nucleotides in length. In some embodiments, the identification region comprises a contiguous non-coding nucleotide sequence. In some embodiments, the identification region comprises a non-coding nucleotide sequence. In some embodiments, the identification region comprises a non-contiguous, non-coding nucleotide sequence. In some embodiments, the length of the sequence of the identification region is less than the length of the sequence of the immunoglobulin region.

[0042] In some embodiments, a composition described herein can include a third region, wherein the third region comprises an adapter region. In some embodiments, the third region comprises an adapter region, and wherein the third region is located between the first region and the second region. In some embodiments, 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 embodiments, 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 embodiments, the composition is present in a container. In some embodiments, a plurality of the compositions are present in a container. In some embodiments, a plurality of the compositions are present in a single well of a single plate comprising a plurality of wells.

[0045] In some embodiments, 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 embodiments, 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 embodiments, obtaining the polynucleotide comprises obtaining a B cell and processing the cell to prepare the polynucleotide. In some embodiments, 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 embodiments, obtaining the polynucleotide library comprises obtaining a plurality of B cells and processing the cells to prepare the polynucleotide library. In some embodiments, 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 embodiments, obtaining the polynucleotide library comprises obtaining a plurality of B cells and processing the cells to prepare the polynucleotide library. In some embodiments, 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 embodiments, the cDNA region comprises an RNA polynucleotide hybridized to a DNA polynucleotide. In some embodiments, the cDNA region comprises an mRNA polynucleotide hybridized to a cDNA polynucleotide. In some embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, the universal primer region is the sequence CACGACCGGTGCTCGATTTAG. In some embodiments, the amplicon region comprises a cDNA region comprising a cDNA nucleotide sequence. In some embodiments, 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 embodiments, 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 embodiments, the sample identification region is at least 10 nucleotides in length. In some embodiments, the sample identification region is at least 1 nucleotide in length. In some embodiments, the sequence of each sample identification region is selected from Tables 2 and 7. In some embodiments, the sequence of the adapter region comprises at least one G nucleotide at its 3' end. In some embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, the first plate identification region is at least 10 nucleotides in length. In some embodiments, the sequence of the first plate identification region is selected from Tables 3 and 6. In some embodiments, 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 embodiments, the sequence of the second plate identification region is identical to the sequence of the first plate identification region on each polynucleotide. In some embodiments, 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 embodiments, 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 embodiments, 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 embodiments, the second plate identification region is at least 10 nucleotides in length. In some embodiments, the sequence of the second plate identification region is selected from Tables 3 and 6. In some embodiments, the first sequencing region comprises GAGAGACTGACAGCGTATCGCCTCCCTCGCGCCATCAG. In some embodiments, 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 embodiments, 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 embodiments, the first restriction site region comprises one or more restriction sites. In some embodiments, 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 embodiments, the second restriction site region comprises one or more restriction sites. In some embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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 embodiments, the vector comprises a plurality of polynucleotides. In some embodiments, 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 embodiments, 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 embodiments, 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 embodiments, obtaining the cDNA library comprises obtaining the plurality of samples and processing the samples to prepare the cDNA library. In some embodiments, 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 embodiments, each cDNA region comprises an mRNA polynucleotide hybridized to a cDNA polynucleotide. In some embodiments, each sample comprises a cell. In some embodiments, the cell is a B cell. In some embodiments, the B cell is a plasmablast, memory B cell, or a plasma cell. In some embodiments, each sample comprises a plurality of cells. In some embodiments, 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 embodiments, the adaptor is added by annealing the adaptor to the '3 tail of a cDNA generated during a reverse transcription reaction. In some embodiments, 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 embodiments, 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 embodiments, 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.

[0081a] In another aspect, there is provided a method for producing one or more polynucleotides of interest, comprising: obtaining the library as described herein; and amplifying the polynucleotide library with a set of primers to produce the one or more polynucleotides of interest, wherein the polynucleotide of interest is the cognate pair of immunoglobulin light chain and heavy chain variable regions.

[0082] In some embodiments, the method further includes sequencing the one or more polynucleotides of interest. In some embodiments, the method further includes sequencing the one or more polynucleotides of interest with 454 sequencing. In some embodiments, the method further includes sequencing the one or more polynucleotides of interest with SMRT sequencing. In some embodiments, the method further includes sequencing the one or more polynucleotides of interest with SOLiD sequencing. In some embodiments, the method further includes sequencing the one or more polynucleotides of interest with SOLEXA sequencing. In some embodiments, the method further includes sequencing the one or more polynucleotides of interest with tSMS sequencing. In some embodiments, the set of primers is selected from the primers shown in Tables 1 and 5. In some embodiments, obtaining the polynucleotide library comprises preparing the polynucleotide library in a laboratory. In some embodiments, 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 embodiments, 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 embodiments, obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of

polynucleotides to experimentally determine the dataset. In some embodiments, 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 embodiments, the dataset is stored on an electronic storage medium. In some embodiments, the single sample is a single cell. In some embodiments, the single sample comprises a single cell. In some embodiments, the single sample comprises a single B cell. In some embodiments, the single sample comprises a plurality of B cells. In some embodiments, 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 embodiments, 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 embodiments, the first single sample comprises a B cell. In some embodiments, the first single sample comprises a plurality of B cells. In some embodiments, 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 embodiments, 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 embodiments, obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of polynucleotides to experimentally determine the dataset. In some embodiments, 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 embodiments, 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 embodiments, obtaining the polynucleotide library comprises preparing the polynucleotide library in a laboratory. In some embodiments, 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 embodiments, obtaining the host cell comprises preparing the host cell comprising the polynucleotide in a laboratory. In some embodiments, 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 embodiments, the molecule of interest is a protein. In some embodiments, the molecule of interest is an antibody. In some embodiments, the molecule of interest is a human monoclonal antibody. In some embodiments, 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 embodiments, the physical linking is by ligation. In some embodiments, the physical linking is by recombination. In some embodiments, the physical linking comprises using an overlap-extension sequence. In some embodiments, the barcode sequence is located at one or both of the ends of the physically linked cDNAs. In some embodiments, the barcode sequence is located in between the physically linked cDNAs. In some embodiments the ligation is performed by annealing and ligation of compatible ends. In some embodiments, the compatible ends are a restriction site. In some embodiments, the ligation is performed by blunt end ligation. In some embodiments, the overlap-extension sequence is added during the course of amplification using a primer comprising the overlap-extension tail. In some embodiments, the overlap-extension sequence is added during the course of reverse transcription using a primer comprising the overlap-extension tail. In some embodiments, 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 embodiments, the barcode sequence is added by ligation. In some embodiments, the ligation is performed by annealing and ligation of compatible ends. In some embodiments, the compatible ends are a restriction site. In some embodiments, the ligation is performed by blunt end ligation of an adaptor comprising the barcode sequence. In some embodiments, the barcode sequence is added during the course of an amplification reaction using a primer comprising the barcode sequence. In some embodiments, the barcode sequence is added during the course of a reverse transcription reaction using a primer comprising the

barcode sequence. In some embodiments, the barcode sequence is added by annealing an adaptor to the 3' tail of a cDNA generated during a reverse transcription reaction. In some embodiments, 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 embodiments, 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 embodiments, the adaptor is incorporated into each cDNA by an MMLV H^- reverse transcriptase. In some embodiments, the overlap-extension sequence comprises a barcode sequence. In some embodiments, the polynucleotide sequences of interest comprise antibody heavy and light chains. In some embodiments, further comprising (d) adding a sequencing region to the cDNAs of interest prior to, during, or after physical linkage. In some embodiments, the sequencing region is added with an adaptor. In some embodiments, further comprising (e) sequencing of the physically linked cDNA molecules of interest using a NextGen sequencing platform. In some embodiments, the NextGen sequencing platform is 454 sequencing. In some embodiments, the NextGen sequencing platform is SMRT sequencing. In some embodiments, the NextGen sequencing platform is SOLiD sequencing. In some aembodiments, the NextGen sequencing platform is SOLEXA sequencing. In some embodiments, the NextGen sequencing platform is tSMS sequencing. In some embodiments, 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 embodiments, 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 embodiments, each sample comprises a cell. In some embodiments, the cell is a B cell. In some embodiments, the B cell is a plasmablast, memory B cell, or a plasma cell. In some embodiments, each sample comprises a plurality of cells. In some embodiments, the polynucleotide sequences of interest comprise antibody heavy and light chains. In some embodiments, said synthesis comprises an

RT-PCR amplification. In some embodiments, said RT-PCR amplification is performed in a single step. In some embodiments, said linkage of the polynucleotide of interest is performed during the course of an RT-PCR amplification using an overlap-extension primer. In some embodiments, 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 embodiments, the sequencing region is added with an adaptor. In some embodiments, further comprising (e) sequencing of the linked polynucleotide sequences of interest using a NextGen sequencing platform. In some embodiments, the NextGen sequencing platform is 454 sequencing. In some embodiments, the NextGen sequencing platform is SMRT sequencing. In some embodiments, the NextGen sequencing platform is SOLiD sequencing. In some embodiments, the NextGen sequencing platform is SOLEXA sequencing. In some embodiments, the NextGen sequencing platform is tSMS sequencing. In some embodiments, 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 embodiments, 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 embodiments, the nucleotide sequences of interest comprise antibody heavy and light chains. In some embodiments, said amplification comprises an RT-PCR amplification. In some embodiments, said RT-PCR amplification is performed in a single step. In some embodiments, said linkage of the nucleotide of interest is performed during the course of an RT-PCR amplification using an overlap-extension primer. In some embodiments, 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 embodiments, the sequencing region is added with an adaptor. In some embodiments, further comprising (e) sequencing of the linked polynucleotide sequences of interest using a NextGen sequencing platform. In some embodiments, the NextGen sequencing platform is 454 sequencing. In some embodiments, the NextGen sequencing platform is SMRT sequencing. In some embodiments, the NextGen sequencing platform is SOLiD sequencing. In some embodiments, the NextGen sequencing platform is SOLEXA sequencing. In some embodiments, the NextGen sequencing

platform is tSMS sequencing. In some embodiments, 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 embodiments, 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 embodiments, said first cDNA region comprises an antibody heavy chain and said second cDNA region comprises an antibody light chain. In some embodiments, 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 embodiments, 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 embodiments, 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 embodiments, 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

Claims

1. A polynucleotide library comprising a plurality of compositions, wherein the library comprises cDNAs encoding cognate pairs of immunoglobulin heavy and light chain variable regions and wherein:

each composition is present in a separate container, each composition comprises

(i) polynucleotides comprising cDNA molecules derived from a single plasmablast, that comprise

cDNA molecules that encode a cognate pair of immunoglobulin heavy chain variable region and immunoglobulin light chain variable region from the plasmablast, and

a sample identification region attached to the cDNA molecules, wherein

(ii) the nucleotide sequence of the sample identification region is distinct from the nucleotide sequence of the sample identification region of the other compositions present in other separate containers in the library.

2. The library of claim 1, wherein the cDNA molecules are attached to the sample identification region by an adapter region.

3. The library of claim 2, wherein the adapter region comprises the nucleotide G at its 3' end and the cDNA molecules comprise the nucleotide C at the 3' end.

4. The library of any one of claims 1- 3, wherein each composition further comprises a universal primer region attached to the sample identification region.

5. The library of claim 1, wherein the plasmablast is a murine B cell.

6. A method for producing one or more polynucleotides of interest, comprising: obtaining the library of any one of claims 2 to 5; and amplifying the polynucleotide library with a set of primers to produce the one or more polynucleotides of interest, wherein the polynucleotide of interest is the cognate pair of immunoglobulin light chain and heavy chain variable regions.