

[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.

[0050a] Disclosed 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 variable region polynucleotide, wherein each sample is associated with a cell, and wherein the polynucleotides associated with each sample are 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 associated with the other samples 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.

[0052a] Disclosed 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 variable region, and wherein the polynucleotides 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.

[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 CACGACCGGTGCTCGATTAG, 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.

[0079a] Disclosed is a method of 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 the cDNA associated with each sample is present in a separate container; (a) 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 the cDNA associated with the samples present in other separate containers; and (b) allowing the adapter region to attach to each cDNA in the library to produce the one or more polynucleotides of interest.

[0079b] Disclosed is a method for producing one or more polynucleotides of interest, comprising: (a) 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 (b) 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 the cDNA associated with the samples present in other separate containers.

[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] Disclosed is a method of producing one or more polynucleotides of interest for sequencing, comprising: a. 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 b. 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 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.

[0083a] Disclosed is a method for analyzing sequencing data, comprising: a. 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 b. analyzing the dataset to match together polynucleotides with the same sample identification region, 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.

[0085a] In one aspect, there is provided a method for identifying a second polynucleotide of interest based on selection of a first polynucleotide of interest, comprising: (a) 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 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 regions of the other polynucleotides in the plurality of polynucleotides from one or more samples distinct from the first single sample; wherein the first single sample comprises a B cell; and (b) selecting a first polynucleotide of interest associated with the 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, wherein the sample identification region of the first and second polynucleotides of interest are identical; and wherein (i) the first polynucleotide of interest comprises an antibody heavy chain nucleotide sequence, and the second polynucleotide of interest comprises an antibody light chain nucleotide sequence; or (ii) the first polynucleotide of interest comprises an antibody light chain nucleotide sequence, and the second polynucleotide of interest comprises an antibody heavy chain nucleotide sequence.

[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

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.

[00100a] Disclosed is a method of sequencing paired heavy and light chains from an individual B cell comprising the steps of: a. sorting individual B cells by flow cytometry; b. adding a barcode during reverse transcription; c. pooling the barcoded cDNA; and d. sequencing the pooled cDNAs.

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 1st 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 method for identifying a second polynucleotide of interest based on selection of a first polynucleotide of interest, comprising:

- (a) 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 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 regions of the other polynucleotides in the plurality of polynucleotides from one or more samples distinct from the first single sample; wherein the first single sample comprises a B cell; and
- (b) selecting a first polynucleotide of interest associated with the 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, wherein the sample identification region of the first and second polynucleotides of interest are identical; and wherein
 - (i) the first polynucleotide of interest comprises an antibody heavy chain nucleotide sequence, and the second polynucleotide of interest comprises an antibody light chain nucleotide sequence; or
 - (ii) the first polynucleotide of interest comprises an antibody light chain nucleotide sequence, and the second polynucleotide of interest comprises an antibody heavy chain nucleotide sequence.

2. A method according to claim 1, wherein the first single sample comprises a plurality of B cells.

3. A method according to claim 1, wherein the first single sample comprises a single B cell.

4. A method according to claim 1, 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.

5. A method according to claim 1, wherein the first polynucleotide of interest comprises an antibody light chain variable region nucleotide sequence, and wherein the second polynucleotide of interest comprises an antibody heavy chain variable nucleotide sequence.
6. A method according to claim 1, wherein the first polynucleotide of interest comprises an antibody heavy chain variable region nucleotide sequence, and wherein the second polynucleotide of interest comprises an antibody light chain variable region nucleotide sequence.
7. A method according to any preceding claim, wherein the polynucleotide is a cDNA.
8. A method according to any preceding claim, wherein obtaining the dataset comprises obtaining the plurality of polynucleotides and sequencing the plurality of polynucleotides to experimentally determine the dataset.
9. A method according to any preceding claim, 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.