

**(12) STANDARD PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

(11) Application No. **AU 2018281745 B2**

(54) Title  
**Sample indexing for single cells**

(51) International Patent Classification(s)  
**C12Q 1/6804** (2018.01) **C12Q 1/6869** (2018.01)

(21) Application No: **2018281745** (22) Date of Filing: **2018.03.27**

(87) WIPO No: **WO18/226293**

(30) Priority Data

| (31) Number       | (32) Date         | (33) Country |
|-------------------|-------------------|--------------|
| <b>62/645,703</b> | <b>2018.03.20</b> | <b>US</b>    |
| <b>62/578,957</b> | <b>2017.10.30</b> | <b>US</b>    |
| <b>62/554,425</b> | <b>2017.09.05</b> | <b>US</b>    |
| <b>62/532,971</b> | <b>2017.07.14</b> | <b>US</b>    |
| <b>62/532,949</b> | <b>2017.07.14</b> | <b>US</b>    |
| <b>62/532,905</b> | <b>2017.07.14</b> | <b>US</b>    |
| <b>62/515,285</b> | <b>2017.06.05</b> | <b>US</b>    |

(43) Publication Date: **2018.12.13**

(44) Accepted Journal Date: **2022.05.19**

(71) Applicant(s)  
**Becton, Dickinson and Company**

(72) Inventor(s)  
**Chang, Christina;Fan, Christina;Shum, Eleen;Martin, Jody;Bansal, Nidhanjali;Ghadiali, James;Lazaruk, Katherine;Lam, Gretchen Yinbon**

(74) Agent / Attorney  
**Pizzseys Patent and Trade Mark Attorneys Pty Ltd, GPO Box 1374, Brisbane, QLD, 4001, AU**

(56) Related Art  
**WO 2015/031691 A1**



## (51) International Patent Classification:

*C12Q 1/6804* (2018.01) *C12Q 1/6869* (2018.01)

## (21) International Application Number:

PCT/US2018/024602

## (22) International Filing Date:

27 March 2018 (27.03.2018)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

|            |                                |    |
|------------|--------------------------------|----|
| 62/515,285 | 05 June 2017 (05.06.2017)      | US |
| 62/532,905 | 14 July 2017 (14.07.2017)      | US |
| 62/532,949 | 14 July 2017 (14.07.2017)      | US |
| 62/532,971 | 14 July 2017 (14.07.2017)      | US |
| 62/554,425 | 05 September 2017 (05.09.2017) | US |
| 62/578,957 | 30 October 2017 (30.10.2017)   | US |
| 62/645,703 | 20 March 2018 (20.03.2018)     | US |

(71) Applicant: **BECTON, DICKINSON AND COMPANY**

[US/US]; 1 Becton Drive, Franklin Lakes, NJ 07417 (US).

(72) Inventors: **CHANG, Christina**; 4040 Campbell Avenue,Suite 110, Menlo Park, California 94025 (US). **FAN,****Christina**; 4040 Campbell Avenue, Suite 110, Menlo Park,California 94025 (US). **SHUM, Elen**; 4040 Campbell

Avenue, Suite 110, Menlo Park, California 94025 (US).

**MARTIN, Jody**; 4040 Campbell Avenue, Suite 110, MenloPark, California 94025 (US). **BANSAL, Nidhanjali**; 4040

Campbell Avenue, Suite 110, Menlo Park, California 94025

(US). **GHADIALI, James**; 4040 Campbell Avenue, Suite110, Menlo Park, California 94025 (US). **LAZARUK,****Katherine**; 4040 Campbell Avenue, Suite 110, Menlo Park,California 94025 (US). **LAM, Gretchen, Yinbon**; 4040

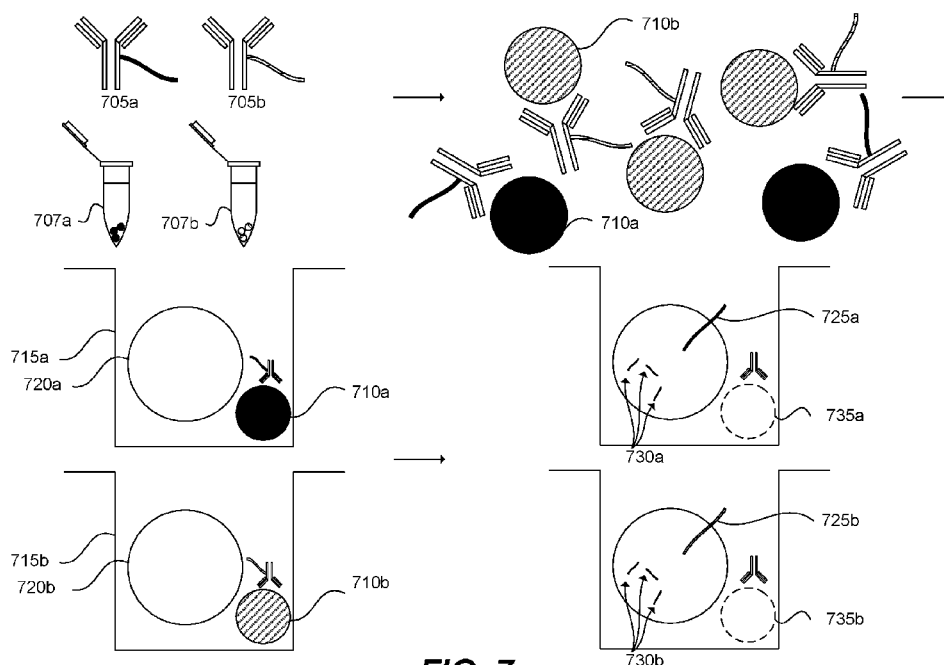
Campbell Avenue, Suite 110, Menlo Park, California 94025

(US).

(74) Agent: **LIU, Jing**; Knobbe, Martens, Olson & Bear, LLP,

2040 Main Street, 14th Floor, Irvine, California 92614 (US).

## (54) Title: SAMPLE INDEXING FOR SINGLE CELLS

**FIG. 7**

(57) **Abstract:** Disclosed herein include systems, methods, compositions, and kits for sample identification. A sample indexing composition can comprise, for example, a protein binding reagent associated with a sample indexing oligonucleotide. Different sample indexing compositions can include sample indexing oligonucleotides with different sequences. Sample origin of cells can be identified based on the sequences of the sample indexing oligonucleotides. Sample indexing oligonucleotides can be barcoded using barcoded and lengthened using daisy-chaining primers.

**(81) Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

**(84) Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

## SAMPLE INDEXING FOR SINGLE CELLS

### RELATED APPLICATIONS

[0001] The present application claims priority to U.S. Provisional Application No. 62/515,285, filed on June 5, 2017; U.S. Provisional Application No. 62/532,905, filed on July 14, 2017; U.S. Provisional Application No. 62/532,949, filed on July 14, 2017; U.S. Provisional Application No. 62/532,971, filed on July 14, 2017; U.S. Provisional Application No. 62/554,425, filed on September 5, 2017; U.S. Provisional Application No. 62/578,957, filed on October 30, 2017; and U.S. Provisional Application No. 62/645,703, filed on March 20, 2018. The content of each of these related applications is incorporated herein by reference in its entirety.

### REFERENCE TO SEQUENCE LISTING

[0002] The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing is provided as a file entitled Sequence\_Listing\_BDCRI\_033A.txt, created March 20, 2018, which is 3,181 bytes in size. The information in the electronic format of the Sequence Listing is incorporated herein by reference in its entirety.

### BACKGROUND

#### Field

[0003] The present disclosure relates generally to the field of molecular biology, for example identifying cells of different samples and detecting interactions between cellular components using molecular barcoding.

#### Description of the Related Art

[0004] Current technology allows measurement of gene expression of single cells in a massively parallel manner (e.g., >10000 cells) by attaching cell specific oligonucleotide barcodes to poly(A) mRNA molecules from individual cells as each of the cells is co-localized with a barcoded reagent bead in a compartment. Gene expression may affect protein expression. Protein-protein interaction may affect gene expression and protein expression. There is a need for systems and methods that can quantitatively analyze protein



expression, simultaneously measure protein expression and gene expression in cells, and determining protein-protein interactions in cells.

#### SUMMARY

[0005] Disclosed herein include methods for sample identification. In some embodiments, the method comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein the one or more cells comprises one or more antigen targets, wherein at least one sample indexing composition of the plurality of sample indexing compositions comprises two or more antigen binding reagents (e.g., protein binding reagents and antibodies), wherein each of the two or more antigen binding reagents is associated with a sample indexing oligonucleotide, wherein at least one of the two or more antigen binding reagents is capable of specifically binding to at least one of the one or more antigen targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides (e.g., identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide). The method can, for example, comprise removing unbound sample indexing compositions of the plurality of sample indexing compositions.

[0006] In some embodiments, the sample indexing sequence is 25-60 nucleotides in length (e.g., 45 nucleotides in length), about 128 nucleotides in length, or at least 128 nucleotides in length. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences. Sample indexing sequences of at least 100 or 1000 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences.

[0007] In some embodiments, the antigen binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the antigen binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the antigen binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

[0008] In some embodiments, at least one sample of the plurality of samples comprises a single cell. The at least one of the one or more antigen targets can be on a cell surface or inside of a cell. A sample of the plurality of samples can comprise a plurality of cells, a tissue, a tumor sample, or any combination thereof. The plurality of samples can comprise a mammalian sample, a bacterial sample, a viral sample, a yeast sample, a fungal sample, or any combination thereof.

[0009] In some embodiments, removing the unbound sample indexing compositions comprises washing the one or more cells from each of the plurality of samples with a washing buffer. The method can comprise lysing the one or more cells from each of the plurality of samples. The sample indexing oligonucleotide can be configured to be detachable or non-detachable from the antigen binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the antigen binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the antigen binding reagent by UV photocleaving, chemical treatment (e.g., using a reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0010] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of the cells of the plurality of samples. The sample indexing oligonucleotide can comprise a molecular label sequence, a poly(A) region, or a combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. A target binding region of the barcode can comprise the capture sequence. The target binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(A) tail. The

sample indexing oligonucleotide can comprise a molecular label.

**[0011]** In some embodiments, the antigen target is, or comprises, an extracellular protein, an intracellular protein, or any combination thereof. The antigen target can be, or comprise, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The antigen target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The antigen target can be selected from a group comprising 10-100 different antigen targets. The antigen binding reagent can be associated with two or more sample indexing oligonucleotides with an identical sequence. The antigen binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second antigen binding reagent not conjugated with the sample indexing oligonucleotide. The antigen binding reagent and the second antigen binding reagent can be identical.

**[0012]** In some embodiments, a barcode of the plurality of barcodes comprises a target binding region and a molecular label sequence. Molecular label sequences of at least two barcodes of the plurality of barcodes comprise different molecule label sequences. The barcode can comprise a cell label, a binding site for a universal primer, or any combination thereof. The target binding region can comprise a poly(dT) region.

**[0013]** In some embodiments, the plurality of barcodes is associated with a particle. At least one barcode of the plurality of barcodes can be immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or any combination thereof. The particle can be degradable. The particle can be a bead. The bead can be selected from the group consisting of streptavidin beads, agarose beads, magnetic beads, conjugated beads, protein A conjugated beads, protein G conjugated beads, protein A/G conjugated beads, protein L conjugated beads, oligo(dT) conjugated beads, silica beads, silica-like beads, anti-biotin microbead, anti-fluorochrome microbead, and any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, and any combination thereof. The

particle can comprise at least 10000 barcodes. In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000 or 10000 different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences.

**[0014]** In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

**[0015]** In some embodiments, the method comprises amplifying the plurality of barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR) at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. Obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide.

**[0016]** In some embodiments, identifying the sample origin of the at least one cell can comprise identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide. Barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides can comprise stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

**[0017]** In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets,

wherein each of the plurality of barcodes comprises a cell label, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets. Prior to obtaining the sequencing data of the plurality of barcoded targets, the method can comprise amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0018] In some embodiments, each of the plurality of sample indexing compositions comprises the antigen binding reagent. The sample indexing sequences of the sample indexing oligonucleotides associated with the two or more antigen binding reagents can be identical. The sample indexing sequences of the sample indexing oligonucleotides associated with the two or more antigen binding reagents can comprise different sequences. Each of the plurality of sample indexing compositions can comprise the two or more antigen binding reagents.

[0019] Disclosed herein include methods for sample identification. In some embodiments, the method comprise: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more cellular component targets, wherein at least one sample indexing composition of the plurality of sample indexing compositions comprises two or more cellular component binding reagents (e.g., antigen binding reagents or antibodies), wherein each of the two or more cellular component binding reagents is associated with a sample indexing oligonucleotide, wherein at least one of the two or more cellular component binding reagents is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least

two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides. The method can comprise removing unbound sample indexing compositions of the plurality of sample indexing compositions.

[0020] In some embodiments, the sample indexing sequence is 25-60 nucleotides in length (e.g., 45 nucleotides in length), about 128 nucleotides in length, or at least 128 nucleotides in length. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions comprise different sequences.

[0021] In some embodiments, the cellular component binding reagent comprises a cell surface binding reagent, an antibody, a tetramer, an aptamers, a protein scaffold, an integrin, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the cellular component binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the cellular component binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

[0022] In some embodiments, at least one sample of the plurality of samples comprises a single cell. The at least one of the one or more cellular component targets can be expressed on a cell surface. A sample of the plurality of samples can comprise a plurality of cells, a tissue, a tumor sample, or any combination thereof. The plurality of samples can comprise a mammalian sample, a bacterial sample, a viral sample, a yeast sample, a fungal sample, or any combination thereof.

[0023] In some embodiments, removing the unbound sample indexing

compositions comprises washing the one or more cells from each of the plurality of samples with a washing buffer. The method can comprise lysing the one or more cells from each of the plurality of samples. The sample indexing oligonucleotide can be configured to be detachable or non-detachable from the cellular component binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent by UV photocleaving, chemical treatment (e.g., using a reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0024] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of the cells of the plurality of samples. The sample indexing oligonucleotide can comprise a molecular label sequence, a poly(A) region, or a combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. A target binding region of the barcode can comprise the capture sequence. The target binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(A) tail. The sample indexing oligonucleotide can comprise a molecular label.

[0025] In some embodiments, the cellular component target is, or comprises, a carbohydrate, a lipid, a protein, an extracellular protein, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, an intracellular protein, or any combination thereof. The cellular component target can be selected from a group comprising 10-100 different cellular component targets. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with an identical sequence. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second cellular component binding reagent not conjugated with the sample indexing oligonucleotide. The cellular component binding reagent and the second cell binding reagent can be identical.

[0026] In some embodiments, a barcode of the plurality of barcodes comprises a

target binding region and a molecular label sequence. Molecular label sequences of at least two barcodes of the plurality of barcodes can comprise different molecule label sequences. The barcode can comprise a cell label, a binding site for a universal primer, or any combination thereof. The target binding region can comprise a poly(dT) region.

[0027] In some embodiments, the plurality of barcodes is enclosed in a particle. The particle can be a bead. At least one barcode of the plurality of barcodes can be immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or any combination thereof. The particle can be degradable. The bead can be selected from the group consisting of streptavidin beads, agarose beads, magnetic beads, conjugated beads, protein A conjugated beads, protein G conjugated beads, protein A/G conjugated beads, protein L conjugated beads, oligo(dT) conjugated beads, silica beads, silica-like beads, anti-biotin microbead, anti-fluorochrome microbead, and any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise at least 10000 barcodes. In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000 or 10000 different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences.

[0028] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0029] In some embodiments, the method comprises amplifying the plurality of barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying



the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR) at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. Obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide.

[0030] In some embodiments, identifying the sample origin of the at least one cell can comprise identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide. Barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides can comprise stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

[0031] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets. The method can comprise: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0032] In some embodiments, each of the plurality of sample indexing compositions comprises the cellular component binding reagent. The sample indexing

sequences of the sample indexing oligonucleotides associated with the two or more cellular component binding reagents can be identical. The sample indexing sequences of the sample indexing oligonucleotides associated with the two or more cellular component binding reagents can comprise different sequences. Each of the plurality of sample indexing compositions can comprise the two or more cellular component binding reagents.

[0033] Disclosed herein include methods for sample identification. In some embodiments, the method comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more cellular component targets, wherein at least one sample indexing composition of the plurality of sample indexing compositions comprises two or more cellular component binding reagents, wherein each of the two or more cellular component binding reagents is associated with a sample indexing oligonucleotide, wherein at least one of the two or more cellular component binding reagents is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one sample indexing oligonucleotide of the plurality of sample indexing compositions. The method can, for example, include removing unbound sample indexing compositions of the plurality of sample indexing compositions.

[0034] In some embodiments, the sample indexing sequence is 25-60 nucleotides in length (e.g., 45 nucleotides in length), about 128 nucleotides in length, or at least 128 nucleotides in length. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences. Sample indexing sequences of at least 100 or 1000 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences.

[0035] In some embodiments, the antigen binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the antigen binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The

chemical group can be reversibly or irreversibly attached to the antigen binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

[0036] In some embodiments, at least one sample of the plurality of samples comprises a single cell. The at least one of the one or more antigen targets can be expressed on a cell surface. A sample of the plurality of samples can comprise a plurality of cells, a tissue, a tumor sample, or any combination thereof. The plurality of samples can comprise a mammalian sample, a bacterial sample, a viral sample, a yeast sample, a fungal sample, or any combination thereof.

[0037] In some embodiments, removing the unbound sample indexing compositions comprises washing the one or more cells from each of the plurality of samples with a washing buffer. The method can comprise lysing the one or more cells from each of the plurality of samples. The sample indexing oligonucleotide can be configured to be detachable or non-detachable from the antigen binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the antigen binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the antigen binding reagent by UV photocleaving, chemical treatment (e.g., using a reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0038] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of the cells of the plurality of samples. The sample indexing oligonucleotide can comprise a molecular label sequence, a poly(A) region, or a combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. A target binding region of the barcode can comprise the capture sequence. The target binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(A) tail. The sample indexing oligonucleotide can comprise a molecular label.

[0039] In some embodiments, the antigen target is, or comprises, a carbohydrate, a lipid, a protein, an extracellular protein, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor,

an intracellular protein, or any combination thereof. The antigen target can be selected from a group comprising 10-100 different antigen targets. The antigen binding reagent can be associated with two or more sample indexing oligonucleotides with an identical sequence. The antigen binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second antigen binding reagent not conjugated with the sample indexing oligonucleotide. The antigen binding reagent and the second antigen binding reagent can be identical.

[0040] In some embodiments, identifying the sample origin of the at least one cell comprises: barcoding sample indexing oligonucleotides of the plurality of sample indexing compositions using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying the sample origin of the cell based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides.

[0041] In some embodiments, a barcode of the plurality of barcodes comprises a target binding region and a molecular label sequence. Molecular label sequences of at least two barcodes of the plurality of barcodes can comprise different molecule label sequences. The barcode can comprise a cell label, a binding site for a universal primer, or any combination thereof. The target binding region can comprise a poly(dT) region.

[0042] In some embodiments, the plurality of barcodes is immobilized on a particle. At least one barcode of the plurality of barcodes can be immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or any combination thereof. The particle can be degradable. The particle can be a bead. The bead can be selected from the group consisting of streptavidin beads, agarose beads, magnetic beads, conjugated beads, protein A conjugated beads, protein G conjugated beads, protein A/G conjugated beads, protein L conjugated beads, oligo(dT) conjugated beads, silica beads, silica-like beads, anti-biotin microbead, anti-fluorochrome microbead, and any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer,

titanium, latex, Sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise at least 10000 barcodes. In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000 or 10000 different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences.

**[0043]** In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes can comprise: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

**[0044]** In some embodiments, the method comprises amplifying the plurality of barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR) at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. Obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide.

**[0045]** In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides comprises stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides. Identifying the sample origin of the at least one cell can comprise identifying the presence or absence of the sample indexing sequence of at least one sample indexing oligonucleotide of the plurality of sample indexing compositions. Identifying the presence or absence of the sample indexing sequence can comprise: replicating the at least

one sample indexing oligonucleotide to generate a plurality of replicated sample indexing oligonucleotides; obtaining sequencing data of the plurality of replicated sample indexing oligonucleotides; and identifying the sample origin of the cell based on the sample indexing sequence of a replicated sample indexing oligonucleotide of the plurality of sample indexing oligonucleotides that correspond to the least one barcoded sample indexing oligonucleotide in the sequencing data.

[0046] In some embodiments, replicating the at least one sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides comprises: prior to replicating the at least one barcoded sample indexing oligonucleotide, ligating a replicating adaptor to the at least one barcoded sample indexing oligonucleotide. Replicating the at least one barcoded sample indexing oligonucleotide can comprise replicating the at least one barcoded sample indexing oligonucleotide using the replicating adaptor ligated to the at least one barcoded sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides.

[0047] In some embodiments, replicating the at least one sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides comprises: prior to replicating the at least one barcoded sample indexing oligonucleotide, contacting a capture probe with the at least one sample indexing oligonucleotide to generate a capture probe hybridized to the sample indexing oligonucleotide; and extending the capture probe hybridized to the sample indexing oligonucleotide to generate a sample indexing oligonucleotide associated with the capture probe. Replicating the at least one sample indexing oligonucleotide can comprise replicating the sample indexing oligonucleotide associated with the capture probe to generate the plurality of replicated sample indexing oligonucleotides.

[0048] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Identifying the sample origin of the at least one barcoded sample indexing oligonucleotide can comprise identifying the sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one

barcoded sample indexing oligonucleotide. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets. The method can comprise: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0049] In some embodiments, each of the plurality of sample indexing compositions comprises the antigen binding reagent. The sample indexing sequences of the sample indexing oligonucleotides associated with the two or more antigen binding reagents can be identical. The sample indexing sequences of the sample indexing oligonucleotides associated with the two or more antigen binding reagents can comprise different sequences. Each of the plurality of sample indexing compositions can comprise the two or more antigen binding reagents.

[0050] Disclosed herein includes a plurality of sample indexing compositions. Each of the plurality of sample indexing compositions can comprise two or more antigen binding reagents. Each of the two or more antigen binding reagents can be associated with a sample indexing oligonucleotide. At least one of the two or more antigen binding reagents can be capable of specifically binding to at least one antigen target. The sample indexing oligonucleotide can comprise a sample indexing sequence for identifying sample origin of one or more cells of a sample. Sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences.

[0051] In some embodiments, the sample indexing sequence comprises a nucleotide sequence of 25-60 nucleotides in length (e.g., 45 nucleotides in length), about 128 nucleotides in length, or at least 128 nucleotides in length. Sample indexing sequences of at

least 10, 100, or 1000 sample indexing compositions of the plurality of sample indexing compositions comprise different sequences.

**[0052]** In some embodiments, the antigen binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the antigen binding reagent through a linker. The at least one sample indexing oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the molecule of the antigen binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

**[0053]** In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of a species. The sample indexing oligonucleotide can comprise a molecular label sequence, a poly(A) region, or a combination thereof. In some embodiments, at least one sample of the plurality of samples can comprise a single cell, a plurality of cells, a tissue, a tumor sample, or any combination thereof. The sample can comprise a mammalian sample, a bacterial sample, a viral sample, a yeast sample, a fungal sample, or any combination thereof.

**[0054]** In some embodiments, the antigen target is, or comprises, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. The antigen target can be selected from a group comprising 10-100 different antigen targets. The antigen binding reagent can be associated with two or more sample indexing oligonucleotides with an identical sequence. The antigen binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second antigen binding reagent not conjugated with the sample indexing oligonucleotide. The antigen binding reagent and the second antigen binding reagent can be identical.

**[0055]** Disclosed herein include control particle compositions. In some embodiments, the control particle composition comprises a plurality of control particle oligonucleotides associated with a control particle, wherein each of the plurality of control particle oligonucleotides comprises a control barcode sequence and a poly(dA) region. At



least two of the plurality of control particle oligonucleotides can comprise different control barcode sequences. The control particle oligonucleotide can comprise a molecular label sequence. The control particle oligonucleotide can comprise a binding site for a universal primer.

[0056] In some embodiments, the control barcode sequence is at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The control particle oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or a combination thereof. The control barcode sequences of at least 5, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can be identical. The control barcode sequences of about 10, 100, 1000, or more of the plurality of control particle oligonucleotides can be identical. At least 3, 5, 10, 100, or more of the plurality of control particle oligonucleotides can comprise different control barcode sequences.

[0057] In some embodiments, the plurality of control particle oligonucleotides comprises a plurality of first control particle oligonucleotides each comprising a first control barcode sequence, and a plurality of second control particle oligonucleotides each comprising a second control barcode sequence, and wherein the first control barcode sequence and the second control barcode sequence have different sequences. The number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be about the same. The number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be different. The number of the plurality of first control particle oligonucleotides can be at least 2 times, 10 times, 100 times, or more greater than the number of the plurality of second control particle oligonucleotides.

[0058] In some embodiments, the control barcode sequence is not homologous to genomic sequences of a species. The control barcode sequence can be homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species can be T7 phage, a PhiX phage, or a combination thereof.

[0059] In some embodiments, at least one of the plurality of control particle oligonucleotides is associated with the control particle through a linker. The at least one of the plurality of control particle oligonucleotides can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly attached to the at least one of the plurality of control particle oligonucleotides. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0060] In some embodiments, the diameter of the control particle is about 1-1000 micrometers, about 10-100 micrometers, 7.5 micrometer, or a combination thereof.

[0061] In some embodiments, the plurality of control particle oligonucleotides is immobilized on the control particle. The plurality of control particle oligonucleotides can be partially immobilized on the control particle. The plurality of control particle oligonucleotides can be enclosed in the control particle. The plurality of control particle oligonucleotides can be partially enclosed in the control particle. The control particle can be disruptable. The control particle can be a bead. The bead can be, or comprise, a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The control particle can comprise a material of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, or any combination thereof. The control particle can comprise a disruptable hydrogel particle.

[0062] In some embodiments, the control particle is associated with a detectable moiety. The control particle oligonucleotide can be associated with a detectable moiety.

[0063] In some embodiments, the control particle is associated with a plurality of first protein binding reagents, and at least one of the plurality of first protein binding reagents is associated with one of the plurality of control particle oligonucleotides. The first protein binding reagent can comprise a first antibody. The control particle oligonucleotide can be conjugated to the first protein binding reagent through a linker. The first control particle oligonucleotide can comprise the linker. The linker can comprise a chemical group. The

chemical group can be reversibly attached to the first protein binding reagent. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0064] In some embodiments, the first protein binding reagent is associated with two or more of the plurality of control particle oligonucleotides with an identical control barcode sequence. The first protein binding reagent can be associated with two or more of the plurality of control particle oligonucleotides with different control barcode sequences. In some embodiments, at least one of the plurality of first protein binding reagents is not associated with any of the plurality of control particle oligonucleotides. The first protein binding reagent associated with the control particle oligonucleotide and the first protein binding reagent not associated with any control particle oligonucleotide can be identical protein binding reagents.

[0065] In some embodiments, the control particle is associated with a plurality of second protein binding reagents. At least one of the plurality of second protein binding reagents can be associated with one of the plurality of control particle oligonucleotides. The control particle oligonucleotide associated with the first protein binding reagent and the control particle oligonucleotide associated with the second protein binding reagent can comprise different control barcode sequences. The first protein binding reagent and the second protein binding reagent can be identical protein binding reagents.

[0066] In some embodiments, the first protein binding reagent can be associated with a partner binding reagent, and wherein the first protein binding reagent is associated with the control particle using the partner binding reagent. The partner binding reagent can comprise a partner antibody. The partner antibody can comprise an anti-cat antibody, an anti-chicken antibody, an anti-cow antibody, an anti-dog antibody, an anti-donkey antibody, an anti-goat antibody, an anti-guinea pig antibody, an anti-hamster antibody, an anti-horse antibody, an anti-human antibody, an anti-llama antibody, an anti-monkey antibody, an anti-mouse antibody, an anti-pig antibody, an anti-rabbit antibody, an anti-rat antibody, an anti-sheep antibody, or a combination thereof. The partner antibody can comprise an immunoglobulin G (IgG), a F(ab') fragment, a F(ab')<sub>2</sub> fragment, a combination thereof, or a fragment thereof.

[0067] In some embodiments, the first protein binding reagent can be associated

with a detectable moiety. The second protein binding reagent can be associated with a detectable moiety.

[0068] Disclosed herein include methods for determining the numbers of targets. In some embodiments, the method comprises: stochastically barcoding a plurality of targets of a cell of a plurality of cells and a plurality of control particle oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets and a plurality of stochastically barcoded control particle oligonucleotides, wherein each of the plurality of stochastic barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two stochastic barcodes of the plurality of stochastic barcodes comprise different sequences, and wherein at least two stochastic barcodes of the plurality of stochastic barcodes comprise an identical cell label sequence, wherein a control particle composition comprises a control particle associated with the plurality of control particle oligonucleotides, wherein each of the plurality of control particle oligonucleotides comprises a control barcode sequence and a pseudo-target region comprising a sequence substantially complementary to the target-binding region of at least one of the plurality of stochastic barcodes. The method can comprise: obtaining sequencing data of the plurality of stochastically barcoded targets and the plurality of stochastically barcoded control particle oligonucleotides; counting the number of molecular label sequences with distinct sequences associated with the plurality of control particle oligonucleotides with the control barcode sequence in the sequencing data. The method can comprise: for at least one target of the plurality of targets: counting the number of molecular label sequences with distinct sequences associated with the target in the sequencing data; and estimating the number of the target, wherein the number of the target estimated correlates with the number of molecular label sequences with distinct sequences associated with the target counted and the number of molecular label sequences with distinct sequences associated with the control barcode sequence.

[0069] In some embodiments, the pseudo-target region comprises a poly(dA) region. The pseudo-target region can comprise a subsequence of the target. In some embodiments, the control barcode sequence can be at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The control particle

oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or any combination thereof. The control barcode sequences of at least 5, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can be identical. At least 3, 5, 10, 100, or more of the plurality of control particle oligonucleotides can comprise different control barcode sequences.

[0070] In some embodiments, the plurality of control particle oligonucleotides comprises a plurality of first control particle oligonucleotides each comprising a first control barcode sequence, and a plurality of second control particle oligonucleotides each comprising a second control barcode sequence. The first control barcode sequence and the second control barcode sequence can have different sequences. The number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be about the same. The number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be different. The number of the plurality of first control particle oligonucleotides can be at least 2 times, 10 times, 100 times, or more greater than the number of the plurality of second control particle oligonucleotides.

[0071] In some embodiments, counting the number of molecular label sequences with distinct sequences associated with the plurality of control particle oligonucleotides with the control barcode sequence in the sequencing data comprises: counting the number of molecular label sequences with distinct sequences associated with the first control barcode sequence in the sequencing data; and counting the number of molecular label sequences with distinct sequences associated with the second control barcode sequence in the sequencing data. The number of the target estimated can correlate with the number of molecular label sequences with distinct sequences associated with the target counted, the number of molecular label sequences with distinct sequences associated with the first control barcode sequence, and the number of molecular label sequences with distinct sequences associated with the second control barcode sequence. The number of the target estimated can correlate with the number of molecular label sequences with distinct sequences associated with the target counted, the number of molecular label sequences with distinct sequences associated with the control barcode sequence, and the number of the plurality of control particle

oligonucleotides comprising the control barcode sequence. The number of the target estimated can correlate with the number of molecular label sequences with distinct sequences associated with the target counted, and a ratio of the number of the plurality of control particle oligonucleotides comprising the control barcode sequence and the number of molecular label sequences with distinct sequences associated with the control barcode sequence.

[0072] In some embodiments, the control particle oligonucleotide is not homologous to genomic sequences of the cell. The control particle oligonucleotide can be not homologous to genomic sequences of the species. The control particle oligonucleotide can be homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species can be T7 phage, a PhiX phage, or a combination thereof.

[0073] In some embodiments, the control particle oligonucleotide can be conjugated to the control particle through a linker. At least one of the plurality of control particle oligonucleotides can be associated with the control particle through a linker. The at least one of the plurality of control particle oligonucleotides can comprise the linker. The chemical group can be reversibly attached to the at least one of the plurality of control particle oligonucleotides. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0074] In some embodiments, the diameter of the control particle is about 1-1000 micrometers, about 10-100 micrometers, about 7.5 micrometer, or a combination thereof. The plurality of control particle oligonucleotides is immobilized on the control particle. The plurality of control particle oligonucleotides can be partially immobilized on the control particle. The plurality of control particle oligonucleotides can be enclosed in the control particle. The plurality of control particle oligonucleotides can be partially enclosed in the control particle.

[0075] In some embodiments, the method comprises releasing the at least one of the plurality of control particle oligonucleotides from the control particle prior to stochastically barcoding the plurality of targets and the control particle and the plurality of control particle oligonucleotides.

[0076] In some embodiments, the control particle is disruptable. The control

particle can be a control particle bead. The control particle bead can comprise a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The control particle can comprise a material of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, or any combination thereof. The control particle can comprise a disruptable hydrogel particle.

[0077] In some embodiments, the control particle is associated with a detectable moiety. The control particle oligonucleotide can be associated with a detectable moiety.

[0078] In some embodiments, the control particle can be associated with a plurality of first protein binding reagents, and at least one of the plurality of first protein binding reagents can be associated with one of the plurality of control particle oligonucleotides. The first protein binding reagent can comprise a first antibody. The control particle oligonucleotide can be conjugated to the first protein binding reagent through a linker. The first control particle oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly attached to the first protein binding reagent. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0079] In some embodiments, the first protein binding reagent can be associated with two or more of the plurality of control particle oligonucleotides with an identical control barcode sequence. The first protein binding reagent can be associated with two or more of the plurality of control particle oligonucleotides with different control barcode sequences. At least one of the plurality of first protein binding reagents can be not associated with any of the plurality of control particle oligonucleotides. The first protein binding reagent associated with the control particle oligonucleotide and the first protein binding reagent not associated with any control particle oligonucleotide can be identical protein binding reagents. The control particle can associated with a plurality of second protein binding reagents. At least one of the plurality of second protein binding reagents can be associated with one of the plurality of control particle oligonucleotides. The control particle oligonucleotide associated

with the first protein binding reagent and the control particle oligonucleotide associated with the second protein binding reagent can comprise different control barcode sequences. The first protein binding reagent and the second protein binding reagent can be identical protein binding reagents.

[0080] In some embodiments, the first protein binding reagent is associated with a partner binding reagent, and wherein the first protein binding reagent is associated with the control particle using the partner binding reagent. The partner binding reagent can comprise a partner antibody. The partner antibody can comprise an anti-cat antibody, an anti-chicken antibody, an anti-cow antibody, an anti-dog antibody, an anti-donkey antibody, an anti-goat antibody, an anti-guinea pig antibody, an anti-hamster antibody, an anti-horse antibody, an anti-human antibody, an anti-llama antibody, an anti-monkey antibody, an anti-mouse antibody, an anti-pig antibody, an anti-rabbit antibody, an anti-rat antibody, an anti-sheep antibody, or a combination thereof. The partner antibody can comprise an immunoglobulin G (IgG), a F(ab') fragment, a F(ab')<sub>2</sub> fragment, a combination thereof, or a fragment thereof.

[0081] In some embodiments, the first protein binding reagent can be associated with a detectable moiety. The second protein binding reagent can be associated with a detectable moiety.

[0082] In some embodiments, the stochastic barcode comprises a binding site for a universal primer. The target-binding region can comprise a poly(dT) region.

[0083] In some embodiments, the plurality of stochastic barcodes is associated with a barcoding particle. At least one stochastic barcode of the plurality of stochastic barcodes can be immobilized on the barcoding particle. At least one stochastic barcode of the plurality of stochastic barcodes can be partially immobilized on the barcoding particle. At least one stochastic barcode of the plurality of stochastic barcodes can be enclosed in the barcoding particle. At least one stochastic barcode of the plurality of stochastic barcodes can be partially enclosed in the barcoding particle.

[0084] In some embodiments, the barcoding particle is disruptable. The barcoding particle can be a barcoding bead. The barcoding bead can comprise a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin



microbead, an anti-fluorochrome microbead, or any combination thereof. The barcoding particle can comprise a material of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, or any combination thereof. The barcoding particle can comprise a disruptable hydrogel particle.

[0085] In some embodiments, the stochastic barcodes of the barcoding particle comprise molecular label sequences selected from at least 1000, 10000, or more different molecular label sequences. In some embodiments, the molecular label sequences of the stochastic barcodes comprise random sequences. In some embodiments, The barcoding particle comprises at least 10000 stochastic barcodes.

[0086] In some embodiments, stochastically barcoding the plurality of targets and the plurality of control particle oligonucleotides using the plurality of stochastic barcodes comprises: contacting the plurality of stochastic barcodes with targets of the plurality of targets and control particle oligonucleotides of the plurality of control particle oligonucleotides to generate stochastic barcodes hybridized to the targets and the control particle oligonucleotides; and extending the stochastic barcodes hybridized to the targets and the control particle oligonucleotides to generate the plurality of stochastically barcoded targets and the plurality of stochastically barcoded control particle oligonucleotides. Extending the stochastic barcodes can comprise extending the stochastic barcodes using a DNA polymerase, a reverse transcriptase, or a combination thereof.

[0087] In some embodiments, the method comprises amplifying the plurality of stochastically barcoded targets and the plurality of stochastically barcoded control particle oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of stochastically barcoded targets and the plurality of stochastically barcoded control particle oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the molecular label sequence and at least a portion of the control particle oligonucleotide or at least a portion of the molecular label sequence and at least a portion of the control particle oligonucleotide. Obtaining the sequencing data can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing the at least a portion of the molecular label sequence and the at least a portion of the control particle oligonucleotide, or the at least a portion of the molecular label sequence

and the at least a portion of the control particle oligonucleotide.

[0088] Disclosed herein also include kits for sequencing control. In some embodiments, the kit comprises: a control particle composition comprising a plurality of control particle oligonucleotides associated with a control particle, wherein each of the plurality of control particle oligonucleotides comprises a control barcode sequence and a poly(dA) region.

[0089] In some embodiments, at least two of the plurality of control particle oligonucleotides comprise different control barcode sequences. In some embodiments, the control barcode sequence can be at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The control particle oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or any combination thereof. The control barcode sequences of at least 5, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can be identical. At least 3, 5, 10, 100, or more of the plurality of control particle oligonucleotides can comprise different control barcode sequences.

[0090] In some embodiments, the plurality of control particle oligonucleotides comprises a plurality of first control particle oligonucleotides each comprising a first control barcode sequence, and a plurality of second control particle oligonucleotides each comprising a second control barcode sequence. The first control barcode sequence and the second control barcode sequence can have different sequences. The number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be about the same. The number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be different. The number of the plurality of first control particle oligonucleotides can be at least 2 times, 10 times, 100 times, or more greater than the number of the plurality of second control particle oligonucleotides.

[0091] In some embodiments, the control particle oligonucleotide is not homologous to genomic sequences of the cell. The control particle oligonucleotide can be not homologous to genomic sequences of the species. The control particle oligonucleotide

can be homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species can be T7 phage, a PhiX phage, or a combination thereof.

[0092] In some embodiments, the control particle oligonucleotide can be conjugated to the control particle through a linker. At least one of the plurality of control particle oligonucleotides can be associated with the control particle through a linker. The at least one of the plurality of control particle oligonucleotides can comprise the linker. The chemical group can be reversibly attached to the at least one of the plurality of control particle oligonucleotides. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0093] In some embodiments, the diameter of the control particle is about 1-1000 micrometers, about 10-100 micrometers, about 7.5 micrometer, or a combination thereof. The plurality of control particle oligonucleotides is immobilized on the control particle. The plurality of control particle oligonucleotides can be partially immobilized on the control particle. The plurality of control particle oligonucleotides can be enclosed in the control particle. The plurality of control particle oligonucleotides can be partially enclosed in the control particle.

[0094] In some embodiments, the kit comprises a plurality of barcodes. A barcode of the plurality of barcodes can comprise a target-binding region and a molecular label sequence, and molecular label sequences of at least two barcodes of the plurality of barcodes can comprise different molecule label sequences. The barcode can comprise a cell label sequence, a binding site for a universal primer, or any combination thereof. The target-binding region comprises a poly(dT) region.

[0095] In some embodiments, the plurality of barcodes can be associated with a barcoding particle. At least one barcode of the plurality of barcodes can be immobilized on the barcoding particle. At least one barcode of the plurality of barcodes is partially immobilized on the barcoding particle. At least one barcode of the plurality of barcodes can be enclosed in the barcoding particle. At least one barcode of the plurality of barcodes can be partially enclosed in the barcoding particle. The barcoding particle can be disruptable. The barcoding particle can be a second bead. The bead can be, or comprise, a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A

conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The barcoding particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, and any combination thereof. The barcoding particle can comprise a disruptable hydrogel particle.

[0096] In some embodiments, the barcodes of the barcoding particle comprise molecular label sequences selected from at least 1000, 10000, or more different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences. The barcoding particle can comprise at least 10000 barcodes. The kit can comprise a DNA polymerase. The kit can comprise reagents for polymerase chain reaction (PCR).

[0097] Methods disclosed herein for cell identification can comprise: contacting a first plurality of cells and a second plurality of cells with two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprises one or more protein targets, wherein each of the two sample indexing compositions comprises a protein binding reagent associated with a sample indexing oligonucleotide, wherein the protein binding reagent is capable of specifically binding to at least one of the one or more protein targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying a cell label sequence associated with two or more sample indexing sequences in the sequencing data obtained; and removing sequencing data

associated with the cell label sequence from the sequencing data obtained. In some embodiments, the sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a universal primer, or a combination thereof. As described herein, the first plurality of cells can be obtained or derived from a different tissue or organ than the second plurality of cells, and the first plurality of cells and the second plurality of cells can be from the same or different subjects (e.g., a mammal). For example, the first plurality of cells can be obtained or derived from a different human subject than the second plurality of cells. In some embodiments, the first plurality of cells and the second plurality of cells are obtained or derived from different tissues of the same human subject.

**[0098]** In some embodiments, contacting the first plurality of cells and the second plurality of cells with the two sample indexing compositions respectively comprises: contacting the first plurality of cells with a first sample indexing compositions of the two sample indexing compositions; and contacting the first plurality of cells with a second sample indexing compositions of the two sample indexing compositions.

**[0099]** As described herein, the sample indexing sequence can be, for example, at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The sample indexing oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or a combination thereof. In some embodiments, sample indexing sequences of at least 10, 100, 1000, or more sample indexing compositions of the plurality of sample indexing compositions comprise different sequences.

**[0100]** In some embodiments, the protein binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the protein binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly attached to the protein binding reagent. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage or any combination thereof.

**[0101]** In some embodiments, at least one sample of the plurality of samples

comprises a single cell. The at least one of the one or more protein targets can be on a cell surface.

[0102] In some embodiments, the method comprises: removing unbound sample indexing compositions of the two sample indexing compositions. Removing the unbound sample indexing compositions can comprise washing cells of the first plurality of cells and the second plurality of cells with a washing buffer. Removing the unbound sample indexing compositions can comprise selecting cells bound to at least one protein binding reagent of the two sample indexing compositions using flow cytometry. In some embodiments, the method comprises: lysing the one or more cells from each of the plurality of samples.

[0103] In some embodiments, the sample indexing oligonucleotide is configured to be detachable or non-detachable from the protein binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the protein binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the protein binding reagent by UV photocleaving, chemical treatment, heating, enzyme treatment, or any combination thereof.

[0104] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of any of the one or more cells. The control barcode sequence may be not homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species is T7 phage, a PhiX phage, or a combination thereof.

[0105] In some embodiments, a sample of the plurality of samples comprises a plurality of cells, a tissue, a tumor sample, or any combination thereof. The plurality of sample can comprise a mammalian cell, a bacterial cell, a viral cell, a yeast cell, a fungal cell, or any combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. The barcode can comprise a target-binding region which comprises the capture sequence. The target-binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(dA) region.

[0106] In some embodiments, the protein target is, or comprises, an extracellular protein, an intracellular protein, or any combination thereof. The protein target can be, or

comprise, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The protein target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The protein target can be selected from a group comprising 10-100 different protein targets.

[0107] In some embodiments, the protein binding reagent is associated with two or more sample indexing oligonucleotides with an identical sequence. The protein binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second protein binding reagent not conjugated with the sample indexing oligonucleotide. The protein binding reagent and the second protein binding reagent can be identical.

[0108] In some embodiments, a barcode of the plurality of barcodes comprises a target-binding region and a molecular label sequence, and molecular label sequences of at least two barcodes of the plurality of barcodes comprise different molecule label sequences. The barcode can comprise a cell label sequence, a binding site for a universal primer, or any combination thereof. The target-binding region can comprise a poly(dT) region.

[0109] In some embodiments, the plurality of barcodes can be associated with a particle. At least one barcode of the plurality of barcodes can be immobilized on the particle. At least one barcode of the plurality of barcodes can be partially immobilized on the particle. At least one barcode of the plurality of barcodes can be enclosed in the particle. At least one barcode of the plurality of barcodes can be partially enclosed in the particle. The particle can be disruptable. The particle can be a bead. The bead can be, or comprise, a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise a disruptable hydrogel particle.

[0110] In some embodiments, the protein binding reagent is associated with a detectable moiety. In some embodiments, the particle is associated with a detectable moiety. The sample indexing oligonucleotide is associated with an optical moiety.

[0111] In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000, 10000, or more different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences. The particle can comprise at least 10000 barcodes.

[0112] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0113] In some embodiments, the method comprises: amplifying the plurality of barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data comprises sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, identifying the sample origin of the at least one cell comprises identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide.

[0114] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides comprises stochastically barcoding the sample indexing oligonucleotides



using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

[0115] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label sequence, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target-binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets.

[0116] In some embodiments, the method comprises: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0117] Also disclosed herein include methods and compositions that can be used for sequencing control. In some embodiments, the method for sequencing control comprises: contacting one or more cells of a plurality of cells with a control composition of a plurality of control compositions, wherein a cell of the plurality of cells comprises a plurality of targets and a plurality of protein targets, wherein each of the plurality of control compositions comprises a protein binding reagent associated with a control oligonucleotide, wherein the protein binding reagent is capable of specifically binding to at least one of the plurality of protein targets, and wherein the control oligonucleotide comprises a control barcode sequence and a pseudo-target region comprising a sequence substantially complementary to the target-binding region of at least one of the plurality of barcodes; barcoding the control oligonucleotides using a plurality of barcodes to create a plurality of barcoded control oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a

molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded control oligonucleotides; determining at least one characteristic of the one or more cells using at least one characteristic of the plurality of barcoded control oligonucleotides in the sequencing data. In some embodiments, the pseudo-target region comprises a poly(dA) region.

**[0118]** In some embodiments, the control barcode sequence is at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The control particle oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or a combination thereof. The control barcode sequences of at least 2, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can be identical. At least 2, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can comprise different control barcode sequences.

**[0119]** In some embodiments, determining the at least one characteristic of the one or more cells comprises: determining the number of cell label sequences with distinct sequences associated with the plurality of barcoded control oligonucleotides in the sequencing data; and determining the number of the one or more cells using the number of cell label sequences with distinct sequences associated with the plurality of barcoded control oligonucleotides. The method can comprise: determining single cell capture efficiency based the number of the one or more cells determined. The method can comprise: comprising determining single cell capture efficiency based on the ratio of the number of the one or more cells determined and the number of the plurality of cells.

**[0120]** In some embodiments, determining the at least one characteristic of the one or more cells using the characteristics of the plurality of barcoded control oligonucleotides in the sequencing data comprises: for each cell label in the sequencing data, determining the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence; and determining the number of the one or

more cells using the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence. Determining the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence can comprise: for each cell label in the sequencing data, determining the number of molecular label sequences with the highest number of distinct sequences associated with the cell label and the control barcode sequence. Determining the number of the one or more cells using the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence can comprise: generating a plot of the number of molecular label sequences with the highest number of distinct sequences with the number of cell labels in the sequencing data associated with the number of molecular label sequences with the highest number of distinct sequences; and determining a cutoff in the plot as the number of the one or more cells.

**[0121]** In some embodiments, the control oligonucleotide is not homologous to genomic sequences of any of the plurality of cells. The control oligonucleotide can be homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species can be T7 phage, a PhiX phage, or a combination thereof.

**[0122]** In some embodiments, the method comprises releasing the control oligonucleotide from the protein binding reagent prior to barcoding the control oligonucleotides. In some embodiments, the method comprises removing unbound control compositions of the plurality of control compositions. Removing the unbound control compositions can comprise washing the one or more cells of the plurality of cells with a washing buffer. Removing the unbound sample indexing compositions can comprise selecting cells bound to at least one protein binding reagent of the control composition using flow cytometry.

**[0123]** In some embodiments, at least one of the plurality of protein targets is on a cell surface. At least one of the plurality of protein targets can comprise a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The protein binding reagent can comprise an antibody. The control oligonucleotide can be conjugated to the protein binding reagent through a linker. The control oligonucleotide can comprise the

linker. The linker can comprise a chemical group. The chemical group can be reversibly attached to the first protein binding reagent. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0124] In some embodiments, the protein binding reagent is associated with two or more control oligonucleotides with an identical control barcode sequence. The protein binding reagent can be associated with two or more control oligonucleotides with different identical control barcode sequences. In some embodiments, a second protein binding reagent of the plurality of control compositions is not associated with the control oligonucleotide. The protein binding reagent and the second protein binding reagent can be identical.

[0125] In some embodiments, the barcode comprises a binding site for a universal primer. The target-binding region can comprise a poly(dT) region. In some embodiments, the plurality of barcodes is associated with a barcoding particle. At least one barcode of the plurality of barcodes can be immobilized on the barcoding particle. At least one barcode of the plurality of barcodes can be partially immobilized on the barcoding particle. At least one barcode of the plurality of barcodes is enclosed in the barcoding particle. At least one barcode of the plurality of barcodes is partially enclosed in the barcoding particle. The barcoding particle can be disruptable. The barcoding particle can be a barcoding bead. The barcoding bead can comprise a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The barcoding particle can comprise a material of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, or any combination thereof. The barcoding particle can comprise a disruptable hydrogel particle.

[0126] In some embodiments, the barcoding particle is associated with an optical moiety. The control oligonucleotide can be associated with an optical moiety.

[0127] In some embodiments, the barcodes of the barcoding particle comprise molecular label sequences selected from at least 1000, 10000, or more different molecular

label sequences. In some embodiments, the molecular label sequences of the barcodes comprise random sequences. In some embodiments, the barcoding particle comprises at least 10000 barcodes.

[0128] In some embodiments, barcoding the control oligonucleotides comprises: barcoding the control oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded control oligonucleotides. In some embodiments, barcoding the plurality of control oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with control oligonucleotides of the plurality of control compositions to generate barcodes hybridized to the control oligonucleotides; and extending the stochastic barcodes hybridized to the control oligonucleotides to generate the plurality of barcoded control oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase, a reverse transcriptase, or a combination thereof. In some embodiments, the method comprises amplifying the plurality of barcoded control oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded control oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the molecular label sequence and at least a portion of the control oligonucleotide. In some embodiments, obtaining the sequencing data comprises obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing the at least a portion of the molecular label sequence and the at least a portion of the control oligonucleotide.

[0129] Disclosed herein include methods for sequencing control. In some embodiments, the method comprises: contacting one or more cells of a plurality of cells with a control composition of a plurality of control compositions, wherein a cell of the plurality of cells comprises a plurality of targets and a plurality of binding targets, wherein each of the plurality of control compositions comprises a cellular component binding reagent associated with a control oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the plurality of binding targets, and wherein the control oligonucleotide comprises a control barcode sequence and a pseudo-target region comprising a sequence substantially complementary to the target-binding region of at least one of the plurality of barcodes; barcoding the control oligonucleotides using a plurality of barcodes to create a plurality of barcoded control oligonucleotides, wherein each of the plurality of

barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded control oligonucleotides; determining at least one characteristic of the one or more cells using at least one characteristic of the plurality of barcoded control oligonucleotides in the sequencing data. In some embodiments, the pseudo-target region comprises a poly(dA) region.

[0130] In some embodiments, the control barcode sequence is at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The control particle oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or a combination thereof. The control barcode sequences of at least 2, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can be identical. At least 2, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can comprise different control barcode sequences.

[0131] In some embodiments, determining the at least one characteristic of the one or more cells comprises: determining the number of cell label sequences with distinct sequences associated with the plurality of barcoded control oligonucleotides in the sequencing data; and determining the number of the one or more cells using the number of cell label sequences with distinct sequences associated with the plurality of barcoded control oligonucleotides. In some embodiments, the method comprises: determining single cell capture efficiency based the number of the one or more cells determined. In some embodiments, the method comprises: determining single cell capture efficiency based on the ratio of the number of the one or more cells determined and the number of the plurality of cells.

[0132] In some embodiments, determining the at least one characteristic of the one or more cells can comprise: for each cell label in the sequencing data, determining the number of molecular label sequences with distinct sequences associated with the cell label

and the control barcode sequence; and determining the number of the one or more cells using the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence. Determining the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence comprises: for each cell label in the sequencing data, determining the number of molecular label sequences with the highest number of distinct sequences associated with the cell label and the control barcode sequence. Determining the number of the one or more cells using the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence can comprise: generating a plot of the number of molecular label sequences with the highest number of distinct sequences with the number of cell labels in the sequencing data associated with the number of molecular label sequences with the highest number of distinct sequences; and determining a cutoff in the plot as the number of the one or more cells.

[0133] In some embodiments, the control oligonucleotide is not homologous to genomic sequences of any of the plurality of cells. The control oligonucleotide can be homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species can be T7 phage, a PhiX phage, or a combination thereof.

[0134] In some embodiments, the method comprises: releasing the control oligonucleotide from the cellular component binding reagent prior to barcoding the control oligonucleotides. At least one of the plurality of binding targets can be expressed on a cell surface. At least one of the plurality of binding targets can comprise a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The cellular component binding reagent can comprise a cell surface binding reagent, an antibody, a tetramer, an aptamers, a protein scaffold, an invasion, or a combination thereof.

[0135] In some embodiments, binding target of the cellular component binding reagent is selected from a group comprising 10-100 different binding targets. Aa binding target of the cellular component binding reagent can comprise a carbohydrate, a lipid, a protein, an extracellular protein, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, an

intracellular protein, or any combination thereof. The control oligonucleotide can be conjugated to the cellular component binding reagent through a linker. The control oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly attached to the first cellular component binding reagent. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0136] In some embodiments, the cellular component binding reagent can be associated with two or more control oligonucleotides with an identical control barcode sequence. The cellular component binding reagent can be associated with two or more control oligonucleotides with different identical control barcode sequences. In some embodiments, a second cellular component binding reagent of the plurality of control compositions is not associated with the control oligonucleotide. The cellular component binding reagent and the second cellular component binding reagent can be identical.

[0137] In some embodiments, the barcode comprises a binding site for a universal primer. In some embodiments, the target-binding region comprises a poly(dT) region.

[0138] In some embodiments, the plurality of barcodes is associated with a barcoding particle. At least one barcode of the plurality of barcodes can be immobilized on the barcoding particle. At least one barcode of the plurality of barcodes can be partially immobilized on the barcoding particle. At least one barcode of the plurality of barcodes can be enclosed in the barcoding particle. At least one barcode of the plurality of barcodes can be partially enclosed in the barcoding particle. The barcoding particle can be disruptable. The barcoding particle can be a barcoding bead. In some embodiments, the barcoding bead comprises a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The barcoding particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, and a combination thereof. The barcoding particle can comprise a disruptable hydrogel particle. The barcoding particle can



be associated with an optical moiety.

[0139] In some embodiments, the control oligonucleotide can be associated with an optical moiety. In some embodiments, the barcodes of the barcoding particle comprise molecular label sequences selected from at least 1000, 10000, or more different molecular label sequences. In some embodiments, the molecular label sequences of the barcodes comprise random sequences. The barcoding particle can comprise at least 10000 barcodes.

[0140] In some embodiments, barcoding the control oligonucleotides comprises: barcoding the control oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded control oligonucleotides. Barcoding the plurality of control oligonucleotides using the plurality of barcodes can comprise: contacting the plurality of barcodes with control oligonucleotides of the plurality of control compositions to generate barcodes hybridized to the control oligonucleotides; and extending the stochastic barcodes hybridized to the control oligonucleotides to generate the plurality of barcoded control oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase, a reverse transcriptase, or a combination thereof. In some embodiment, the method comprises amplifying the plurality of barcoded control oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded control oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the molecular label sequence and at least a portion of the control oligonucleotide. Obtaining the sequencing data can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing the at least a portion of the molecular label sequence and the at least a portion of the control oligonucleotide.

[0141] The methods for sequencing control can, in some embodiments, comprise: contacting one or more cells of a plurality of cells with a control composition of a plurality of control compositions, wherein a cell of the plurality of cells comprises a plurality of targets and a plurality of protein targets, wherein each of the plurality of control compositions comprises a protein binding reagent associated with a control oligonucleotide, wherein the protein binding reagent is capable of specifically binding to at least one of the plurality of protein targets, and wherein the control oligonucleotide comprises a control barcode sequence and a pseudo-target region comprising a sequence substantially complementary to the target-binding region of at least one of the plurality of barcodes; and determining at least

one characteristic of the one or more cells using at least one characteristic of the plurality of control oligonucleotides. The pseudo-target region can comprise a poly(dA) region.

[0142] In some embodiments, the control barcode sequence is at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The control particle oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or a combination thereof. The control barcode sequences of at least 2, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can be identical. At least 2, 10, 100, 1000, or more of the plurality of control particle oligonucleotides can comprise different control barcode sequences.

[0143] In some embodiments, determining the at least one characteristic of the one or more cells comprises: determining the number of cell label sequences with distinct sequences associated with the plurality of barcoded control oligonucleotides in the sequencing data; and determining the number of the one or more cells using the number of cell label sequences with distinct sequences associated with the plurality of barcoded control oligonucleotides. The method can comprise: determining single cell capture efficiency based the number of the one or more cells determined. The method can comprise: comprising determining single cell capture efficiency based on the ratio of the number of the one or more cells determined and the number of the plurality of cells.

[0144] In some embodiments, determining the at least one characteristic of the one or more cells using the characteristics of the plurality of barcoded control oligonucleotides in the sequencing data comprises: for each cell label in the sequencing data, determining the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence; and determining the number of the one or more cells using the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence. Determining the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence can comprise: for each cell label in the sequencing data, determining the number of molecular label sequences with the highest number of distinct sequences associated with the

cell label and the control barcode sequence. Determining the number of the one or more cells using the number of molecular label sequences with distinct sequences associated with the cell label and the control barcode sequence can comprise: generating a plot of the number of molecular label sequences with the highest number of distinct sequences with the number of cell labels in the sequencing data associated with the number of molecular label sequences with the highest number of distinct sequences; and determining a cutoff in the plot as the number of the one or more cells.

[0145] In some embodiments, the control oligonucleotide is not homologous to genomic sequences of any of the plurality of cells. The control oligonucleotide can be homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species can be T7 phage, a PhiX phage, or a combination thereof.

[0146] In some embodiments, the method comprises releasing the control oligonucleotide from the protein binding reagent prior to barcoding the control oligonucleotides. In some embodiments, the method comprises removing unbound control compositions of the plurality of control compositions. Removing the unbound control compositions can comprise washing the one or more cells of the plurality of cells with a washing buffer. Removing the unbound sample indexing compositions can comprise selecting cells bound to at least one protein binding reagent of the control composition using flow cytometry.

[0147] In some embodiments, at least one of the plurality of protein targets is on a cell surface. At least one of the plurality of protein targets can comprise a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The protein binding reagent can comprise an antibody. The control oligonucleotide can be conjugated to the protein binding reagent through a linker. The control oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly attached to the first protein binding reagent. The chemical group can comprise a UV photocleavable group, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0148] In some embodiments, the protein binding reagent is associated with two

or more control oligonucleotides with an identical control barcode sequence. The protein binding reagent can be associated with two or more control oligonucleotides with different identical control barcode sequences. In some embodiments, a second protein binding reagent of the plurality of control compositions is not associated with the control oligonucleotide. The protein binding reagent and the second protein binding reagent can be identical.

[0149] In some embodiments, the barcode comprises a binding site for a universal primer. The target-binding region can comprise a poly(dT) region. In some embodiments, the plurality of barcodes is associated with a barcoding particle. At least one barcode of the plurality of barcodes can be immobilized on the barcoding particle. At least one barcode of the plurality of barcodes can be partially immobilized on the barcoding particle. At least one barcode of the plurality of barcodes is enclosed in the barcoding particle. At least one barcode of the plurality of barcodes is partially enclosed in the barcoding particle. The barcoding particle can be disruptable. The barcoding particle can be a barcoding bead. The barcoding bead can comprise a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The barcoding particle can comprise a material of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, or any combination thereof. The barcoding particle can comprise a disruptable hydrogel particle.

[0150] In some embodiments, the barcoding particle is associated with an optical moiety. The control oligonucleotide can be associated with an optical moiety.

[0151] In some embodiments, the method comprises: barcoding the control oligonucleotides using a plurality of barcodes to create a plurality of barcoded control oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the plurality of barcoded control

oligonucleotides;

[0152] In some embodiments, the barcodes of the barcoding particle comprise molecular label sequences selected from at least 1000, 10000, or more different molecular label sequences. In some embodiments, the molecular label sequences of the barcodes comprise random sequences. In some embodiments, the barcoding particle comprises at least 10000 barcodes.

[0153] In some embodiments, barcoding the control oligonucleotides comprises: barcoding the control oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded control oligonucleotides. In some embodiments, barcoding the plurality of control oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with control oligonucleotides of the plurality of control compositions to generate barcodes hybridized to the control oligonucleotides; and extending the stochastic barcodes hybridized to the control oligonucleotides to generate the plurality of barcoded control oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase, a reverse transcriptase, or a combination thereof. In some embodiments, the method comprises amplifying the plurality of barcoded control oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded control oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the molecular label sequence and at least a portion of the control oligonucleotide. In some embodiments, obtaining the sequencing data comprises obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing the at least a portion of the molecular label sequence and the at least a portion of the control oligonucleotide.

[0154] Methods for cell identification can, in some embodiments, comprise: contacting a first plurality of cells and a second plurality of cells with two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprise one or more antigen targets, wherein each of the two sample indexing compositions comprises an antigen binding reagent associated with a sample indexing oligonucleotide, wherein the antigen binding reagent is capable of specifically binding to at least one of the one or more antigen targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing

sequences of the two sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying a cell label sequence associated with two or more sample indexing sequences in the sequencing data obtained; and removing sequencing data associated with the cell label sequence from the sequencing data obtained and/or excluding the sequencing data associated with the cell label sequence from subsequent analysis. In some embodiments, the sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a universal primer, or a combination thereof.

[0155] Disclosed herein also includes methods for multiplet identification. A multiplet expression profile, or a multiplet, can be an expression profile comprising expression profiles of multiplet cells. When determining expression profiles of single cells,  $n$  cells may be identified as one cell and the expression profiles of the  $n$  cells may be identified as the expression profile for one cell (referred to as a multiplet or  $n$ -plet expression profile). For example, when determining expression profiles of two cells using barcoding (e.g., stochastic barcoding), the mRNA molecules of the two cells may be associated with barcodes having the same cell label. As another example, two cells may be associated with one particle (e.g., a bead). The particle can include barcodes with the same cell label. After lysing the cells, the mRNA molecules in the two cells can be associated with the barcodes of the particle, thus the same cell label. Doublet expression profiles can skew the interpretation of the expression profiles. Multiplets can be different in different implementations. In some embodiments, the plurality of multiplets can include a doublet, a triplet, a quartet, a quintet, a sextet, a septet, an octet, a nonet, or any combination thereof.

[0156] A multiplet can be any  $n$ -plet. In some embodiments,  $n$  is, or is about, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or a range between any two of these values. In some embodiments,  $n$  is at least, or is at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,

12, 13, 14, 15, 16, 17, 18, 19, or 20. A singlet can be an expression profile that is not a multiplet expression profile.

[0157] The performance of using sample indexing oligonucleotides to index samples and identify multiplets can be comparable to the performance of using synthetic multiplet expression profiles to identify multiplets (described in U.S. Application No. 15/926977, filed on March 20, 2018, entitled “SYNTHETIC MULTIPLETS FOR MULTIPLETS DETERMINATION,” the content of which is incorporated herein in its entirety). In some embodiments, multiplets can be identified using both sample indexing oligonucleotides and synthetic multiplet expression profiles.

[0158] In some embodiments, the methods of multiplet identification disclosed herein comprise: contacting a first plurality of cells and a second plurality of cells with two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprise one or more antigen targets, wherein each of the two sample indexing compositions comprises an antigen binding reagent associated with a sample indexing oligonucleotide, wherein the antigen binding reagent is capable of specifically binding to at least one of the one or more antigen targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying one or more multiplet cell label sequences that is each associated with two or more sample indexing sequences in the sequencing data obtained. In some embodiments, the method comprises: removing the sequencing data associated with the one or more multiplet cell label sequences from the sequencing data obtained and/or excluding the sequencing data associated with the one or more multiplet cell label sequences from subsequent analysis. In some embodiments, the sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a

universal primer, or a combination thereof.

[0159] In some embodiments, contacting the first plurality of cells and the second plurality of cells with the two sample indexing compositions respectively comprises: contacting the first plurality of cells with a first sample indexing compositions of the two sample indexing compositions; and contacting the first plurality of cells with a second sample indexing compositions of the two sample indexing compositions.

[0160] In some embodiments, the sample indexing sequence is at least 6 nucleotides in length, 25-60 nucleotides in length (e.g., 45 nucleotides in length), about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The sample indexing oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or a combination thereof. In some embodiments, sample indexing sequences of at least 10, 100, 1000, or more sample indexing compositions of the plurality of sample indexing compositions comprise different sequences.

[0161] In some embodiments, the antigen binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the antigen binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the antigen binding reagent. The chemical group can comprise a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, a disulfide linkage or any combination thereof.

[0162] In some embodiments, at least one of the first plurality of cells and the second plurality of cells comprises single cells. The at least one of the one or more antigen targets can be on a cell surface.

[0163] In some embodiments, the method comprises: removing unbound sample indexing compositions of the two sample indexing compositions. Removing the unbound sample indexing compositions can comprise washing cells of the first plurality of cells and the second plurality of cells with a washing buffer. Removing the unbound sample indexing compositions can comprise selecting cells bound to at least one antigen binding reagent of the two sample indexing compositions using flow cytometry. In some embodiments, the



method comprises: lysing one or more cells of the first plurality of cells and the second plurality of cells.

[0164] In some embodiments, the sample indexing oligonucleotide is configured to be detachable or non-detachable from the antigen binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the antigen binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the antigen binding reagent by UV photocleaving, chemical treatment (e.g., using reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0165] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of any of the one or more cells. The control barcode sequence may be not homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species is T7 phage, a PhiX phage, or a combination thereof.

[0166] In some embodiments, the first plurality of cells and the second plurality of cells comprise a tumor cells, a mammalian cell, a bacterial cell, a viral cell, a yeast cell, a fungal cell, or any combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. The barcode can comprise a target-binding region which comprises the capture sequence. The target-binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(dA) region.

[0167] In some embodiments, the antigen target is, or comprises, an extracellular protein, an intracellular protein, or any combination thereof. The antigen target can be, or comprise, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The antigen target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The antigen target can be selected from a group comprising 10-100 different antigen targets.

[0168] In some embodiments, the antigen binding reagent is associated with two or more sample indexing oligonucleotides with an identical sequence. The antigen binding

reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second antigen binding reagent not conjugated with the sample indexing oligonucleotide. The antigen binding reagent and the second antigen binding reagent can be identical.

[0169] In some embodiments, a barcode of the plurality of barcodes comprises a target-binding region and a molecular label sequence, and molecular label sequences of at least two barcodes of the plurality of barcodes comprise different molecule label sequences. The barcode can comprise a cell label sequence, a binding site for a universal primer, or any combination thereof. The target-binding region can comprise a poly(dT) region.

[0170] In some embodiments, the plurality of barcodes can be associated with a particle. At least one barcode of the plurality of barcodes can be immobilized on the particle. At least one barcode of the plurality of barcodes can be partially immobilized on the particle. At least one barcode of the plurality of barcodes can be enclosed in the particle. At least one barcode of the plurality of barcodes can be partially enclosed in the particle. The particle can be disruptable. The particle can be a bead. The bead can be, or comprise, a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise a disruptable hydrogel particle.

[0171] In some embodiments, the antigen binding reagent is associated with a detectable moiety. In some embodiments, the particle is associated with a detectable moiety. The sample indexing oligonucleotide is associated with an optical moiety. In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000, 10000, or more different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences. The particle can comprise at least 10000 barcodes.

[0172] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0173] In some embodiments, the method comprises: amplifying the plurality of barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data comprises sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, identifying the sample origin of the at least one cell comprises identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide.

[0174] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides comprises stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

[0175] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label sequence, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using

the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target-binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets.

[0176] In some embodiments, the method comprises: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0177] Methods for cell identification can, in some embodiments, comprise: contacting a first plurality of cells and a second plurality of cells with two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprise one or more cellular component targets, wherein each of the two sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; identifying one or more cell label sequences that is each associated with two or more sample indexing sequences in the sequencing data obtained; and removing the sequencing data associated with the one or more cell label sequences that is

each associated with two or more sample indexing sequences from the sequencing data obtained and/or excluding the sequencing data associated with the one or more cell label sequences that is each associated with two or more sample indexing sequences from subsequent analysis. In some embodiments, the sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a universal primer, or a combination thereof.

[0178] Disclosed herein includes methods for multiplet identification. In some embodiments, the method comprises: contacting a first plurality of cells and a second plurality of cells with two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprise one or more cellular component targets, wherein each of the two sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; identifying one or more multiplet cell label sequences that is each associated with two or more sample indexing sequences in the sequencing data obtained. In some embodiments, the method comprises: removing the sequencing data associated with the one or more multiplet cell label sequences from the sequencing data obtained and/or excluding the sequencing data associated with the one or more multiplet cell label sequences from subsequent analysis. In some embodiments, the sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a universal primer, or a combination thereof.

[0179] In some embodiments, contacting the first plurality of cells and the second

plurality of cells with the two sample indexing compositions respectively comprises: contacting the first plurality of cells with a first sample indexing compositions of the two sample indexing compositions; and contacting the first plurality of cells with a second sample indexing compositions of the two sample indexing compositions.

[0180] In some embodiments, the sample indexing sequence is at least 6 nucleotides in length, 25-60 nucleotides in length (e.g., 45 nucleotides in length), about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The sample indexing oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or a combination thereof. In some embodiments, sample indexing sequences of at least 10, 100, 1000, or more sample indexing compositions of the plurality of sample indexing compositions comprise different sequences.

[0181] In some embodiments, the cellular component binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the cellular component binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the cellular component binding reagent. The chemical group can comprise a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, a disulfide linkage or any combination thereof.

[0182] In some embodiments, at least one of the first plurality of cells and the second plurality of cells comprises a single cell. The at least one of the one or more cellular component targets can be on a cell surface.

[0183] In some embodiments, the method comprises: removing unbound sample indexing compositions of the two sample indexing compositions. Removing the unbound sample indexing compositions can comprise washing cells of the first plurality of cells and the second plurality of cells with a washing buffer. Removing the unbound sample indexing compositions can comprise selecting cells bound to at least one cellular component binding reagent of the two sample indexing compositions using flow cytometry. In some embodiments, the method comprises: lysing one or more cells of the first plurality of cells

and the second plurality of cells.

[0184] In some embodiments, the sample indexing oligonucleotide is configured to be detachable or non-detachable from the cellular component binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent by UV photocleaving, chemical treatment (e.g., using reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0185] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of any of the one or more cells. The control barcode sequence may be not homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species is T7 phage, a PhiX phage, or a combination thereof.

[0186] In some embodiments, the first plurality of cells and the second plurality of cells comprise a tumor cell, a mammalian cell, a bacterial cell, a viral cell, a yeast cell, a fungal cell, or any combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. The barcode can comprise a target-binding region which comprises the capture sequence. The target-binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(dA) region.

[0187] In some embodiments, the antigen target is, or comprises, an extracellular protein, an intracellular protein, or any combination thereof. The antigen target can be, or comprise, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The antigen target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The antigen target can be selected from a group comprising 10-100 different antigen targets.

[0188] In some embodiments, the cellular component binding reagent is associated with two or more sample indexing oligonucleotides with an identical sequence. The cellular component binding reagent can be associated with two or more sample indexing

oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second cellular component binding reagent not conjugated with the sample indexing oligonucleotide. The cellular component binding reagent and the second cellular component binding reagent can be identical.

[0189] In some embodiments, a barcode of the plurality of barcodes comprises a target-binding region and a molecular label sequence, and molecular label sequences of at least two barcodes of the plurality of barcodes comprise different molecule label sequences. The barcode can comprise a cell label sequence, a binding site for a universal primer, or any combination thereof. The target-binding region can comprise a poly(dT) region.

[0190] In some embodiments, the plurality of barcodes can be associated with a particle. At least one barcode of the plurality of barcodes can be immobilized on the particle. At least one barcode of the plurality of barcodes can be partially immobilized on the particle. At least one barcode of the plurality of barcodes can be enclosed in the particle. At least one barcode of the plurality of barcodes can be partially enclosed in the particle. The particle can be disruptable. The particle can be a bead. The bead can be, or comprise, a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise a disruptable hydrogel particle.

[0191] In some embodiments, the cellular component binding reagent is associated with a detectable moiety. In some embodiments, the particle is associated with a detectable moiety. The sample indexing oligonucleotide is associated with an optical moiety.

[0192] In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000, 10000, or more different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences. The particle can comprise at least 10000 barcodes.



[0193] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0194] In some embodiments, the method comprises: amplifying the plurality of barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data comprises sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, identifying the sample origin of the at least one cell comprises identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide.

[0195] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides comprises stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

[0196] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label sequence, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using

the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target-binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets.

**[0197]** In some embodiments, the method comprises: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

**[0198]** Disclosed herein includes methods for cell identification. In some embodiments, the method comprises: contacting one or more cells from each of a first plurality of cells and a second plurality of cells with a sample indexing composition of a plurality of two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprises one or more antigen targets, wherein each of the two sample indexing compositions comprises an antigen binding reagent associated with a sample indexing oligonucleotide, wherein the antigen binding reagent is capable of specifically binding to at least one of the one or more antigen targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences; and identifying one or more cells that is each associated with two or more sample indexing sequences. In some embodiments, the sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a universal primer, or a combination thereof.

**[0199]** Disclosed herein include methods for multiplet identification. In some embodiments, the method comprises: contacting one or more cells from each of a first plurality of cells and a second plurality of cells with a sample indexing composition of a plurality of two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprises one or more antigen

targets, wherein each of the two sample indexing compositions comprises an antigen binding reagent associated with a sample indexing oligonucleotide, wherein the antigen binding reagent is capable of specifically binding to at least one of the one or more antigen targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences; and identifying one or more cells that is each associated with two or more sample indexing sequences as multiplet cells.

[0200] In some embodiments, identifying the cells that is each associated with two or more sample indexing sequences comprises: barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a molecular label sequence, and a target-binding region, wherein the molecular label sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying one or more cell label sequences that is each associated with two or more sample indexing sequences in the sequencing data obtained. The method can comprise removing the sequencing data associated with the one or more cell label sequences that is each associated with two or more sample indexing sequences from the sequencing data obtained and/or excluding the sequencing data associated with the one or more cell label sequences that is each associated with the two or more sample indexing sequences from subsequent analysis.

[0201] In some embodiments, contacting the first plurality of cells and the second plurality of cells with the two sample indexing compositions respectively comprises: contacting the first plurality of cells with a first sample indexing compositions of the two sample indexing compositions; and contacting the first plurality of cells with a second sample indexing compositions of the two sample indexing compositions.

[0202] In some embodiments, the sample indexing sequence is at least 6 nucleotides in length, 25-60 nucleotides in length (e.g., 45 nucleotides in length), about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, or a combination thereof. The sample indexing oligonucleotide can be about 50 nucleotides

in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or a combination thereof. In some embodiments, sample indexing sequences of at least 10, 100, 1000, or more sample indexing compositions of the plurality of sample indexing compositions comprise different sequences.

[0203] In some embodiments, the antigen binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the antigen binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the antigen binding reagent. The chemical group can comprise a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, a disulfide linkage or any combination thereof.

[0204] In some embodiments, at least one of the first plurality of cells and the second plurality of cells comprises single cells. The at least one of the one or more antigen targets can be on a cell surface. In some embodiments, the method comprises: removing unbound sample indexing compositions of the two sample indexing compositions. Removing the unbound sample indexing compositions can comprise washing cells of the first plurality of cells and the second plurality of cells with a washing buffer. Removing the unbound sample indexing compositions can comprise selecting cells bound to at least one antigen binding reagent of the two sample indexing compositions using flow cytometry. In some embodiments, the method comprises: lysing one or more cells of the first plurality of cells and the second plurality of cells.

[0205] In some embodiments, the sample indexing oligonucleotide is configured to be detachable or non-detachable from the antigen binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the antigen binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the antigen binding reagent by UV photocleaving, chemical treatment (e.g., using reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0206] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of any of the one or more cells. The control barcode

sequence may be not homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species is T7 phage, a PhiX phage, or a combination thereof.

[0207] In some embodiments, the first plurality of cells and the second plurality of cells comprise a tumor cells, a mammalian cell, a bacterial cell, a viral cell, a yeast cell, a fungal cell, or any combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. The barcode can comprise a target-binding region which comprises the capture sequence. The target-binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(dA) region.

[0208] In some embodiments, the antigen target is, or comprises, an extracellular protein, an intracellular protein, or any combination thereof. The antigen target can be, or comprise, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The antigen target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The antigen target can be selected from a group comprising 10-100 different antigen targets.

[0209] In some embodiments, the antigen binding reagent is associated with two or more sample indexing oligonucleotides with an identical sequence. The antigen binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second antigen binding reagent not conjugated with the sample indexing oligonucleotide. The antigen binding reagent and the second antigen binding reagent can be identical.

[0210] In some embodiments, a barcode of the plurality of barcodes comprises a target-binding region and a molecular label sequence, and molecular label sequences of at least two barcodes of the plurality of barcodes comprise different molecule label sequences. The barcode can comprise a cell label sequence, a binding site for a universal primer, or any combination thereof. The target-binding region can comprise a poly(dT) region.

[0211] In some embodiments, the plurality of barcodes can be associated with a

particle. At least one barcode of the plurality of barcodes can be immobilized on the particle. At least one barcode of the plurality of barcodes can be partially immobilized on the particle. At least one barcode of the plurality of barcodes can be enclosed in the particle. At least one barcode of the plurality of barcodes can be partially enclosed in the particle. The particle can be disruptable. The particle can be a bead. The bead can be, or comprise, a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise a disruptable hydrogel particle.

[0212] In some embodiments, the antigen binding reagent is associated with a detectable moiety. In some embodiments, the particle is associated with a detectable moiety. The sample indexing oligonucleotide is associated with an optical moiety.

[0213] In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000, 10000, or more different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences. The particle can comprise at least 10000 barcodes.

[0214] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0215] In some embodiments, the method comprises: amplifying the plurality of barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying

the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data comprises sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, identifying the sample origin of the at least one cell comprises identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide.

[0216] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides comprises stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

[0217] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label sequence, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target-binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets.

[0218] In some embodiments, the method comprises: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create

a plurality of stochastically barcoded targets.

[0219] Disclosed herein include systems, methods, and kits for determining interactions between cellular component targets, for example interactions between proteins. In some embodiments, the method comprises: contacting a cell with a first pair of interaction determination compositions, wherein the cell comprises a first protein target and a second protein target, wherein each of the first pair of interaction determination compositions comprises a protein binding reagent associated with an interaction determination oligonucleotide, wherein the protein binding reagent of one of the first pair of interaction determination compositions is capable of specifically binding to the first protein target and the protein binding reagent of the other of the first pair of interaction determination compositions is capable of specifically binding to the second protein target, and wherein the interaction determination oligonucleotide comprises an interaction determination sequence and a bridge oligonucleotide hybridization region, and wherein the interaction determination sequences of the first pair of interaction determination compositions comprise different sequences; ligating the interaction determination oligonucleotides of the first pair of interaction determination compositions using a bridge oligonucleotide to generate a ligated interaction determination oligonucleotide, wherein the bridge oligonucleotide comprises two hybridization regions capable of specifically binding to the bridge oligonucleotide hybridization regions of the first pair of interaction determination compositions; barcoding the ligated interaction determination oligonucleotide using a plurality of barcodes to create a plurality of barcoded interaction determination oligonucleotides, wherein each of the plurality of barcodes comprises a barcode sequence and a capture sequence; obtaining sequencing data of the plurality of barcoded interaction determination oligonucleotides; and determining an interaction between the first and second protein targets based on the association of the interaction determination sequences of the first pair of interaction determination compositions in the obtained sequencing data.

[0220] Disclosed herein include systems, methods, and kits for determining interactions between cellular component targets. In some embodiments, the method comprises: contacting a cell with a first pair of interaction determination compositions, wherein the cell comprises a first cellular component target and a second cellular component target, wherein each of the first pair of interaction determination compositions comprises a



cellular component binding reagent associated with an interaction determination oligonucleotide, wherein the cellular component binding reagent of one of the first pair of interaction determination compositions is capable of specifically binding to the first cellular component target and the cellular component binding reagent of the other of the first pair of interaction determination compositions is capable of specifically binding to the second cellular component target, and wherein the interaction determination oligonucleotide comprises an interaction determination sequence and a bridge oligonucleotide hybridization region, and wherein the interaction determination sequences of the first pair of interaction determination compositions comprise different sequences; ligating the interaction determination oligonucleotides of the first pair of interaction determination compositions using a bridge oligonucleotide to generate a ligated interaction determination oligonucleotide, wherein the bridge oligonucleotide comprises two hybridization regions capable of specifically binding to the bridge oligonucleotide hybridization regions of the first pair of interaction determination compositions; barcoding the ligated interaction determination oligonucleotide using a plurality of barcodes to create a plurality of barcoded interaction determination oligonucleotides, wherein each of the plurality of barcodes comprises a barcode sequence and a capture sequence; obtaining sequencing data of the plurality of barcoded interaction determination oligonucleotides; and determining an interaction between the first and second cellular component targets based on the association of the interaction determination sequences of the first pair of interaction determination compositions in the obtained sequencing data. At least one of the two cellular component binding reagent can comprise a protein binding reagent. The protein binding reagent can be associated with one of the two interaction determination oligonucleotides. The one or more cellular component targets can comprise at least one protein target.

[0221] In some embodiments, contacting the cell with the first pair of interaction determination compositions comprises: contacting the cell with each of the first pair of interaction determination compositions sequentially or simultaneously. The first protein target can be the same as the second protein target, or the first protein target can be different from the second protein target.

[0222] The length of the interaction determination sequence can vary. For example, the interaction determination sequence can be 2 nucleotides to about 1000

nucleotides in length. In some embodiments, the interaction determination sequence can be at least 6 nucleotides in length, 25-60 nucleotides in length, about 45 nucleotides in length, about 50 nucleotides in length, about 100 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, about 200-300 nucleotides in length, about 200-500 nucleotides in length, about 500 nucleotides in length, or any combination thereof.

[0223] In some embodiments, the method comprises: contacting the cell with a second pair of interaction determination compositions, wherein the cell comprises a third protein target and a fourth protein target, wherein each of the second pair of interaction determination compositions comprises a protein binding reagent associated with an interaction determination oligonucleotide, wherein the protein binding reagent of one of the second pair of interaction determination compositions is capable of specifically binding to the third protein target and the protein binding reagent of the other of the second pair of interaction determination compositions is capable of specifically binding to the fourth protein target. At least one of the third and fourth protein targets can be different from one of the first and second protein targets. In some embodiments, at least one of the third and fourth protein targets and at least one of the first and second protein targets can be identical.

[0224] In some embodiments, the method comprises: contacting the cell with three or more pairs of interaction determination compositions. The interaction determination sequences of at least 10 interaction determination compositions of the plurality of pairs of interaction determination compositions can comprise different sequences. The interaction determination sequences of at least 100 interaction determination compositions of the plurality of pairs of interaction determination compositions can comprise different sequences. The interaction determination sequences of at least 1000 interaction determination compositions of the plurality of pairs of interaction determination compositions can comprise different sequences.

[0225] In some embodiments, the bridge oligonucleotide hybridization regions of the first pair of interaction determination compositions comprise different sequences. At least one of the bridge oligonucleotide hybridization regions can be complementary to at least one of the two hybridization regions of the bridge oligonucleotide.

[0226] In some embodiments, ligating the interaction determination

oligonucleotides of the first pair of interaction determination compositions using the bridge oligonucleotide comprises: hybridizing a first hybridization regions of the bridge oligonucleotide with a first bridge oligonucleotide hybridization region of the bridge oligonucleotide hybridization regions of the interaction determination oligonucleotides; hybridizing a second hybridization region of the bridge oligonucleotide with a second bridge oligonucleotide hybridization region of the bridge oligonucleotide hybridization regions of the interaction determination oligonucleotides; and ligating the interaction determination oligonucleotides hybridized to the bridge oligonucleotide to generate a ligated interaction determination oligonucleotide.

[0227] In some embodiments, the protein binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, an integrin, or a combination thereof. The interaction determination oligonucleotide can be conjugated to the protein binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the protein binding reagent. The chemical group can comprise a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, a disulfide linkage or any combination thereof.

[0228] The location of the protein targets in the cell can vary, for example, on the cell surface or inside the cell. In some embodiments, the at least one of the one or more protein targets is on a cell surface. In some embodiments, the at least one of the one or more protein targets is an intracellular protein. In some embodiments, the at least one of the one or more protein targets is a transmembrane protein. In some embodiments, the at least one of the one or more protein targets is an extracellular protein. In some embodiments, the method can comprise: fixating the cell prior to contacting the cell with the first pair of interaction determination compositions. In some embodiments, the method can comprise: removing unbound interaction determination compositions of the first pair of interaction determination compositions. Removing the unbound interaction determination compositions can comprise washing the cell with a washing buffer. Removing the unbound interaction determination compositions can comprise selecting the cell using flow cytometry. In some embodiments, the method can comprise: lysing the cell.

[0229] In some embodiments, the interaction determination oligonucleotide is configured to be detachable or non-detachable from the protein binding reagent. The method

can comprise: detaching the interaction determination oligonucleotide from the protein binding reagent. Detaching the interaction determination oligonucleotide can comprise detaching the interaction determination oligonucleotide from the protein binding reagent by UV photocleaving, chemical treatment, heating, enzyme treatment, or any combination thereof. The interaction determination oligonucleotide can be not homologous to genomic sequences of the cell. The interaction determination oligonucleotide can be homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species can be T7 phage, a PhiX phage, or a combination thereof. In some embodiments, the interaction determination oligonucleotide of the one of the first pair of interaction determination compositions comprises a sequence complementary to the capture sequence. The capture sequence can comprise a poly(dT) region. The sequence of the interaction determination oligonucleotide complementary to the capture sequence can comprise a poly(dA) region. In some embodiments, the interaction determination oligonucleotide comprises a second barcode sequence. The interaction determination oligonucleotide of the other of the first pair of interaction identification compositions can comprise a binding site for a universal primer. The interaction determination oligonucleotide can be associated with a detectable moiety.

[0230] In some embodiments, the protein binding reagent can be associated with two or more interaction determination oligonucleotides with different interaction determination sequences. In some embodiments, the one of the plurality of interaction determination compositions comprises a second protein binding reagent not associated with the interaction determination oligonucleotide. The protein binding reagent and the second protein binding reagent can be identical. The protein binding reagent can be associated with a detectable moiety.

[0231] In some embodiments, the cell is a tumor cell or non-tumor cell. For example, the cell can be a mammalian cell, a bacterial cell, a viral cell, a yeast cell, a fungal cell, or any combination thereof. In some embodiments, the method comprises: contacting two or more cells with the first pair of interaction determination compositions, and wherein each of the two or more cells comprises the first and the second protein targets. At least one of the two or more cells can comprise a single cell.

[0232] In some embodiments, the protein target is, or comprises, an extracellular

protein, an intracellular protein, or any combination thereof. The protein target can be, or comprise, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The protein target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The protein target can be selected from a group comprising 10-100 different protein targets. The protein binding reagent can be associated with two or more interaction determination oligonucleotides with an identical sequence.

[0233] In some embodiments, the barcode comprises a cell label sequence, a binding site for a universal primer, or any combination thereof. At least two barcodes of the plurality of barcodes can comprise an identical cell label sequence. In some embodiments, the plurality of barcodes is associated with a particle. At least one barcode of the plurality of barcodes can be immobilized on the particle. At least one barcode of the plurality of barcodes can be partially immobilized on the particle. At least one barcode of the plurality of barcodes can be enclosed in the particle. At least one barcode of the plurality of barcodes can be partially enclosed in the particle. The particle can be disruptable. The particle can comprise a bead. The particle can comprise a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise a disruptable hydrogel particle. The particle can be associated with a detectable moiety.

[0234] In some embodiments, the barcodes of the particle comprise barcode sequences selected from at least 1000 different barcode sequences. The barcodes of the particle can comprise barcode sequences selected from at least 10000 different barcode sequences. The barcodes sequences of the barcodes can comprise random sequences. The particle can comprise at least 10000 barcodes.

[0235] In some embodiments, barcoding the interaction determination

oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the interaction determination oligonucleotides to generate barcodes hybridized to the interaction determination oligonucleotides; and extending the barcodes hybridized to the interaction determination oligonucleotides to generate the plurality of barcoded interaction determination oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded interaction determination oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded interaction determination oligonucleotides. Extending the barcodes can comprise extending the barcodes using a Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase or a Taq DNA polymerase to generate the plurality of barcoded interaction determination oligonucleotides. Extending the barcodes can comprise displacing the bridge oligonucleotide from the ligated interaction determination oligonucleotide.

[0236] In some embodiments, the method comprises: amplifying the plurality of barcoded interaction determination oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded interaction determination oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the barcode sequence and at least a portion of the interaction determination oligonucleotide. Obtaining the sequencing data of the plurality of barcoded interaction determination oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing at least a portion of the barcode sequence and at least a portion of the interaction determination oligonucleotide.

[0237] In some embodiments, obtaining sequencing data of the plurality of barcoded interaction determination oligonucleotides comprises obtaining partial and/or complete sequences of the plurality of barcoded interaction determination oligonucleotides.

[0238] In some embodiments, the plurality of barcodes comprises a plurality of stochastic barcodes. The barcode sequence of each of the plurality of stochastic barcodes can comprise a molecular label sequence. The molecular label sequences of at least two stochastic barcodes of the plurality of stochastic barcodes can comprise different sequences. Barcoding the interaction determination oligonucleotides using the plurality of barcodes to create the plurality of barcoded interaction determination oligonucleotides can comprise

stochastically barcoding the interaction determination oligonucleotides using the plurality of stochastic barcodes to create a plurality of stochastically barcoded interaction determination oligonucleotides.

[0239] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target-binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets. The method can comprise: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using the plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0240] Disclosed herein include kits for identifying interactions between cellular components, for example protein-protein interactions. In some embodiments, the kit comprises: a first pair of interaction determination compositions, wherein each of the first pair of interaction determination compositions comprises a protein binding reagent associated with an interaction determination oligonucleotide, wherein the protein binding reagent of one of the first pair of interaction determination compositions is capable of specifically binding to a first protein target and a protein binding reagent of the other of the first pair of interaction determination compositions is capable of specifically binding to the second protein target, wherein the interaction determination oligonucleotide comprises an interaction determination sequence and a bridge oligonucleotide hybridization region, and wherein the interaction determination sequences of the first pair of interaction determination compositions comprise different sequences; and a plurality of bridge oligonucleotides each comprising two hybridization regions capable of specifically binding to the bridge oligonucleotide hybridization regions of the first pair of interaction determination compositions.

[0241] The length of the interaction determination sequence can vary. In some embodiments, the interaction determination sequence is at least 6 nucleotides in length, 25-60 nucleotides in length, about 45 nucleotides in length, about 50 nucleotides in length, about 100 nucleotides in length, about 128 nucleotides in length, at least 128 nucleotides in length, about 200-500 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 500 nucleotides in length, or any combination thereof.

[0242] In some embodiments, the kit further comprises: a second pair of interaction determination compositions, wherein each of the second pair of interaction determination compositions comprises a protein binding reagent associated with an interaction determination oligonucleotide, wherein the protein binding reagent of one of the second pair of interaction determination compositions is capable of specifically binding to a third protein target and the protein binding reagent of the other of the second pair of interaction determination compositions is capable of specifically binding to a fourth protein target. At least one of the third and fourth protein targets can be different from one of the first and second protein targets. At least one of the third and fourth protein targets and at least one of the first and second protein targets can be identical.

[0243] In some embodiments, the kit can comprise three or more pairs of interaction determination compositions. The interaction determination sequences of at least 10 interaction determination compositions of the three or more pairs of interaction determination compositions can comprise different sequences. The interaction determination sequences of at least 100 interaction determination compositions of the three or more pairs of interaction determination compositions can comprise different sequences. The interaction determination sequences of at least 1000 interaction determination compositions of the three or more pairs of interaction determination compositions can comprise different sequences.

[0244] In some embodiments, the bridge oligonucleotide hybridization regions of two interaction determination compositions of the plurality of interaction determination compositions comprise different sequences. At least one of the bridge oligonucleotide hybridization regions can be complementary to at least one of the two hybridization regions of the bridge oligonucleotide.

[0245] In some embodiments, the protein binding reagent comprises an antibody,



a tetramer, an aptamers, a protein scaffold, or a combination thereof. The interaction determination oligonucleotide can be conjugated to the protein binding reagent through a linker. The at least one interaction determination oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the protein binding reagent. The chemical group can comprise a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof. The interaction determination oligonucleotide can be not homologous to genomic sequences of any cell of interest. The cell of interest can be a tumor cell or non-tumor cell. The cell of interest can be a single cell, a mammalian cell, a bacterial cell, a viral cell, a yeast cell, a fungal cell, or any combination thereof.

[0246] In some embodiments, the kit further comprises: a plurality of barcodes, wherein each of the plurality of barcodes comprises a barcode sequence and a capture sequence. The interaction determination oligonucleotide of the one of the first pair of interaction determination compositions can comprise a sequence complementary to the capture sequence of at least one barcode of a plurality of barcodes. The capture sequence can comprise a poly(dT) region. The sequence of the interaction determination oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(dA) region. The interaction determination oligonucleotide of the other of the first pair of interaction identification compositions can comprise a cell label sequence, a binding site for a universal primer, or any combination thereof. The plurality of barcodes can comprise a plurality of stochastic barcodes, wherein the barcode sequence of each of the plurality of stochastic barcodes comprises a molecular label sequence, wherein the molecular label sequences of at least two stochastic barcodes of the plurality of stochastic barcodes comprise different sequences.

[0247] In some embodiments, the protein target is, or comprises, an extracellular protein, an intracellular protein, or any combination thereof. The protein target can be, or comprise, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. The protein target can be selected from a group comprising 10-100 different protein targets. The protein binding reagent can be associated with two or more interaction determination oligonucleotides with an identical sequence. The protein binding reagent can be associated

with two or more interaction determination oligonucleotides with different interaction determination sequences. In some embodiments, the protein binding reagent can be associated with a detectable moiety.

[0248] In some embodiments, the one of the first pair of interaction determination compositions comprises a second protein binding reagent not associated with the interaction determination oligonucleotide. The first protein binding reagent and the second protein binding reagent can be identical. The interaction determination oligonucleotide can be associated with a detectable moiety.

[0249] In some embodiments, the plurality of barcodes is associated with a particle. At least one barcode the plurality of barcodes can be immobilized on the particle. At least one barcode of the plurality of barcodes can be partially immobilized on the particle. At least one barcode of the plurality of barcodes can be enclosed in the particle. At least one barcode of the plurality of barcodes can be partially enclosed in the particle. The particle can be disruptable. The particle can comprise a bead. The particle can comprise a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise a disruptable hydrogel particle. The particle can be associated with a detectable moiety.

[0250] In some embodiments, the barcodes of the particle comprise barcode sequences selected from at least 1000 different barcode sequences. The barcodes of the particle can comprise barcode sequences selected from least 10000 different barcode sequences. The barcodes sequences of the barcodes can comprise random sequences. The particle can comprise at least 10000 barcodes.

[0251] In some embodiments, the kit comprises: a DNA polymerase, a reverse transcriptase, a Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase, a Taq DNA polymerase, or any combination thereof. The kit can comprise a fixation agent.

[0252] Disclosed herein include methods for sample identification. In some embodiments, the method comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein the one or more cells comprises one or more cellular component targets, wherein each of the plurality of sample indexing compositions comprises a cellular component binding reagent (e.g., an antibody) associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets (e.g., proteins), wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides; amplifying the plurality of barcoded sample indexing oligonucleotides using a plurality of daisy-chaining amplification primers to create a plurality of daisy-chaining elongated amplicons; obtaining sequencing data of the plurality of daisy-chaining elongated amplicons comprising sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides. The method can comprise removing unbound sample indexing compositions of the plurality of sample indexing compositions.

[0253] In some embodiments, the sample indexing sequence is at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, or at least 128 nucleotides in length, or a combination thereof. The sample indexing oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 200-500 nucleotides in length, about 500 nucleotides in length, or a combination thereof. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences. Sample indexing sequences of at least 100 or 1000 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences.

[0254] In some embodiments, the cellular component binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the cellular component binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the cellular component binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

[0255] In some embodiments, at least one sample of the plurality of samples comprises a single cell. The at least one of the one or more cellular component targets can be on a cell surface. A sample of the plurality of samples can comprise a plurality of cells, a tissue, a tumor sample, or any combination thereof. The plurality of samples can comprise a mammalian sample, a bacterial sample, a viral sample, a yeast sample, a fungal sample, or any combination thereof.

[0256] In some embodiments, removing the unbound sample indexing compositions comprises washing the one or more cells from each of the plurality of samples with a washing buffer. The method can comprise lysing the one or more cells from each of the plurality of samples. The sample indexing oligonucleotide can be configured to be detachable or non-detachable from the cellular component binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent by UV photocleaving, chemical treatment (e.g., using a reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0257] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of the cells of the plurality of samples. The sample indexing oligonucleotide can comprise a molecular label sequence, a poly(A) region, or a combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. A target binding region of the barcode can comprise the capture sequence. The target binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide

complementary to the capture sequence of the barcode can comprise a poly(A) tail. The sample indexing oligonucleotide can comprise a molecular label.

**[0258]** In some embodiments, the cellular component target is, or comprises, an extracellular protein, an intracellular protein, or any combination thereof. The cellular component target can be, or comprise, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The cellular component target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The cellular component target can be selected from a group comprising 10-100 different cellular component targets. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with an identical sequence. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second cellular component binding reagent not conjugated with the sample indexing oligonucleotide. The cellular component binding reagent and the second cellular component binding reagent can be identical.

**[0259]** In some embodiments, a barcode of the plurality of barcodes comprises a target binding region and a molecular label sequence. Molecular label sequences of at least two barcodes of the plurality of barcodes comprise different molecule label sequences. The barcode can comprise a cell label, a binding site for a universal primer, or any combination thereof. The target binding region can comprise a poly(dT) region.

**[0260]** In some embodiments, the plurality of barcodes is associated with a particle. At least one barcode of the plurality of barcodes can be immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or any combination thereof. The particle can be degradable. The particle can be a bead. The bead can be selected from the group consisting of streptavidin beads, agarose beads, magnetic beads, conjugated beads, protein A conjugated beads, protein G conjugated beads, protein A/G conjugated beads, protein L conjugated beads, oligo(dT) conjugated beads, silica beads, silica-like beads, anti-biotin microbead, anti-fluorochrome microbead, and any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose,

gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise at least 10000 barcodes. In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000 or 10000 different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences.

[0261] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0262] In some embodiments, amplifying the plurality of barcoded sample indexing oligonucleotides comprises amplifying the plurality of barcoded sample indexing oligonucleotides using polymerase chain reaction (PCR). Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. Obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of daisy-chaining elongated amplicons. Obtaining the sequencing data can comprise sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide.

[0263] In some embodiments, each daisy-chaining amplification primer of the plurality of daisy-chaining amplification primers comprises a barcoded sample indexing oligonucleotide-binding region and an overhang region, wherein the barcoded sample indexing oligonucleotide-binding region is capable of binding to a daisy-chaining sample indexing region of the sample indexing oligonucleotide. The barcoded sample indexing oligonucleotide-binding region can be at least 20 nucleotides in length, at least 30 nucleotides in length, about 40 nucleotides in length, at least 40 nucleotides in length, about 50

nucleotides in length, or a combination thereof. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise barcoded sample indexing oligonucleotide-binding regions with an identical sequence. The plurality of daisy-chaining amplification primers can comprise barcoded sample indexing oligonucleotide-binding regions with an identical sequence. The overhang region can be at least 50 nucleotides in length, at least 100 nucleotides in length, at least 150 nucleotides in length, about 150 nucleotides in length, at least 200 nucleotides in length, or a combination thereof. The overhang region can comprise a daisy-chaining amplification primer barcode sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise overhang regions with an identical daisy-chaining amplification primer barcode sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise overhang regions with two daisy-chaining amplification primer barcode sequences. Overhang regions of the plurality of daisy-chaining amplification primers can comprise different daisy-chaining amplification primer barcode sequences. A daisy-chaining elongated amplicon of the plurality of daisy-chaining elongated amplicons can be at least 250 nucleotides in length, at least 300 nucleotides in length, at least 350 nucleotides in length, at least 400 nucleotides in length, about 400 nucleotides in length, at least 450 nucleotides in length, at least 500 nucleotides in length, or a combination thereof.

[0264] In some embodiments, identifying the sample origin of the at least one cell can comprise identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide. Barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides can comprise stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

[0265] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the

targets with target binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets. Prior to obtaining the sequencing data of the plurality of barcoded targets, the method can comprise amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0266] In some embodiments, each of the plurality of sample indexing compositions comprises the cellular component binding reagent. The sample indexing sequences of the sample indexing oligonucleotides associated with two or more cellular component binding reagents can be identical. The sample indexing sequences of the sample indexing oligonucleotides associated with two or more cellular component binding reagents can comprise different sequences. Each of the plurality of sample indexing compositions can comprise two or more cellular component binding reagents. In some embodiments, the cellular component is an antigen. The cellular component binding reagent can be an antibody.

[0267] Disclosed herein include methods for sample identification. In some embodiments, the method comprise: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more binding targets, wherein each of the plurality of sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more binding targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides using a plurality of daisy-chaining amplification



primers; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides. The method can comprise removing unbound sample indexing compositions of the plurality of sample indexing compositions.

[0268] In some embodiments, the sample indexing sequence is at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, or at least 128 nucleotides in length, or a combination thereof. The sample indexing oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 200-500 nucleotides in length, about 500 nucleotides in length, or a combination thereof. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions comprise different sequences.

[0269] In some embodiments, the cellular component binding reagent comprises a cell surface binding reagent, an antibody, a tetramer, an aptamers, a protein scaffold, an integrin, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the cellular component binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the cellular component binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

[0270] In some embodiments, at least one sample of the plurality of samples comprises a single cell. The at least one of the one or more cellular component targets can be expressed on a cell surface. A sample of the plurality of samples can comprise a plurality of cells, a tissue, a tumor sample, or any combination thereof. The plurality of samples can comprise a mammalian sample, a bacterial sample, a viral sample, a yeast sample, a fungal sample, or any combination thereof.

[0271] In some embodiments, removing the unbound sample indexing compositions comprises washing the one or more cells from each of the plurality of samples with a washing buffer. The method can comprise lysing the one or more cells from each of the plurality of samples. The sample indexing oligonucleotide can be configured to be detachable or non-detachable from the cellular component binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent by UV photocleaving, chemical treatment (e.g., using a reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0272] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of the cells of the plurality of samples. The sample indexing oligonucleotide can comprise a molecular label sequence, a poly(A) region, or a combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. A target binding region of the barcode can comprise the capture sequence. The target binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(A) tail. The sample indexing oligonucleotide can comprise a molecular label.

[0273] In some embodiments, the cellular component target is, or comprises, a carbohydrate, a lipid, a protein, an extracellular protein, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, an intracellular protein, or any combination thereof. The cellular component target can be selected from a group comprising 10-100 different cellular component targets. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with an identical sequence. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second cellular component binding reagent not conjugated with the sample indexing oligonucleotide. The cellular component binding reagent and the second cell binding reagent can be identical.

[0274] In some embodiments, a barcode of the plurality of barcodes comprises a target binding region and a molecular label sequence. Molecular label sequences of at least two barcodes of the plurality of barcodes can comprise different molecule label sequences. The barcode can comprise a cell label, a binding site for a universal primer, or any combination thereof. The target binding region can comprise a poly(dT) region.

[0275] In some embodiments, the plurality of barcodes is enclosed in a particle. The particle can be a bead. At least one barcode of the plurality of barcodes can be immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or any combination thereof. The particle can be degradable. The bead can be selected from the group consisting of streptavidin beads, agarose beads, magnetic beads, conjugated beads, protein A conjugated beads, protein G conjugated beads, protein A/G conjugated beads, protein L conjugated beads, oligo(dT) conjugated beads, silica beads, silica-like beads, anti-biotin microbead, anti-fluorochrome microbead, and any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise at least 10000 barcodes. In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000 or 10000 different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences.

[0276] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0277] In some embodiments, the method comprises amplifying the plurality of

barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR) at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying the plurality of barcoded sample indexing oligonucleotides using the plurality of daisy-chaining amplification primers to produce the plurality of daisy-chaining elongated amplicons. Obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of daisy-chaining elongated amplicons. Obtaining the sequencing data can comprise sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide.

[0278] In some embodiments, a daisy-chaining amplification primer of the plurality of daisy-chaining amplification primers comprises a barcoded sample indexing oligonucleotide-binding region and an overhang region, wherein the barcoded sample indexing oligonucleotide-binding region is capable of binding to a daisy-chaining sample indexing region of the sample indexing oligonucleotide. The barcoded sample indexing oligonucleotide-binding region can be at least 20 nucleotides in length, at least 30 nucleotides in length, about 40 nucleotides in length, at least 40 nucleotides in length, about 50 nucleotides in length, or a combination thereof. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise barcoded sample indexing oligonucleotide-binding regions with an identical sequence. The plurality of daisy-chaining amplification primers can comprise barcoded sample indexing oligonucleotide-binding regions with an identical sequence. The overhang region can be at least 50 nucleotides in length, at least 100 nucleotides in length, at least 150 nucleotides in length, about 150 nucleotides in length, at least 200 nucleotides in length, or a combination thereof. The overhang region can comprise a daisy-chaining amplification primer barcode sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise overhang regions with an identical daisy-chaining amplification primer barcode sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise overhang regions with two daisy-chaining amplification primer barcode sequences. Overhang regions of the plurality of daisy-chaining amplification

primers can comprise different daisy-chaining amplification primer barcode sequences. A daisy-chaining elongated amplicon of the plurality of daisy-chaining elongated amplicons can be at least 250 nucleotides in length, at least 300 nucleotides in length, at least 350 nucleotides in length, at least 400 nucleotides in length, about 400 nucleotides in length, at least 450 nucleotides in length, at least 500 nucleotides in length, or a combination thereof.

[0279] In some embodiments, identifying the sample origin of the at least one cell can comprise identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide. Barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides can comprise stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

[0280] In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets. The method can comprise: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0281] In some embodiments, each of the plurality of sample indexing compositions comprises the cellular component binding reagent. The sample indexing sequences of the sample indexing oligonucleotides associated with two or more cellular component binding reagents can be identical. The sample indexing sequences of the sample

indexing oligonucleotides associated with two or more cellular component binding reagents can comprise different sequences. Each of the plurality of sample indexing compositions can comprise two or more cellular component binding reagents.

[0282] Disclosed herein include methods for sample identification. In some embodiments, the method comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more cellular component targets, wherein each of the plurality of sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one sample indexing oligonucleotide of the plurality of sample indexing compositions and a plurality of daisy-chaining amplification primers. Identifying the sample origin of the at least one cell can comprise: barcoding sample indexing oligonucleotides of the plurality of sample indexing compositions using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying the sample origin of the cell based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides in the sequencing data. The method can, for example, include removing unbound sample indexing compositions of the plurality of sample indexing compositions.

[0283] In some embodiments, the sample indexing sequence is at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, or at least 128 nucleotides in length, or a combination thereof. The sample indexing oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 200-500 nucleotides in length, about 500 nucleotides in length, or a combination

thereof. Sample indexing sequences of at least 10 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences. Sample indexing sequences of at least 100 or 1000 sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences.

[0284] In some embodiments, the cellular component binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the cellular component binding reagent through a linker. The oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the cellular component binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

[0285] In some embodiments, at least one sample of the plurality of samples comprises a single cell. The at least one of the one or more cellular component targets can be expressed on a cell surface. A sample of the plurality of samples can comprise a plurality of cells, a tissue, a tumor sample, or any combination thereof. The plurality of samples can comprise a mammalian sample, a bacterial sample, a viral sample, a yeast sample, a fungal sample, or any combination thereof.

[0286] In some embodiments, removing the unbound sample indexing compositions comprises washing the one or more cells from each of the plurality of samples with a washing buffer. The method can comprise lysing the one or more cells from each of the plurality of samples. The sample indexing oligonucleotide can be configured to be detachable or non-detachable from the cellular component binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent by UV photocleaving, chemical treatment (e.g., using a reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0287] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of the cells of the plurality of samples. The sample indexing oligonucleotide can comprise a molecular label sequence, a poly(A) region, or a

combination thereof. The sample indexing oligonucleotide can comprise a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. A target binding region of the barcode can comprise the capture sequence. The target binding region can comprise a poly(dT) region. The sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(A) tail. The sample indexing oligonucleotide can comprise a molecular label.

[0288] In some embodiments, the cellular component target is, or comprises, a carbohydrate, a lipid, a protein, an extracellular protein, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an intracellular protein, or any combination thereof. The cellular component target can be selected from a group comprising 10-100 different cellular component targets. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with an identical sequence. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second cellular component binding reagent not conjugated with the sample indexing oligonucleotide. The cellular component binding reagent and the second cellular component binding reagent can be identical.

[0289] In some embodiments, identifying the sample origin of the at least one cell comprises: barcoding sample indexing oligonucleotides of the plurality of sample indexing compositions using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying the sample origin of the cell based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides.

[0290] In some embodiments, a barcode of the plurality of barcodes comprises a target binding region and a molecular label sequence. Molecular label sequences of at least two barcodes of the plurality of barcodes can comprise different molecule label sequences. The barcode can comprise a cell label, a binding site for a universal primer, or any combination thereof. The target binding region can comprise a poly(dT) region.

[0291] In some embodiments, the plurality of barcodes is immobilized on a



particle. At least one barcode of the plurality of barcodes can be immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or any combination thereof. The particle can be degradable. The particle can be a bead. The bead can be selected from the group consisting of streptavidin beads, agarose beads, magnetic beads, conjugated beads, protein A conjugated beads, protein G conjugated beads, protein A/G conjugated beads, protein L conjugated beads, oligo(dT) conjugated beads, silica beads, silica-like beads, anti-biotin microbead, anti-fluorochrome microbead, and any combination thereof. The particle can comprise a material selected from the group consisting of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, and any combination thereof. The particle can comprise at least 10000 barcodes. In some embodiments, the barcodes of the particle can comprise molecular label sequences selected from at least 1000 or 10000 different molecular label sequences. The molecular label sequences of the barcodes can comprise random sequences.

[0292] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes can comprise: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0293] In some embodiments, amplifying the plurality of barcoded sample indexing oligonucleotides comprises amplifying the plurality of barcoded sample indexing oligonucleotides using polymerase chain reaction (PCR). Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying the plurality of barcoded sample indexing oligonucleotides using the plurality of

daisy-chaining amplification primers to produce a plurality of daisy-chaining elongated amplicons. Obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of daisy-chaining elongated amplicons. Obtaining the sequencing data can comprise sequencing at least a portion of the molecular label sequence and at least a portion of the sample indexing oligonucleotide.

[0294] In some embodiments, a daisy-chaining amplification primer of the plurality of daisy-chaining amplification primers comprises a barcoded sample indexing oligonucleotide-binding region and an overhang region, wherein the barcoded sample indexing oligonucleotide-binding region is capable of binding to a daisy-chaining sample indexing region of the sample indexing oligonucleotide. The barcoded sample indexing oligonucleotide-binding region can be at least 20 nucleotides in length, at least 30 nucleotides in length, about 40 nucleotides in length, at least 40 nucleotides in length, about 50 nucleotides in length, or a combination thereof. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise barcoded sample indexing oligonucleotide-binding regions with an identical sequence. The plurality of daisy-chaining amplification primers can comprise barcoded sample indexing oligonucleotide-binding regions with an identical sequence. The overhang region can be at least 50 nucleotides in length, at least 100 nucleotides in length, at least 150 nucleotides in length, about 150 nucleotides in length, at least 200 nucleotides in length, or a combination thereof. The overhang region can comprise a daisy-chaining amplification primer barcode sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise overhang regions with an identical daisy-chaining amplification primer barcode sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise overhang regions with two daisy-chaining amplification primer barcode sequences. Overhang regions of the plurality of daisy-chaining amplification primers can comprise different daisy-chaining amplification primer barcode sequences. A daisy-chaining elongated amplicon of the plurality of daisy-chaining elongated amplicons can be at least 250 nucleotides in length, at least 300 nucleotides in length, at least 350 nucleotides in length, at least 400 nucleotides in length, at least 450 nucleotides in length, at least 500 nucleotides in length, or a combination thereof. A daisy-chaining elongated

amplicon of the plurality of daisy-chaining elongated amplicons can be about 200 nucleotides in length, about 250 nucleotides in length, about 300 nucleotides in length, about 350 nucleotides in length, about 400 nucleotides in length, about 450 nucleotides in length, about 500 nucleotides in length, about 600 nucleotides in length, or a range between any two of these values.

[0295] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides comprises stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides. Identifying the sample origin of the at least one cell can comprise identifying the presence or absence of the sample indexing sequence of at least one sample indexing oligonucleotide of the plurality of sample indexing compositions. Identifying the presence or absence of the sample indexing sequence can comprise: replicating the at least one sample indexing oligonucleotide to generate a plurality of replicated sample indexing oligonucleotides; obtaining sequencing data of the plurality of replicated sample indexing oligonucleotides; and identifying the sample origin of the cell based on the sample indexing sequence of a replicated sample indexing oligonucleotide of the plurality of sample indexing oligonucleotides that correspond to the least one barcoded sample indexing oligonucleotide in the sequencing data.

[0296] In some embodiments, replicating the at least one sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides comprises: prior to replicating the at least one barcoded sample indexing oligonucleotide, ligating a replicating adaptor to the at least one barcoded sample indexing oligonucleotide. Replicating the at least one barcoded sample indexing oligonucleotide can comprise replicating the at least one barcoded sample indexing oligonucleotide using the replicating adaptor ligated to the at least one barcoded sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides.

[0297] In some embodiments, replicating the at least one sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides comprises: prior to replicating the at least one barcoded sample indexing oligonucleotide, contacting a capture probe with the at least one sample indexing oligonucleotide to generate a

capture probe hybridized to the sample indexing oligonucleotide; and extending the capture probe hybridized to the sample indexing oligonucleotide to generate a sample indexing oligonucleotide associated with the capture probe. Replicating the at least one sample indexing oligonucleotide can comprise replicating the sample indexing oligonucleotide associated with the capture probe to generate the plurality of replicated sample indexing oligonucleotides.

**[0298]** In some embodiments, the method comprises: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Identifying the sample origin of the at least one barcoded sample indexing oligonucleotide can comprise identifying the sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can comprise: contacting copies of the targets with target binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets. The method can comprise: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

**[0299]** In some embodiments, each of the plurality of sample indexing compositions comprises the cellular component binding reagent. The sample indexing sequences of the sample indexing oligonucleotides associated with the two or more cellular component binding reagents can be identical. The sample indexing sequences of the sample indexing oligonucleotides associated with the two or more cellular component binding reagents can comprise different sequences. Each of the plurality of sample indexing

compositions can comprise the two or more cellular component binding reagents.

[0300] Disclosed herein includes a kit comprising: a plurality of sample indexing compositions; and a plurality of daisy-chaining amplification primers. Each of the plurality of sample indexing compositions can comprise two or more cellular component binding reagents. Each of the two or more cellular component binding reagents can be associated with a sample indexing oligonucleotide. At least one of the two or more cellular component binding reagents can be capable of specifically binding to at least one cellular component target. The sample indexing oligonucleotide can comprise a sample indexing sequence for identifying sample origin of one or more cells of a sample. The sample indexing oligonucleotide can comprise a daisy-chaining amplification primer binding sequence. Sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences.

[0301] A daisy-chaining amplification primer of the plurality of daisy-chaining amplification primers can comprise a barcoded sample indexing oligonucleotide-binding region and an overhang region, wherein the barcoded sample indexing oligonucleotide-binding region is capable of binding to a daisy-chaining sample indexing region of the sample indexing oligonucleotide. The barcoded sample indexing oligonucleotide-binding region can be at least 20 nucleotides in length, at least 30 nucleotides in length, about 40 nucleotides in length, at least 40 nucleotides in length, about 50 nucleotides in length, or a combination thereof. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise barcoded sample indexing oligonucleotide-binding regions with an identical sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise barcoded sample indexing oligonucleotide-binding regions with different sequences. The barcoded sample indexing oligonucleotide-binding regions can comprise the daisy-chaining amplification primer binding sequence, a complement thereof, a reverse complement thereof, or a combination thereof. Daisy-chaining amplification primer binding sequences of at least two sample indexing compositions of the plurality of sample indexing compositions can comprise an identical sequence. Daisy-chaining amplification primer binding sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences. The overhang region can be at least 50 nucleotides in length, at least

100 nucleotides in length, at least 150 nucleotides in length, about 150 nucleotides in length, or a combination thereof. The overhang region can be at least 200 nucleotides in length. The overhang region can comprise a daisy-chaining amplification primer barcode sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise overhang regions with an identical daisy-chaining amplification primer barcode sequence. Two daisy-chaining amplification primers of the plurality of daisy-chaining amplification primers can comprise overhang regions with two daisy-chaining amplification primer barcode sequences. Overhang regions of the plurality of daisy-chaining amplification primers can comprise different daisy-chaining amplification primer barcode sequences.

[0302] In some embodiments, the sample indexing sequence is at least 6 nucleotides in length, 25-45 nucleotides in length, about 128 nucleotides in length, or at least 128 nucleotides in length, or a combination thereof. The sample indexing oligonucleotide can be about 50 nucleotides in length, about 100 nucleotides in length, about 200 nucleotides in length, at least 200 nucleotides in length, less than about 200-300 nucleotides in length, about 200-500 nucleotides in length, about 500 nucleotides in length, or a combination thereof. Sample indexing sequences of at least 10, 100, or 1000 sample indexing compositions of the plurality of sample indexing compositions comprise different sequences.

[0303] In some embodiments, the cellular component binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The sample indexing oligonucleotide can be conjugated to the cellular component binding reagent through a linker. The at least one sample indexing oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the molecule of the cellular component binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

[0304] In some embodiments, the sample indexing oligonucleotide is not homologous to genomic sequences of a species. The sample indexing oligonucleotide can comprise a molecular label sequence, a poly(A) region, or a combination thereof. In some embodiments, at least one sample of the plurality of samples can comprise a single cell, a plurality of cells, a tissue, a tumor sample, or any combination thereof. The sample can comprise a mammalian sample, a bacterial sample, a viral sample, a yeast sample, a fungal

sample, or any combination thereof.

[0305] In some embodiments, the cellular component target is, or comprises, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. The cellular component target can be selected from a group comprising 10-100 different cellular component targets. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with an identical sequence. The cellular component binding reagent can be associated with two or more sample indexing oligonucleotides with different sample indexing sequences. The sample indexing composition of the plurality of sample indexing compositions can comprise a second cellular component binding reagent not conjugated with the sample indexing oligonucleotide. The cellular component binding reagent and the second cellular component binding reagent can be identical.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0306] FIG. 1 illustrates a non-limiting exemplary stochastic barcode.

[0307] FIG. 2 shows a non-limiting exemplary workflow of stochastic barcoding and digital counting.

[0308] FIG. 3 is a schematic illustration showing a non-limiting exemplary process for generating an indexed library of the stochastically barcoded targets from a plurality of targets.

[0309] FIG. 4 shows a schematic illustration of an exemplary protein binding reagent (antibody illustrated here) conjugated with an oligonucleotide comprising a unique identifier for the protein binding reagent.

[0310] FIG. 5 shows a schematic illustration of an exemplary binding reagent (antibody illustrated here) conjugated with an oligonucleotide comprising a unique identifier for sample indexing to determine cells from the same or different samples.

[0311] FIG. 6 shows a schematic illustration of an exemplary workflow using oligonucleotide-conjugated antibodies to determine protein expression and gene expression simultaneously in a high throughput manner.

[0312] FIG. 7 shows a schematic illustration of an exemplary workflow of using oligonucleotide-conjugated antibodies for sample indexing.

[0313] FIGS. 8A1 and 8A2 show a schematic illustration of an exemplary workflow of daisy-chaining a barcoded oligonucleotide. FIGS. 8B1 and 8B2 show a schematic illustration of another exemplary workflow of daisy-chaining a barcoded oligonucleotide.

[0314] FIGS. 9A-9E show non-limiting exemplary schematic illustrations of particles functionalized with oligonucleotides.

[0315] FIG. 10A is a schematic illustration of an exemplary workflow of using particles functionalized with oligonucleotides for single cell sequencing control. FIG. 10B is a schematic illustration of another exemplary workflow of using particles functionalized with oligonucleotides for single cell sequencing control.

[0316] FIG. 11 shows a schematic illustration of an exemplary workflow of using control oligonucleotide-conjugated antibodies for determining single cell sequencing efficiency.

[0317] FIG. 12 shows another schematic illustration of an exemplary workflow of using control oligonucleotide-conjugated antibodies for determining single cell sequencing efficiency.

[0318] FIGS. 13A-13C are plots showing that control oligonucleotides can be used for cell counting.

[0319] FIGS. 14A-14F show a schematic illustration of an exemplary workflow of determining interactions between cellular components (e.g., proteins) using a pair of interaction determination compositions.

[0320] FIGS. 15A-15D show non-limiting exemplary designs of oligonucleotides for determining protein expression and gene expression simultaneously and for sample indexing.

[0321] FIG. 16 shows a schematic illustration of a non-limiting exemplary oligonucleotide sequence for determining protein expression and gene expression simultaneously and for sample indexing.

[0322] FIGS. 17A-17F are non-limiting exemplary tSNE projection plots showing results of using oligonucleotide-conjugated antibodies to measure CD4 protein expression and gene expression simultaneously in a high throughput manner.

[0323] FIGS. 18A-18F are non-limiting exemplary bar charts showing the



expressions of CD4 mRNA and protein in CD4 T cells, CD8 T cells, and Myeloid cells.

[0324] FIG. 19 is a non-limiting exemplary bar chart showing that, with similar sequencing depth, detection sensitivity for CD4 protein level increased with higher ratios of antibody:oligonucleotide, with the 1:3 ratio performing better than the 1:1 and 1:2 ratios.

[0325] FIGS. 20A-20D are plots showing the CD4 protein expression on cell surface of cells sorted using flow cytometry.

[0326] FIG. 21A-21F are non-limiting exemplary bar charts showing the expressions of CD4 mRNA and protein in CD4 T cells, CD8 T cells, and Myeloid cells of two samples.

[0327] FIG. 22 is a non-limiting exemplary bar chart showing detection sensitivity for CD4 protein level determined using different sample preparation protocols with an antibody:oligonucleotide ratio of 1:3.

[0328] FIG. 23 shows a non-limiting exemplary experimental design for performing sample indexing and determining the effects of the lengths of the sample indexing oligonucleotides and the cleavability of sample indexing oligonucleotides on sample indexing.

[0329] FIGS. 24A-24C are non-limiting exemplary tSNE plots showing that the three types of anti-CD147 antibody conjugated with different sample indexing oligonucleotides (cleavable 95mer, non-cleavable 95mer, and cleavable 200mer) can be used for determining the protein expression level of CD147.

[0330] FIG. 25 is a non-limiting exemplary tSNE plot with an overlay of GAPDH expression per cell.

[0331] FIGS. 26A-26C are non-limiting exemplary histograms showing the numbers of molecules of sample indexing oligonucleotides detected using the three types of sample indexing oligonucleotides.

[0332] FIGS. 27A-27C are non-limiting exemplary plots and bar charts showing that CD147 expression was higher in dividing cells.

[0333] FIGS. 28A-28C are non-limiting exemplary tSNE projection plots showing that sample indexing can be used to identify cells of different samples.

[0334] FIGS. 29A-29C are non-limiting exemplary histograms of the sample indexing sequences per cell based on the numbers of molecules of the sample indexing

oligonucleotides determined.

[0335] FIGS. 30A-30D are non-limiting exemplary plots comparing annotations of cell types determined based on mRNA expression of CD3D (for Jurkat cells) and JCHAIN (for Ramos cells) and sample indexing of Jurkat and Ramos cells.

[0336] FIGS. 31A-31C are non-limiting tSNE projection plots of the mRNA expression profiles of Jurkat and Ramos cells with overlays of the mRNA expressions of CD3D and JCHAIN (FIG. 31A), JCHAIN (FIG. 31B), and CD3D (FIG. 31C).

[0337] FIGS. 32 is a non-limiting exemplary tSNE projection plot of expression profiles of Jurkat and Ramos cells with an overlay of the cell types determined using sample indexing with a DBEC cutoff of 250.

[0338] FIGS. 33A-33C are non-limiting exemplary bar charts of the numbers of molecules of sample indexing oligonucleotides per cell for Ramos & Jurkat cells (FIG. 33), Ramos cells (FIG. 33B), and Jurkat cells (FIG. 33C) that were not labeled or labeled with “Short 3” sample indexing oligonucleotides, “Short 2” & “Short 3” sample indexing oligonucleotides, and “Short 2” sample indexing oligonucleotides.

[0339] FIGS. 34A-34C are non-limiting exemplary plots showing that less than 1% of single cells were labeled with both the “Short 2” and “Short 3” sample indexing oligonucleotides.

[0340] FIGS. 35A-35C are non-limiting exemplary tSNE plots showing batch effects on expression profiles of Jurkat and Ramos cells among samples prepared using different flowcells as outlined in FIG. 23.

[0341] FIGS. 36A1-36A9, 36B and 36C are non-limiting exemplary plots showing determination of an optimal dilution of an antibody stock using dilution titration.

[0342] FIG. 37 shows a non-limiting exemplary experimental design for determining a staining concentration of oligonucleotide-conjugated antibodies such that the antibody oligonucleotides account for a desired percentage of total reads in sequencing data.

[0343] FIGS. 38A-38D are non-limiting exemplary bioanalyzer traces showing peaks (indicated by arrows) consistent with the expected size of the antibody oligonucleotide.

[0344] FIGS. 39A1-39A3 and 39B1-39B3 are non-limiting exemplary histograms showing the numbers of molecules of antibody oligonucleotides detected for samples stained with different antibody dilutions and different percentage of the antibody molecules

conjugated with the antibody oligonucleotides (“hot antibody”).

[0345] FIGS. 40A-40C are non-limiting exemplary plots showing that oligonucleotide-conjugated anti-CD147 antibody molecules can be used to label various cell types. The cell types were determined using the expression profiles of 488 genes in a blood panel (FIG. 40A). The cells were stained with a mixture of 10% hot antibody:90% cold antibody prepared using a 1:100 diluted stock, resulting in a clear signal in a histogram showing the numbers of molecules of antibody oligonucleotides detected (FIG. 40B). The labeling of the various cell types by the antibody oligonucleotide is shown in FIG. 40C.

[0346] FIGS. 41A-41C are non-limiting exemplary plots showing that oligonucleotide-conjugated anti-CD147 antibodies can be used to label various cell types. The cell types were determined using the expression profiles of 488 genes in a blood panel (FIG. 41A). The cells were stained with a mixture of 1% hot antibody:99% cold antibody prepared using a 1:100 diluted stock, resulting in no clear signal in a histogram showing the numbers of molecules of antibody oligonucleotides detected (FIG. 41B). The labeling of the various cell types by the antibody oligonucleotide is shown in FIG. 41C.

[0347] FIGS. 42A-42C are non-limiting exemplary plots showing that oligonucleotide-conjugated anti-CD147 antibody molecules can be used to label various cell types. The cell types were determined using the expression profiles of 488 genes in a blood panel (FIG. 42A). The cells were stained with a 1:800 diluted stock, resulting in a clear signal in a histogram showing the numbers of molecules of antibody oligonucleotides detected (FIG. 42B). The labeling of the various cell types by the antibody oligonucleotide is shown in FIG. 42C.

[0348] FIGS. 43A-43B are plots showing the composition of control particle oligonucleotides in a staining buffer and control particle oligonucleotides associated with control particles detected using the workflow illustrated in FIG. 10.

[0349] FIGS. 44A-44B are brightfield images of cells (FIG. 44A, white circles) and control particles (FIG. 44B, black circles) in a hemocytometer.

[0350] FIGS. 45A-45B are phase contrast (FIG. 45A, 10X) and fluorescent (FIG. 45B, 10X) images of control particles bound to oligonucleotide-conjugated antibodies associated with fluorophores.

[0351] FIG. 46 is an image showing cells and a control particle being loaded into

microwells of a cartridge.

[0352] FIGS. 47A-47C are plots showing using an antibody cocktail for sample indexing can increase labeling sensitivity.

[0353] FIG. 48 is a non-limiting exemplary plot showing that multiplets can be identified and removed from sequencing data using sample indexing.

[0354] FIG. 49A is a non-limiting exemplary tSNE projection plot of expression profiles of CD45+ single cells from 12 samples of six tissues from two mice identified as singlets or multiplets using sample indexing oligonucleotides.

[0355] FIG. 49B is a non-limiting exemplary tSNE projection plot of expression profiles of CD45+ single cells of 12 samples of six tissues from two mice with multiplets identified using sample indexing oligonucleotides and shown in FIG. 49A removed.

[0356] FIG. 50A is a non-limiting exemplary tSNE projection plot of expression profiles of CD45+ single cells from two mice with multiplets identified using sample indexing oligonucleotides removed.

[0357] FIGS. 50B and 50C are non-limiting exemplary pie charts showing that after multiplet expression profiles were removed, the two mice, which were biological replicates, exhibited similar expression profiles.

[0358] FIGS. 51A-51F are non-limiting exemplary pie charts showing immune cell profiles of six different tissues with multiplet expression profiles in sequencing data identified and removed using sample indexing oligonucleotides.

[0359] FIGS. 52A-52C are non-limiting exemplary graphs showing expression profiles of macrophages, T cells, and B cells from six different tissues after multiplet expression profiles in sequencing data identified and removed using sample indexing oligonucleotides.

[0360] FIGS. 53A-53D are non-limiting exemplary plots comparing macrophages in the colon and spleen.

[0361] FIG. 54 is a non-limiting exemplary plot showing that tagging cells with sample indexing compositions did not alter mRNA expression profiles in PBMCs.

#### DETAILED DESCRIPTION

[0362] In the following detailed description, reference is made to the

accompanying drawings, which form a part hereof. In the drawings, similar symbols typically identify similar components, unless context dictates otherwise. The illustrative embodiments described in the detailed description, drawings, and claims are not meant to be limiting. Other embodiments may be utilized, and other changes may be made, without departing from the spirit or scope of the subject matter presented herein. It will be readily understood that the aspects of the present disclosure, as generally described herein, and illustrated in the Figures, can be arranged, substituted, combined, separated, and designed in a wide variety of different configurations, all of which are explicitly contemplated herein and made part of the disclosure herein.

[0363] All patents, published patent applications, other publications, and sequences from GenBank, and other databases referred to herein are incorporated by reference in their entirety with respect to the related technology.

[0364] Quantifying small numbers of nucleic acids, for example messenger ribonucleotide acid (mRNA) molecules, is clinically important for determining, for example, the genes that are expressed in a cell at different stages of development or under different environmental conditions. However, it can also be very challenging to determine the absolute number of nucleic acid molecules (e.g., mRNA molecules), especially when the number of molecules is very small. One method to determine the absolute number of molecules in a sample is digital polymerase chain reaction (PCR). Ideally, PCR produces an identical copy of a molecule at each cycle. However, PCR can have disadvantages such that each molecule replicates with a stochastic probability, and this probability varies by PCR cycle and gene sequence, resulting in amplification bias and inaccurate gene expression measurements. Stochastic barcodes with unique molecular labels (also referred to as molecular indexes (MIs)) can be used to count the number of molecules and correct for amplification bias. Stochastic barcoding such as the Precise<sup>TM</sup> assay (Cellular Research, Inc. (Palo Alto, CA)) can correct for bias induced by PCR and library preparation steps by using molecular labels (MLs) to label mRNAs during reverse transcription (RT).

[0365] The Precise<sup>TM</sup> assay can utilize a non-depleting pool of stochastic barcodes with large number, for example 6561 to 65536, unique molecular labels on poly(T) oligonucleotides to hybridize to all poly(A)-mRNAs in a sample during the RT step. A stochastic barcode can comprise a universal PCR priming site. During RT, target gene

molecules react randomly with stochastic barcodes. Each target molecule can hybridize to a stochastic barcode resulting to generate stochastically barcoded complementary ribonucleotide acid (cDNA) molecules). After labeling, stochastically barcoded cDNA molecules from microwells of a microwell plate can be pooled into a single tube for PCR amplification and sequencing. Raw sequencing data can be analyzed to produce the number of reads, the number of stochastic barcodes with unique molecular labels, and the numbers of mRNA molecules.

[0366] Methods for determining mRNA expression profiles of single cells can be performed in a massively parallel manner. For example, the Precise<sup>TM</sup> assay can be used to determine the mRNA expression profiles of more than 10000 cells simultaneously. The number of single cells (e.g., 100s or 1000s of singles) for analysis per sample can be lower than the capacity of the current single cell technology. Pooling of cells from different samples enables improved utilization of the capacity of the current single technology, thus lowering reagents wasted and the cost of single cell analysis. The disclosure provides methods of sample indexing for distinguishing cells of different samples for cDNA library preparation for cell analysis, such as single cell analysis. Pooling of cells from different samples can minimize the variations in cDNA library preparation of cells of different samples, thus enabling more accurate comparisons of different samples.

[0367] Disclosed herein include methods for sample identification. In some embodiments, the method comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more antigen targets, wherein each of the plurality of sample indexing compositions comprises a protein binding reagent associated with a sample indexing oligonucleotide, wherein the protein binding reagent is capable of specifically binding to at least one of the one or more antigen targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; removing unbound sample indexing compositions of the plurality of sample indexing compositions; barcoding (e.g., stochastically barcoding) the sample indexing oligonucleotides using a plurality of barcodes (e.g., stochastic barcodes) to create a plurality of barcoded sample indexing oligonucleotides

(e.g., stochastically barcoded sample indexing oligonucleotides); obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides.

[0368] In some embodiments, the method for sample identification disclosed herein comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more cellular component targets, wherein each of the plurality of sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; removing unbound sample indexing compositions of the plurality of sample indexing compositions; barcoding (e.g., stochastically barcoding) the sample indexing oligonucleotides using a plurality of barcodes (e.g., stochastic barcodes) to create a plurality of barcoded sample indexing oligonucleotides (e.g., stochastically barcoded sample indexing oligonucleotides); obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides.

[0369] Disclosed herein include methods for sample identification. In some embodiments, the method comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more antigen targets, wherein each of the plurality of sample indexing compositions comprises a protein binding reagent associated with a sample indexing oligonucleotide, wherein the protein binding reagent is capable of specifically binding to at least one of the one or more antigen targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample

indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; removing unbound sample indexing compositions of the plurality of sample indexing compositions; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one sample indexing oligonucleotide of the plurality of sample indexing compositions.

[0370] Disclosed herein is a plurality of sample indexing compositions. Each of the plurality of sample indexing compositions can comprise a protein binding reagent associated with a sample indexing oligonucleotide. The protein binding reagent can be capable of specifically binding to at least one antigen target. The sample indexing oligonucleotide can comprise a sample indexing sequence for identifying sample origin of one or more cells of a sample. Sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences.

[0371] Disclosed herein include methods for sample identification. In some embodiments, the method comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein the one or more cells comprises one or more cellular component targets, wherein each of the plurality of sample indexing compositions comprises a cellular component binding reagent (e.g., an antibody) associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets (e.g., proteins), wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides; amplifying the plurality of barcoded sample indexing oligonucleotides using a plurality of daisy-chaining amplification primers to create a plurality of daisy-chaining elongated amplicons; obtaining sequencing data of the plurality of daisy-chaining elongated amplicons comprising sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample



indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides. The method can comprise removing unbound sample indexing compositions of the plurality of sample indexing compositions.

[0372] In some embodiments, the sample indexing methods, kits, and compositions disclosed herein can increase sample throughput (e.g., for rare samples of low cell number, hard to isolate cells, and heterogeneous cells), lower reagent costs, reduce technical errors and batch effects by performing library preparation in a single tube reaction, and/or identify inter-sample multiplet cells during data analysis. In some embodiments, cells of tissues from different lymphoid organs and non-lymphoid organs can be tagged using different sample indexing compositions to, for example, increase sample throughput and reduce, or minimize, batch effects. In some embodiments, immune defense and tissue homeostasis and functions can be investigated using the methods, kits, and compositions of the disclosure.

#### Definitions

[0373] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the present disclosure belongs. *See, e.g.*, Singleton et al., Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, NY 1994); Sambrook et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, NY 1989). For purposes of the present disclosure, the following terms are defined below.

[0374] As used herein, the term “adaptor” can mean a sequence to facilitate amplification or sequencing of associated nucleic acids. The associated nucleic acids can comprise target nucleic acids. The associated nucleic acids can comprise one or more of spatial labels, target labels, sample labels, indexing label, or barcode sequences (e.g., molecular labels). The adapters can be linear. The adaptors can be pre-adenylated adapters. The adaptors can be double- or single-stranded. One or more adaptor can be located on the 5' or 3' end of a nucleic acid. When the adaptors comprise known sequences on the 5' and 3' ends, the known sequences can be the same or different sequences. An adaptor located on the 5' and/or 3' ends of a polynucleotide can be capable of hybridizing to one or more oligonucleotides immobilized on a surface. An adapter can, in some embodiments, comprise

a universal sequence. A universal sequence can be a region of nucleotide sequence that is common to two or more nucleic acid molecules. The two or more nucleic acid molecules can also have regions of different sequence. Thus, for example, the 5' adapters can comprise identical and/or universal nucleic acid sequences and the 3' adapters can comprise identical and/or universal sequences. A universal sequence that may be present in different members of a plurality of nucleic acid molecules can allow the replication or amplification of multiple different sequences using a single universal primer that is complementary to the universal sequence. Similarly, at least one, two (e.g., a pair) or more universal sequences that may be present in different members of a collection of nucleic acid molecules can allow the replication or amplification of multiple different sequences using at least one, two (e.g., a pair) or more single universal primers that are complementary to the universal sequences. Thus, a universal primer includes a sequence that can hybridize to such a universal sequence. The target nucleic acid sequence-bearing molecules may be modified to attach universal adapters (e.g., non-target nucleic acid sequences) to one or both ends of the different target nucleic acid sequences. The one or more universal primers attached to the target nucleic acid can provide sites for hybridization of universal primers. The one or more universal primers attached to the target nucleic acid can be the same or different from each other.

[0375] As used herein, an antibody can be a full-length (e.g., naturally occurring or formed by normal immunoglobulin gene fragment recombinatorial processes) immunoglobulin molecule (e.g., an IgG antibody) or an immunologically active (i.e., specifically binding) portion of an immunoglobulin molecule, like an antibody fragment.

[0376] In some embodiments, an antibody is a functional antibody fragment. For example, an antibody fragment can be a portion of an antibody such as F(ab')<sub>2</sub>, Fab', Fab, Fv, sFv and the like. An antibody fragment can bind with the same antigen that is recognized by the full-length antibody. An antibody fragment can include isolated fragments consisting of the variable regions of antibodies, such as the "Fv" fragments consisting of the variable regions of the heavy and light chains and recombinant single chain polypeptide molecules in which light and heavy variable regions are connected by a peptide linker ("scFv proteins"). Exemplary antibodies can include, but are not limited to, antibodies for cancer cells, antibodies for viruses, antibodies that bind to cell surface receptors (for example, CD8, CD34, and CD45), and therapeutic antibodies.

[0377] As used herein the term “associated” or “associated with” can mean that two or more species are identifiable as being co-located at a point in time. An association can mean that two or more species are or were within a similar container. An association can be an informatics association. For example, digital information regarding two or more species can be stored and can be used to determine that one or more of the species were co-located at a point in time. An association can also be a physical association. In some embodiments, two or more associated species are “tethered”, “attached”, or “immobilized” to one another or to a common solid or semisolid surface. An association may refer to covalent or non-covalent means for attaching labels to solid or semi-solid supports such as beads. An association may be a covalent bond between a target and a label. An association can comprise hybridization between two molecules (such as a target molecule and a label).

[0378] As used herein, the term “complementary” can refer to the capacity for precise pairing between two nucleotides. For example, if a nucleotide at a given position of a nucleic acid is capable of hydrogen bonding with a nucleotide of another nucleic acid, then the two nucleic acids are considered to be complementary to one another at that position. Complementarity between two single-stranded nucleic acid molecules may be “partial,” in which only some of the nucleotides bind, or it may be complete when total complementarity exists between the single-stranded molecules. A first nucleotide sequence can be said to be the “complement” of a second sequence if the first nucleotide sequence is complementary to the second nucleotide sequence. A first nucleotide sequence can be said to be the “reverse complement” of a second sequence, if the first nucleotide sequence is complementary to a sequence that is the reverse (i.e., the order of the nucleotides is reversed) of the second sequence. As used herein, the terms “complement”, “complementary”, and “reverse complement” can be used interchangeably. It is understood from the disclosure that if a molecule can hybridize to another molecule it may be the complement of the molecule that is hybridizing.

[0379] As used herein, the term “digital counting” can refer to a method for estimating a number of target molecules in a sample. Digital counting can include the step of determining a number of unique labels that have been associated with targets in a sample. This methodology, which can be stochastic in nature, transforms the problem of counting molecules from one of locating and identifying identical molecules to a series of yes/no

digital questions regarding detection of a set of predefined labels.

[0380] As used herein, the term “label” or “labels” can refer to nucleic acid codes associated with a target within a sample. A label can be, for example, a nucleic acid label. A label can be an entirely or partially amplifiable label. A label can be entirely or partially sequencable label. A label can be a portion of a native nucleic acid that is identifiable as distinct. A label can be a known sequence. A label can comprise a junction of nucleic acid sequences, for example a junction of a native and non-native sequence. As used herein, the term “label” can be used interchangeably with the terms, “index”, “tag,” or “label-tag.” Labels can convey information. For example, in various embodiments, labels can be used to determine an identity of a sample, a source of a sample, an identity of a cell, and/or a target.

[0381] As used herein, the term “non-depleting reservoirs” can refer to a pool of barcodes (e.g., stochastic barcodes) made up of many different labels. A non-depleting reservoir can comprise large numbers of different barcodes such that when the non-depleting reservoir is associated with a pool of targets each target is likely to be associated with a unique barcode. The uniqueness of each labeled target molecule can be determined by the statistics of random choice, and depends on the number of copies of identical target molecules in the collection compared to the diversity of labels. The size of the resulting set of labeled target molecules can be determined by the stochastic nature of the barcoding process, and analysis of the number of barcodes detected then allows calculation of the number of target molecules present in the original collection or sample. When the ratio of the number of copies of a target molecule present to the number of unique barcodes is low, the labeled target molecules are highly unique (i.e., there is a very low probability that more than one target molecule will have been labeled with a given label).

[0382] As used herein, the term “nucleic acid” refers to a polynucleotide sequence, or fragment thereof. A nucleic acid can comprise nucleotides. A nucleic acid can be exogenous or endogenous to a cell. A nucleic acid can exist in a cell-free environment. A nucleic acid can be a gene or fragment thereof. A nucleic acid can be DNA. A nucleic acid can be RNA. A nucleic acid can comprise one or more analogs (e.g., altered backbone, sugar, or nucleobase). Some non-limiting examples of analogs include: 5-bromouracil, peptide nucleic acid, xeno nucleic acid, morpholinos, locked nucleic acids, glycol nucleic acids, threose nucleic acids, dideoxynucleotides, cordycepin, 7-deaza-GTP, fluorophores

(e.g., rhodamine or fluorescein linked to the sugar), thiol containing nucleotides, biotin linked nucleotides, fluorescent base analogs, CpG islands, methyl-7-guanosine, methylated nucleotides, inosine, thiouridine, pseudouridine, dihydrouridine, queuosine, and wyosine. “Nucleic acid”, “polynucleotide”, “target polynucleotide”, and “target nucleic acid” can be used interchangeably.

[0383] A nucleic acid can comprise one or more modifications (e.g., a base modification, a backbone modification), to provide the nucleic acid with a new or enhanced feature (e.g., improved stability). A nucleic acid can comprise a nucleic acid affinity tag. A nucleoside can be a base-sugar combination. The base portion of the nucleoside can be a heterocyclic base. The two most common classes of such heterocyclic bases are the purines and the pyrimidines. Nucleotides can be nucleosides that further include a phosphate group covalently linked to the sugar portion of the nucleoside. For those nucleosides that include a pentofuranosyl sugar, the phosphate group can be linked to the 2', the 3', or the 5' hydroxyl moiety of the sugar. In forming nucleic acids, the phosphate groups can covalently link adjacent nucleosides to one another to form a linear polymeric compound. In turn, the respective ends of this linear polymeric compound can be further joined to form a circular compound; however, linear compounds are generally suitable. In addition, linear compounds may have internal nucleotide base complementarity and may therefore fold in a manner as to produce a fully or partially double-stranded compound. Within nucleic acids, the phosphate groups can commonly be referred to as forming the internucleoside backbone of the nucleic acid. The linkage or backbone can be a 3' to 5' phosphodiester linkage.

[0384] A nucleic acid can comprise a modified backbone and/or modified internucleoside linkages. Modified backbones can include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone. Suitable modified nucleic acid backbones containing a phosphorus atom therein can include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkyl phosphotriesters, methyl and other alkyl phosphonate such as 3'-alkylene phosphonates, 5'-alkylene phosphonates, chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkyl phosphoramidates, phosphorodiamidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, selenophosphates, and boranophosphates having normal 3'-5'

linkages, 2'-5' linked analogs, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', a 5' to 5' or a 2' to 2' linkage.

[0385] A nucleic acid can comprise polynucleotide backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These can include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH<sub>2</sub> component parts.

[0386] A nucleic acid can comprise a nucleic acid mimetic. The term "mimetic" can be intended to include polynucleotides wherein only the furanose ring or both the furanose ring and the internucleotide linkage are replaced with non-furanose groups, replacement of only the furanose ring can also be referred as being a sugar surrogate. The heterocyclic base moiety or a modified heterocyclic base moiety can be maintained for hybridization with an appropriate target nucleic acid. One such nucleic acid can be a peptide nucleic acid (PNA). In a PNA, the sugar-backbone of a polynucleotide can be replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleotides can be retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. The backbone in PNA compounds can comprise two or more linked aminoethylglycine units which gives PNA an amide containing backbone. The heterocyclic base moieties can be bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

[0387] A nucleic acid can comprise a morpholino backbone structure. For example, a nucleic acid can comprise a 6-membered morpholino ring in place of a ribose ring. In some of these embodiments, a phosphorodiamidate or other non-phosphodiester internucleoside linkage can replace a phosphodiester linkage.

[0388] A nucleic acid can comprise linked morpholino units (e.g., morpholino nucleic acid) having heterocyclic bases attached to the morpholino ring. Linking groups can link the morpholino monomeric units in a morpholino nucleic acid. Non-ionic morpholino-

based oligomeric compounds can have less undesired interactions with cellular proteins. Morpholino-based polynucleotides can be nonionic mimics of nucleic acids. A variety of compounds within the morpholino class can be joined using different linking groups. A further class of polynucleotide mimetic can be referred to as cyclohexenyl nucleic acids (CeNA). The furanose ring normally present in a nucleic acid molecule can be replaced with a cyclohexenyl ring. CeNA DMT protected phosphoramidite monomers can be prepared and used for oligomeric compound synthesis using phosphoramidite chemistry. The incorporation of CeNA monomers into a nucleic acid chain can increase the stability of a DNA/RNA hybrid. CeNA oligoadenylates can form complexes with nucleic acid complements with similar stability to the native complexes. A further modification can include Locked Nucleic Acids (LNAs) in which the 2'-hydroxyl group is linked to the 4' carbon atom of the sugar ring thereby forming a 2'-C, 4'-C-oxymethylene linkage thereby forming a bicyclic sugar moiety. The linkage can be a methylene (-CH<sub>2</sub>), group bridging the 2' oxygen atom and the 4' carbon atom wherein *n* is 1 or 2. LNA and LNA analogs can display very high duplex thermal stabilities with complementary nucleic acid (*T<sub>m</sub>*=+3 to +10 °C), stability towards 3'-exonucleolytic degradation and good solubility properties.

[0389] A nucleic acid may also include nucleobase (often referred to simply as "base") modifications or substitutions. As used herein, "unmodified" or "natural" nucleobases can include the purine bases, (e.g., adenine (A) and guanine (G)), and the pyrimidine bases, (e.g., thymine (T), cytosine (C) and uracil (U)). Modified nucleobases can include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl (—C≡C—CH<sub>3</sub>) uracil and cytosine and other alkynyl derivatives of pyrimidine bases, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 2-F-adenine, 2-aminoadenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Modified nucleobases can include tricyclic pyrimidines such as phenoxazine cytidine(1H-

pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido(5,4-b)(1,4)benzothiazin-2(3H)-one), G-clamps such as a substituted phenoxazine cytidine (e.g., 9-(2-aminoethoxy)-H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido(5,4-b)(1,4)benzothiazin-2(3H)-one), G-clamps such as a substituted phenoxazine cytidine (e.g., 9-(2-aminoethoxy)-H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido(4,5-b)indol-2-one), pyridoindole cytidine (H-pyrido(3',2':4,5)pyrrolo[2,3-d]pyrimidin-2-one).

**[0390]** As used herein, the term “sample” can refer to a composition comprising targets. Suitable samples for analysis by the disclosed methods, devices, and systems include cells, tissues, organs, or organisms.

**[0391]** As used herein, the term “sampling device” or “device” can refer to a device which may take a section of a sample and/or place the section on a substrate. A sample device can refer to, for example, a fluorescence activated cell sorting (FACS) machine, a cell sorter machine, a biopsy needle, a biopsy device, a tissue sectioning device, a microfluidic device, a blade grid, and/or a microtome.

**[0392]** As used herein, the term “solid support” can refer to discrete solid or semi-solid surfaces to which a plurality of barcodes (e.g., stochastic barcodes) may be attached. A solid support may encompass any type of solid, porous, or hollow sphere, ball, bearing, cylinder, or other similar configuration composed of plastic, ceramic, metal, or polymeric material (e.g., hydrogel) onto which a nucleic acid may be immobilized (e.g., covalently or non-covalently). A solid support may comprise a discrete particle that may be spherical (e.g., microspheres) or have a non-spherical or irregular shape, such as cubic, cuboid, pyramidal, cylindrical, conical, oblong, or disc-shaped, and the like. A bead can be non-spherical in shape. A plurality of solid supports spaced in an array may not comprise a substrate. A solid support may be used interchangeably with the term “bead.”

**[0393]** As used herein, the term “stochastic barcode” can refer to a polynucleotide sequence comprising labels of the present disclosure. A stochastic barcode can be a polynucleotide sequence that can be used for stochastic barcoding. Stochastic barcodes can be used to quantify targets within a sample. Stochastic barcodes can be used to control for errors which may occur after a label is associated with a target. For example, a stochastic barcode can be used to assess amplification or sequencing errors. A stochastic barcode



associated with a target can be called a stochastic barcode-target or stochastic barcode-tag-target.

[0394] As used herein, the term “gene-specific stochastic barcode” can refer to a polynucleotide sequence comprising labels and a target-binding region that is gene-specific. A stochastic barcode can be a polynucleotide sequence that can be used for stochastic barcoding. Stochastic barcodes can be used to quantify targets within a sample. Stochastic barcodes can be used to control for errors which may occur after a label is associated with a target. For example, a stochastic barcode can be used to assess amplification or sequencing errors. A stochastic barcode associated with a target can be called a stochastic barcode-target or stochastic barcode-tag-target.

[0395] As used herein, the term “stochastic barcoding” can refer to the random labeling (e.g., barcoding) of nucleic acids. Stochastic barcoding can utilize a recursive Poisson strategy to associate and quantify labels associated with targets. As used herein, the term “stochastic barcoding” can be used interchangeably with “stochastic labeling.”

[0396] As used here, the term “target” can refer to a composition which can be associated with a barcode (e.g., a stochastic barcode). Exemplary suitable targets for analysis by the disclosed methods, devices, and systems include oligonucleotides, DNA, RNA, mRNA, microRNA, tRNA, and the like. Targets can be single or double stranded. In some embodiments, targets can be proteins, peptides, or polypeptides. In some embodiments, targets are lipids. As used herein, “target” can be used interchangeably with “species.”

[0397] As used herein, the term “reverse transcriptases” can refer to a group of enzymes having reverse transcriptase activity (i.e., that catalyze synthesis of DNA from an RNA template). In general, such enzymes include, but are not limited to, retroviral reverse transcriptase, retrotransposon reverse transcriptase, retroplasmid reverse transcriptases, retron reverse transcriptases, bacterial reverse transcriptases, group II intron-derived reverse transcriptase, and mutants, variants or derivatives thereof. Non-retroviral reverse transcriptases include non-LTR retrotransposon reverse transcriptases, retroplasmid reverse transcriptases, retron reverse transcriptases, and group II intron reverse transcriptases. Examples of group II intron reverse transcriptases include the *Lactococcus lactis* LI.LtrB intron reverse transcriptase, the *Thermosynechococcus elongatus* TeI4c intron reverse transcriptase, or the *Geobacillus stearothermophilus* GsI-IIC intron reverse transcriptase.

Other classes of reverse transcriptases can include many classes of non-retroviral reverse transcriptases (i.e., retrons, group II introns, and diversity-generating retroelements among others).

[0398] The terms “universal adaptor primer,” “universal primer adaptor” or “universal adaptor sequence” are used interchangeably to refer to a nucleotide sequence that can be used to hybridize to barcodes (e.g., stochastic barcodes) to generate gene-specific barcodes. A universal adaptor sequence can, for example, be a known sequence that is universal across all barcodes used in methods of the disclosure. For example, when multiple targets are being labeled using the methods disclosed herein, each of the target-specific sequences may be linked to the same universal adaptor sequence. In some embodiments, more than one universal adaptor sequences may be used in the methods disclosed herein. For example, when multiple targets are being labeled using the methods disclosed herein, at least two of the target-specific sequences are linked to different universal adaptor sequences. A universal adaptor primer and its complement may be included in two oligonucleotides, one of which comprises a target-specific sequence and the other comprises a barcode. For example, a universal adaptor sequence may be part of an oligonucleotide comprising a target-specific sequence to generate a nucleotide sequence that is complementary to a target nucleic acid. A second oligonucleotide comprising a barcode and a complementary sequence of the universal adaptor sequence may hybridize with the nucleotide sequence and generate a target-specific barcode (e.g., a target-specific stochastic barcode). In some embodiments, a universal adaptor primer has a sequence that is different from a universal PCR primer used in the methods of this disclosure.

### Barcodes

[0399] Barcoding, such as stochastic barcoding, has been described in, for example, US20150299784, WO2015031691, and Fu et al, Proc Natl Acad Sci U.S.A. 2011 May 31;108(22):9026-31, the content of these publications is incorporated hereby in its entirety. In some embodiments, the barcode disclosed herein can be a stochastic barcode which can be a polynucleotide sequence that may be used to stochastically label (e.g., barcode, tag) a target. Barcodes can be referred to stochastic barcodes if the ratio of the number of different barcode sequences of the stochastic barcodes and the number of

occurrence of any of the targets to be labeled can be, or be about, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, or a number or a range between any two of these values. A target can be an mRNA species comprising mRNA molecules with identical or nearly identical sequences. Barcodes can be referred to as stochastic barcodes if the ratio of the number of different barcode sequences of the stochastic barcodes and the number of occurrence of any of the targets to be labeled is at least, or is at most, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, or 100:1. Barcode sequences of stochastic barcodes can be referred to as molecular labels.

**[0400]** A barcode, for example a stochastic barcode, can comprise one or more labels. Exemplary labels can include a universal label, a cell label, a barcode sequence (e.g., a molecular label), a sample label, a plate label, a spatial label, and/or a pre-spatial label. FIG. 1 illustrates an exemplary barcode 104 with a spatial label. The barcode 104 can comprise a 5' amine that may link the barcode to a solid support 105. The barcode can comprise a universal label, a dimension label, a spatial label, a cell label, and/or a molecular label. The order of different labels (including but not limited to the universal label, the dimension label, the spatial label, the cell label, and the molecule label) in the barcode can vary. For example, as shown in FIG. 1, the universal label may be the 5'-most label, and the molecular label may be the 3'-most label. The spatial label, dimension label, and the cell label may be in any order. In some embodiments, the universal label, the spatial label, the dimension label, the cell label, and the molecular label are in any order. The barcode can comprise a target-binding region. The target-binding region can interact with a target (e.g., target nucleic acid, RNA, mRNA, DNA) in a sample. For example, a target-binding region can comprise an oligo(dT) sequence which can interact with poly(A) tails of mRNAs. In some instances, the labels of the barcode (e.g., universal label, dimension label, spatial label, cell label, and barcode sequence) may be separated by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 or more nucleotides.

**[0401]** A label, for example the cell label, can comprise a unique set of nucleic acid sub-sequences of defined length, e.g., seven nucleotides each (equivalent to the number of bits used in some Hamming error correction codes), which can be designed to provide

error correction capability. The set of error correction sub-sequences comprise seven nucleotide sequences can be designed such that any pairwise combination of sequences in the set exhibits a defined “genetic distance” (or number of mismatched bases), for example, a set of error correction sub-sequences can be designed to exhibit a genetic distance of three nucleotides. In this case, review of the error correction sequences in the set of sequence data for labeled target nucleic acid molecules (described more fully below) can allow one to detect or correct amplification or sequencing errors. In some embodiments, the length of the nucleic acid sub-sequences used for creating error correction codes can vary, for example, they can be, or be about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 31, 40, 50, or a number or a range between any two of these values, nucleotides in length. In some embodiments, nucleic acid sub-sequences of other lengths can be used for creating error correction codes.

[0402] The barcode can comprise a target-binding region. The target-binding region can interact with a target in a sample. The target can be, or comprise, ribonucleic acids (RNAs), messenger RNAs (mRNAs), microRNAs, small interfering RNAs (siRNAs), RNA degradation products, RNAs each comprising a poly(A) tail, or any combination thereof. In some embodiments, the plurality of targets can include deoxyribonucleic acids (DNAs).

[0403] In some embodiments, a target-binding region can comprise an oligo(dT) sequence which can interact with poly(A) tails of mRNAs. One or more of the labels of the barcode (e.g., the universal label, the dimension label, the spatial label, the cell label, and the barcode sequences (e.g., molecular label)) can be separated by a spacer from another one or two of the remaining labels of the barcode. The spacer can be, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20, or more nucleotides. In some embodiments, none of the labels of the barcode is separated by spacer.

#### Universal Labels

[0404] A barcode can comprise one or more universal labels. In some embodiments, the one or more universal labels can be the same for all barcodes in the set of barcodes attached to a given solid support. In some embodiments, the one or more universal labels can be the same for all barcodes attached to a plurality of beads. In some embodiments, a universal label can comprise a nucleic acid sequence that is capable of hybridizing to a sequencing primer. Sequencing primers can be used for sequencing

barcodes comprising a universal label. Sequencing primers (e.g., universal sequencing primers) can comprise sequencing primers associated with high-throughput sequencing platforms. In some embodiments, a universal label can comprise a nucleic acid sequence that is capable of hybridizing to a PCR primer. In some embodiments, the universal label can comprise a nucleic acid sequence that is capable of hybridizing to a sequencing primer and a PCR primer. The nucleic acid sequence of the universal label that is capable of hybridizing to a sequencing or PCR primer can be referred to as a primer binding site. A universal label can comprise a sequence that can be used to initiate transcription of the barcode. A universal label can comprise a sequence that can be used for extension of the barcode or a region within the barcode. A universal label can be, or be about, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or a number or a range between any two of these values, nucleotides in length. For example, a universal label can comprise at least about 10 nucleotides. A universal label can be at least, or be at most, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, or 300 nucleotides in length. In some embodiments, a cleavable linker or modified nucleotide can be part of the universal label sequence to enable the barcode to be cleaved off from the support.

#### Dimension Labels

[0405] A barcode can comprise one or more dimension labels. In some embodiments, a dimension label can comprise a nucleic acid sequence that provides information about a dimension in which the labeling (e.g., stochastic labeling) occurred. For example, a dimension label can provide information about the time at which a target was barcoded. A dimension label can be associated with a time of barcoding (e.g., stochastic barcoding) in a sample. A dimension label can be activated at the time of labeling. Different dimension labels can be activated at different times. The dimension label provides information about the order in which targets, groups of targets, and/or samples were barcoded. For example, a population of cells can be barcoded at the G0 phase of the cell cycle. The cells can be pulsed again with barcodes (e.g., stochastic barcodes) at the G1 phase of the cell cycle. The cells can be pulsed again with barcodes at the S phase of the cell cycle, and so on. Barcodes at each pulse (e.g., each phase of the cell cycle), can comprise different dimension labels. In this way, the dimension label provides information about which targets were labelled at which phase of the cell cycle. Dimension labels can interrogate many

different biological times. Exemplary biological times can include, but are not limited to, the cell cycle, transcription (e.g., transcription initiation), and transcript degradation. In another example, a sample (e.g., a cell, a population of cells) can be labeled before and/or after treatment with a drug and/or therapy. The changes in the number of copies of distinct targets can be indicative of the sample's response to the drug and/or therapy.

[0406] A dimension label can be activatable. An activatable dimension label can be activated at a specific time point. The activatable label can be, for example, constitutively activated (e.g., not turned off). The activatable dimension label can be, for example, reversibly activated (e.g., the activatable dimension label can be turned on and turned off). The dimension label can be, for example, reversibly activatable at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more times. The dimension label can be reversibly activatable, for example, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more times. In some embodiments, the dimension label can be activated with fluorescence, light, a chemical event (e.g., cleavage, ligation of another molecule, addition of modifications (e.g., pegylated, sumoylated, acetylated, methylated, deacetylated, demethylated), a photochemical event (e.g., photocaging), and introduction of a non-natural nucleotide.

[0407] The dimension label can, in some embodiments, be identical for all barcodes (e.g., stochastic barcodes) attached to a given solid support (e.g., a bead), but different for different solid supports (e.g., beads). In some embodiments, at least 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99% or 100%, of barcodes on the same solid support can comprise the same dimension label. In some embodiments, at least 60% of barcodes on the same solid support can comprise the same dimension label. In some embodiments, at least 95% of barcodes on the same solid support can comprise the same dimension label.

[0408] There can be as many as  $10^6$  or more unique dimension label sequences represented in a plurality of solid supports (e.g., beads). A dimension label can be, or be about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or a number or a range between any two of these values, nucleotides in length. A dimension label can be at least, or be at most, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, or 300, nucleotides in length. A dimension label can comprise between about 5 to about 200 nucleotides. A dimension label can comprise between about 10 to about 150 nucleotides. A dimension label can comprise between about 20 to about 125 nucleotides in length.

*Spatial Labels*

[0409] A barcode can comprise one or more spatial labels. In some embodiments, a spatial label can comprise a nucleic acid sequence that provides information about the spatial orientation of a target molecule which is associated with the barcode. A spatial label can be associated with a coordinate in a sample. The coordinate can be a fixed coordinate. For example, a coordinate can be fixed in reference to a substrate. A spatial label can be in reference to a two or three-dimensional grid. A coordinate can be fixed in reference to a landmark. The landmark can be identifiable in space. A landmark can be a structure which can be imaged. A landmark can be a biological structure, for example an anatomical landmark. A landmark can be a cellular landmark, for instance an organelle. A landmark can be a non-natural landmark such as a structure with an identifiable identifier such as a color code, bar code, magnetic property, fluorescents, radioactivity, or a unique size or shape. A spatial label can be associated with a physical partition (e.g., A well, a container, or a droplet). In some embodiments, multiple spatial labels are used together to encode one or more positions in space.

[0410] The spatial label can be identical for all barcodes attached to a given solid support (e.g., a bead), but different for different solid supports (e.g., beads). In some embodiments, the percentage of barcodes on the same solid support comprising the same spatial label can be, or be about, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of barcodes on the same solid support comprising the same spatial label can be at least, or be at most, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, or 100%. In some embodiments, at least 60% of barcodes on the same solid support can comprise the same spatial label. In some embodiments, at least 95% of barcodes on the same solid support can comprise the same spatial label.

[0411] There can be as many as  $10^6$  or more unique spatial label sequences represented in a plurality of solid supports (e.g., beads). A spatial label can be, or be about, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or a number or a range between any two of these values, nucleotides in length. A spatial label can be at least or at most 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, or 300 nucleotides in length. A spatial label can comprise between about 5 to about 200 nucleotides. A spatial label can comprise between

about 10 to about 150 nucleotides. A spatial label can comprise between about 20 to about 125 nucleotides in length.

Cell labels

[0412] A barcode (e.g., a stochastic barcode) can comprise one or more cell labels. In some embodiments, a cell label can comprise a nucleic acid sequence that provides information for determining which target nucleic acid originated from which cell. In some embodiments, the cell label is identical for all barcodes attached to a given solid support (e.g., a bead), but different for different solid supports (e.g., beads). In some embodiments, the percentage of barcodes on the same solid support comprising the same cell label can be, or be about 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of barcodes on the same solid support comprising the same cell label can be, or be about 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, or 100%. For example, at least 60% of barcodes on the same solid support can comprise the same cell label. As another example, at least 95% of barcodes on the same solid support can comprise the same cell label.

[0413] There can be as many as  $10^6$  or more unique cell label sequences represented in a plurality of solid supports (e.g., beads). A cell label can be, or be about, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or a number or a range between any two of these values, nucleotides in length. A cell label can be at least, or be at most, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, or 300 nucleotides in length. For example, a cell label can comprise between about 5 to about 200 nucleotides. As another example, a cell label can comprise between about 10 to about 150 nucleotides. As yet another example, a cell label can comprise between about 20 to about 125 nucleotides in length.

Barcode Sequences

[0414] A barcode can comprise one or more barcode sequences. In some embodiments, a barcode sequence can comprise a nucleic acid sequence that provides identifying information for the specific type of target nucleic acid species hybridized to the barcode. A barcode sequence can comprise a nucleic acid sequence that provides a counter (e.g., that provides a rough approximation) for the specific occurrence of the target nucleic acid species hybridized to the barcode (e.g., target-binding region).

[0415] In some embodiments, a diverse set of barcode sequences are attached to a



given solid support (e.g., a bead). In some embodiments, there can be, or be about,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ , or a number or a range between any two of these values, unique molecular label sequences. For example, a plurality of barcodes can comprise about 6561 barcodes sequences with distinct sequences. As another example, a plurality of barcodes can comprise about 65536 barcode sequences with distinct sequences. In some embodiments, there can be at least, or be at most,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ , or  $10^9$ , unique barcode sequences. The unique molecular label sequences can be attached to a given solid support (e.g., a bead).

[0416] The length of a barcode can be different in different implementations. For example, a barcode can be, or be about, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or a number or a range between any two of these values, nucleotides in length. As another example, a barcode can be at least, or be at most, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, or 300 nucleotides in length.

#### Molecular Labels

[0417] A barcode (e.g., a stochastic barcode) can comprise one or more molecular labels. Molecular labels can include barcode sequences. In some embodiments, a molecular label can comprise a nucleic acid sequence that provides identifying information for the specific type of target nucleic acid species hybridized to the barcode. A molecular label can comprise a nucleic acid sequence that provides a counter for the specific occurrence of the target nucleic acid species hybridized to the barcode (e.g., target-binding region).

[0418] In some embodiments, a diverse set of molecular labels are attached to a given solid support (e.g., a bead). In some embodiments, there can be, or be about,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ , or a number or a range between any two of these values, of unique molecular label sequences. For example, a plurality of barcodes can comprise about 6561 molecular labels with distinct sequences. As another example, a plurality of barcodes can comprise about 65536 molecular labels with distinct sequences. In some embodiments, there can be at least, or be at most,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ , or  $10^9$ , unique molecular label sequences. Barcodes with unique molecular label sequences can be attached to a given solid support (e.g., a bead).

[0419] For stochastic barcoding using a plurality of stochastic barcodes, the ratio of the number of different molecular label sequences and the number of occurrence of any of

the targets can be, or be about, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, 100:1, or a number or a range between any two of these values. A target can be an mRNA species comprising mRNA molecules with identical or nearly identical sequences. In some embodiments, the ratio of the number of different molecular label sequences and the number of occurrence of any of the targets is at least, or is at most, 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 30:1, 40:1, 50:1, 60:1, 70:1, 80:1, 90:1, or 100:1.

[0420] A molecular label can be, or be about, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or a number or a range between any two of these values, nucleotides in length. A molecular label can be at least, or be at most, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200, or 300 nucleotides in length.

#### Target-Binding Region

[0421] A barcode can comprise one or more target binding regions, such as capture probes. In some embodiments, a target-binding region can hybridize with a target of interest. In some embodiments, the target binding regions can comprise a nucleic acid sequence that hybridizes specifically to a target (e.g., target nucleic acid, target molecule, e.g., a cellular nucleic acid to be analyzed), for example to a specific gene sequence. In some embodiments, a target binding region can comprise a nucleic acid sequence that can attach (e.g., hybridize) to a specific location of a specific target nucleic acid. In some embodiments, the target binding region can comprise a nucleic acid sequence that is capable of specific hybridization to a restriction enzyme site overhang (e.g., an EcoRI sticky-end overhang). The barcode can then ligate to any nucleic acid molecule comprising a sequence complementary to the restriction site overhang.

[0422] In some embodiments, a target binding region can comprise a non-specific target nucleic acid sequence. A non-specific target nucleic acid sequence can refer to a sequence that can bind to multiple target nucleic acids, independent of the specific sequence of the target nucleic acid. For example, target binding region can comprise a random multimer sequence, or an oligo(dT) sequence that hybridizes to the poly(A) tail on mRNA molecules. A random multimer sequence can be, for example, a random dimer, trimer, tetramer, pentamer, hexamer, septamer, octamer, nonamer, decamer, or higher multimer

sequence of any length. In some embodiments, the target binding region is the same for all barcodes attached to a given bead. In some embodiments, the target binding regions for the plurality of barcodes attached to a given bead can comprise two or more different target binding sequences. A target binding region can be, or be about, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, or a number or a range between any two of these values, nucleotides in length. A target binding region can be at most about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length.

[0423] In some embodiments, a target-binding region can comprise an oligo(dT) which can hybridize with mRNAs comprising polyadenylated ends. A target-binding region can be gene-specific. For example, a target-binding region can be configured to hybridize to a specific region of a target. A target-binding region can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or a number or a range between any two of these values, nucleotides in length. A target-binding region can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30, nucleotides in length. A target-binding region can be about 5-30 nucleotides in length. When a barcode comprises a gene-specific target-binding region, the barcode can be referred to herein as a gene-specific barcode.

#### Orientation Property

[0424] A stochastic barcode (e.g., a stochastic barcode) can comprise one or more orientation properties which can be used to orient (e.g., align) the barcodes. A barcode can comprise a moiety for isoelectric focusing. Different barcodes can comprise different isoelectric focusing points. When these barcodes are introduced to a sample, the sample can undergo isoelectric focusing in order to orient the barcodes into a known way. In this way, the orientation property can be used to develop a known map of barcodes in a sample. Exemplary orientation properties can include, electrophoretic mobility (e.g., based on size of the barcode), isoelectric point, spin, conductivity, and/or self-assembly. For example, barcodes with an orientation property of self-assembly, can self-assemble into a specific orientation (e.g., nucleic acid nanostructure) upon activation.

#### Affinity Property

[0425] A barcode (e.g., a stochastic barcode) can comprise one or more affinity properties. For example, a spatial label can comprise an affinity property. An affinity

property can include a chemical and/or biological moiety that can facilitate binding of the barcode to another entity (e.g., cell receptor). For example, an affinity property can comprise an antibody, for example, an antibody specific for a specific moiety (e.g., receptor) on a sample. In some embodiments, the antibody can guide the barcode to a specific cell type or molecule. Targets at and/or near the specific cell type or molecule can be labeled (e.g., stochastically labeled). The affinity property can, in some embodiments, provide spatial information in addition to the nucleotide sequence of the spatial label because the antibody can guide the barcode to a specific location. The antibody can be a therapeutic antibody, for example a monoclonal antibody or a polyclonal antibody. The antibody can be humanized or chimeric. The antibody can be a naked antibody or a fusion antibody.

[0426] The antibody can be a full-length (i.e., naturally occurring or formed by normal immunoglobulin gene fragment recombinatorial processes) immunoglobulin molecule (e.g., an IgG antibody) or an immunologically active (i.e., specifically binding) portion of an immunoglobulin molecule, like an antibody fragment.

[0427] The antibody fragment can be, for example, a portion of an antibody such as F(ab')<sub>2</sub>, Fab', Fab, Fv, sFv and the like. In some embodiments, the antibody fragment can bind with the same antigen that is recognized by the full-length antibody. The antibody fragment can include isolated fragments consisting of the variable regions of antibodies, such as the "Fv" fragments consisting of the variable regions of the heavy and light chains and recombinant single chain polypeptide molecules in which light and heavy variable regions are connected by a peptide linker ("scFv proteins"). Exemplary antibodies can include, but are not limited to, antibodies for cancer cells, antibodies for viruses, antibodies that bind to cell surface receptors (CD8, CD34, CD45), and therapeutic antibodies.

#### Universal Adaptor Primer

[0428] A barcode can comprise one or more universal adaptor primers. For example, a gene-specific barcode, such as a gene-specific stochastic barcode, can comprise a universal adaptor primer. A universal adaptor primer can refer to a nucleotide sequence that is universal across all barcodes. A universal adaptor primer can be used for building gene-specific barcodes. A universal adaptor primer can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or a number or a range between any two of these nucleotides in length. A universal adaptor primer can be at

least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length. A universal adaptor primer can be from 5-30 nucleotides in length.

#### Linker

[0429] When a barcode comprises more than one of a type of label (*e.g.*, more than one cell label or more than one barcode sequence, such as one molecular label), the labels may be interspersed with a linker label sequence. A linker label sequence can be at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. A linker label sequence can be at most about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 or more nucleotides in length. In some instances, a linker label sequence is 12 nucleotides in length. A linker label sequence can be used to facilitate the synthesis of the barcode. The linker label can comprise an error-correcting (*e.g.*, Hamming) code.

#### Solid Supports

[0430] Barcodes, such as stochastic barcodes, disclosed herein can, in some embodiments, be associated with a solid support. The solid support can be, for example, a synthetic particle. In some embodiments, some or all of the barcode sequences, such as molecular labels for stochastic barcodes (*e.g.*, the first barcode sequences) of a plurality of barcodes (*e.g.*, the first plurality of barcodes) on a solid support differ by at least one nucleotide. The cell labels of the barcodes on the same solid support can be the same. The cell labels of the barcodes on different solid supports can differ by at least one nucleotide. For example, first cell labels of a first plurality of barcodes on a first solid support can have the same sequence, and second cell labels of a second plurality of barcodes on a second solid support can have the same sequence. The first cell labels of the first plurality of barcodes on the first solid support and the second cell labels of the second plurality of barcodes on the second solid support can differ by at least one nucleotide. A cell label can be, for example, about 5-20 nucleotides long. A barcode sequence can be, for example, about 5-20 nucleotides long. The synthetic particle can be, for example, a bead.

[0431] The bead can be, for example, a silica gel bead, a controlled pore glass bead, a magnetic bead, a Dynabead, a Sephadex/Sepharose bead, a cellulose bead, a polystyrene bead, or any combination thereof. The bead can comprise a material such as

polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, or any combination thereof.

[0432] In some embodiments, the bead can be a polymeric bead, for example a deformable bead or a gel bead, functionalized with barcodes or stochastic barcodes (such as gel beads from 10X Genomics (San Francisco, CA). In some implementation, a gel bead can comprise a polymer based gels. Gel beads can be generated, for example, by encapsulating one or more polymeric precursors into droplets. Upon exposure of the polymeric precursors to an accelerator (e.g., tetramethylethylenediamine (TEMED)), a gel bead may be generated.

[0433] In some embodiments, the particle can be degradable. For example, the polymeric bead can dissolve, melt, or degrade, for example, under a desired condition. The desired condition can include an environmental condition. The desired condition may result in the polymeric bead dissolving, melting, or degrading in a controlled manner. A gel bead may dissolve, melt, or degrade due to a chemical stimulus, a physical stimulus, a biological stimulus, a thermal stimulus, a magnetic stimulus, an electric stimulus, a light stimulus, or any combination thereof.

[0434] Analytes and/or reagents, such as oligonucleotide barcodes, for example, may be coupled/immobilized to the interior surface of a gel bead (e.g., the interior accessible via diffusion of an oligonucleotide barcode and/or materials used to generate an oligonucleotide barcode) and/or the outer surface of a gel bead or any other microcapsule described herein. Coupling/immobilization may be via any form of chemical bonding (e.g., covalent bond, ionic bond) or physical phenomena (e.g., Van der Waals forces, dipole-dipole interactions, etc.). In some embodiments, coupling/immobilization of a reagent to a gel bead or any other microcapsule described herein may be reversible, such as, for example, via a labile moiety (e.g., via a chemical cross-linker, including chemical cross-linkers described herein). Upon application of a stimulus, the labile moiety may be cleaved and the immobilized reagent set free. In some embodiments, the labile moiety is a disulfide bond. For example, in the case where an oligonucleotide barcode is immobilized to a gel bead via a disulfide bond, exposure of the disulfide bond to a reducing agent can cleave the disulfide bond and free the oligonucleotide barcode from the bead. The labile moiety may be included as part of a gel bead or microcapsule, as part of a chemical linker that links a reagent or

analyte to a gel bead or microcapsule, and/or as part of a reagent or analyte. In some embodiments, at least one barcode of the plurality of barcodes can be immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or any combination thereof.

[0435] In some embodiments, a gel bead can comprise a wide range of different polymers including but not limited to: polymers, heat sensitive polymers, photosensitive polymers, magnetic polymers, pH sensitive polymers, salt-sensitive polymers, chemically sensitive polymers, polyelectrolytes, polysaccharides, peptides, proteins, and/or plastics. Polymers may include but are not limited to materials such as poly(N-isopropylacrylamide) (PNIPAAm), poly(styrene sulfonate) (PSS), poly(allyl amine) (PAAm), poly(acrylic acid) (PAA), poly(ethylene imine) (PEI), poly(diallyldimethyl-ammonium chloride) (PDADMAC), poly(pyrolle) (PPy), poly(vinylpyrrolidone) (PVPON), poly(vinyl pyridine) (PVP), poly(methacrylic acid) (PMAA), poly(methyl methacrylate) (PMMA), polystyrene (PS), poly(tetrahydrofuran) (PTHF), poly(phthalaldehyde) (PTHF), poly(hexyl viologen) (PHV), poly(L-lysine) (PLL), poly(L-arginine) (PARG), poly(lactic-co-glycolic acid) (PLGA).

[0436] Numerous chemical stimuli can be used to trigger the disruption, dissolution, or degradation of the beads. Examples of these chemical changes may include, but are not limited to pH-mediated changes to the bead wall, disintegration of the bead wall via chemical cleavage of crosslink bonds, triggered depolymerization of the bead wall, and bead wall switching reactions. Bulk changes may also be used to trigger disruption of the beads.

[0437] Bulk or physical changes to the microcapsule through various stimuli also offer many advantages in designing capsules to release reagents. Bulk or physical changes occur on a macroscopic scale, in which bead rupture is the result of mechano-physical forces induced by a stimulus. These processes may include, but are not limited to pressure induced rupture, bead wall melting, or changes in the porosity of the bead wall.

[0438] Biological stimuli may also be used to trigger disruption, dissolution, or degradation of beads. Generally, biological triggers resemble chemical triggers, but many examples use biomolecules, or molecules commonly found in living systems such as enzymes, peptides, saccharides, fatty acids, nucleic acids and the like. For example, beads

may comprise polymers with peptide cross-links that are sensitive to cleavage by specific proteases. More specifically, one example may comprise a microcapsule comprising GFLGK peptide cross links. Upon addition of a biological trigger such as the protease Cathepsin B, the peptide cross links of the shell wall are cleaved and the contents of the beads are released. In other cases, the proteases may be heat-activated. In another example, beads comprise a shell wall comprising cellulose. Addition of the hydrolytic enzyme chitosan serves as biologic trigger for cleavage of cellulosic bonds, depolymerization of the shell wall, and release of its inner contents.

[0439] The beads may also be induced to release their contents upon the application of a thermal stimulus. A change in temperature can cause a variety changes to the beads. A change in heat may cause melting of a bead such that the bead wall disintegrates. In other cases, the heat may increase the internal pressure of the inner components of the bead such that the bead ruptures or explodes. In still other cases, the heat may transform the bead into a shrunken dehydrated state. The heat may also act upon heat-sensitive polymers within the wall of a bead to cause disruption of the bead.

[0440] Inclusion of magnetic nanoparticles to the bead wall of microcapsules may allow triggered rupture of the beads as well as guide the beads in an array. A device of this disclosure may comprise magnetic beads for either purpose. In one example, incorporation of  $\text{Fe}_3\text{O}_4$  nanoparticles into polyelectrolyte containing beads triggers rupture in the presence of an oscillating magnetic field stimulus.

[0441] A bead may also be disrupted, dissolved, or degraded as the result of electrical stimulation. Similar to magnetic particles described in the previous section, electrically sensitive beads can allow for both triggered rupture of the beads as well as other functions such as alignment in an electric field, electrical conductivity or redox reactions. In one example, beads containing electrically sensitive material are aligned in an electric field such that release of inner reagents can be controlled. In other examples, electrical fields may induce redox reactions within the bead wall itself that may increase porosity.

[0442] A light stimulus may also be used to disrupt the beads. Numerous light triggers are possible and may include systems that use various molecules such as nanoparticles and chromophores capable of absorbing photons of specific ranges of wavelengths. For example, metal oxide coatings can be used as capsule triggers. UV



irradiation of polyelectrolyte capsules coated with SiO<sub>2</sub> may result in disintegration of the bead wall. In yet another example, photo switchable materials such as azobenzene groups may be incorporated in the bead wall. Upon the application of UV or visible light, chemicals such as these undergo a reversible cis-to-trans isomerization upon absorption of photons. In this aspect, incorporation of photon switches result in a bead wall that may disintegrate or become more porous upon the application of a light trigger.

[0443] For example, in a non-limiting example of barcoding (e.g., stochastic barcoding) illustrated in FIG. 2, after introducing cells such as single cells onto a plurality of microwells of a microwell array at block 208, beads can be introduced onto the plurality of microwells of the microwell array at block 212. Each microwell can comprise one bead. The beads can comprise a plurality of arcodes. A barcode can comprise a 5' amine region attached to a bead. The barcode can comprise a universal label, a barcode sequence (e.g., a molecular label), a target-binding region, or any combination thereof.

[0444] The barcodes disclosed herein can be associated with (e.g., attached to) a solid support (e.g., a bead). The barcodes associated with a solid support can each comprise a barcode sequence selected from a group comprising at least 100 or 1000 barcode sequences with unique sequences. In some embodiments, different barcodes associated with a solid support can comprise barcode with different sequences. In some embodiments, a percentage of barcodes associated with a solid support comprises the same cell label. For example, the percentage can be, or be about 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, 100%, or a number or a range between any two of these values. As another example, the percentage can be at least, or be at most 60%, 70%, 80%, 85%, 90%, 95%, 97%, 99%, or 100%. In some embodiments, barcodes associated with a solid support can have the same cell label. The barcodes associated with different solid supports can have different cell labels selected from a group comprising at least 100 or 1000 cell labels with unique sequences.

[0445] The barcodes disclosed herein can be associated to (e.g., attached to) a solid support (e.g., a bead). In some embodiments, barcoding the plurality of targets in the sample can be performed with a solid support including a plurality of synthetic particles associated with the plurality of barcodes. In some embodiments, the solid support can include a plurality of synthetic particles associated with the plurality of barcodes. The spatial labels of the plurality of barcodes on different solid supports can differ by at least one

nucleotide. The solid support can, for example, include the plurality of barcodes in two dimensions or three dimensions. The synthetic particles can be beads. The beads can be silica gel beads, controlled pore glass beads, magnetic beads, Dynabeads, Sephadex/Sepharose beads, cellulose beads, polystyrene beads, or any combination thereof. The solid support can include a polymer, a matrix, a hydrogel, a needle array device, an antibody, or any combination thereof. In some embodiments, the solid supports can be free floating. In some embodiments, the solid supports can be embedded in a semi-solid or solid array. The barcodes may not be associated with solid supports. The barcodes can be individual nucleotides. The barcodes can be associated with a substrate.

[0446] As used herein, the terms “tethered,” “attached,” and “immobilized,” are used interchangeably, and can refer to covalent or non-covalent means for attaching barcodes to a solid support. Any of a variety of different solid supports can be used as solid supports for attaching pre-synthesized barcodes or for *in situ* solid-phase synthesis of barcode.

[0447] In some embodiments, the solid support is a bead. The bead can comprise one or more types of solid, porous, or hollow sphere, ball, bearing, cylinder, or other similar configuration which a nucleic acid can be immobilized (e.g., covalently or non-covalently). The bead can be, for example, composed of plastic, ceramic, metal, polymeric material, or any combination thereof. A bead can be, or comprise, a discrete particle that is spherical (e.g., microspheres) or have a non-spherical or irregular shape, such as cubic, cuboid, pyramidal, cylindrical, conical, oblong, or disc-shaped, and the like. In some embodiments, a bead can be non-spherical in shape.

[0448] Beads can comprise a variety of materials including, but not limited to, paramagnetic materials (e.g., magnesium, molybdenum, lithium, and tantalum), superparamagnetic materials (e.g., ferrite ( $\text{Fe}_3\text{O}_4$ ; magnetite) nanoparticles), ferromagnetic materials (e.g., iron, nickel, cobalt, some alloys thereof, and some rare earth metal compounds), ceramic, plastic, glass, polystyrene, silica, methylstyrene, acrylic polymers, titanium, latex, Sepharose, agarose, hydrogel, polymer, cellulose, nylon, or any combination thereof.

[0449] In some embodiments, the bead (e.g., the bead to which the labels are attached) is a hydrogel bead. In some embodiments, the bead comprises hydrogel.

[0450] Some embodiments disclosed herein include one or more particles (for

example, beads). Each of the particles can comprise a plurality of oligonucleotides (e.g., barcodes). Each of the plurality of oligonucleotides can comprise a barcode sequence (e.g., a molecular label sequence), a cell label, and a target-binding region (e.g., an oligo(dT) sequence, a gene-specific sequence, a random multimer, or a combination thereof). The cell label sequence of each of the plurality of oligonucleotides can be the same. The cell label sequences of oligonucleotides on different particles can be different such that the oligonucleotides on different particles can be identified. The number of different cell label sequences can be different in different implementations. In some embodiments, the number of cell label sequences can be, or be about 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ , a number or a range between any two of these values, or more. In some embodiments, the number of cell label sequences can be at least, or be at most 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000,  $10^6$ ,  $10^7$ ,  $10^8$ , or  $10^9$ . In some embodiments, no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or more of the plurality of the particles include oligonucleotides with the same cell sequence. In some embodiment, the plurality of particles that include oligonucleotides with the same cell sequence can be at most 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, or more. In some embodiments, none of the plurality of the particles has the same cell label sequence.

**[0451]** The plurality of oligonucleotides on each particle can comprise different barcode sequences (e.g., molecular labels). In some embodiments, the number of barcode sequences can be, or be about 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ , or a number or a range between any two of these values. In some embodiments, the number of barcode sequences can be at least, or be at most 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000,  $10^6$ ,  $10^7$ ,  $10^8$ , or  $10^9$ . For example, at least 100 of the plurality of oligonucleotides comprise different barcode sequences. As another example, in a single particle, at least 100,

500, 1000, 5000, 10000, 15000, 20000, 50000, a number or a range between any two of these values, or more of the plurality of oligonucleotides comprise different barcode sequences. Some embodiments provide a plurality of the particles comprising barcodes. In some embodiments, the ratio of an occurrence (or a copy or a number) of a target to be labeled and the different barcode sequences can be at least 1:1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:30, 1:40, 1:50, 1:60, 1:70, 1:80, 1:90, or more. In some embodiments, each of the plurality of oligonucleotides further comprises a sample label, a universal label, or both. The particle can be, for example, a nanoparticle or microparticle.

**[0452]** The size of the beads can vary. For example, the diameter of the bead can range from 0.1 micrometer to 50 micrometer. In some embodiments, the diameter of the bead can be, or be about, 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50 micrometer, or a number or a range between any two of these values.

**[0453]** The diameter of the bead can be related to the diameter of the wells of the substrate. In some embodiments, the diameter of the bead can be, or be about, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, or a number or a range between any two of these values, longer or shorter than the diameter of the well. The diameter of the beads can be related to the diameter of a cell (e.g., a single cell entrapped by a well of the substrate). In some embodiments, the diameter of the bead can be at least, or be at most, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% longer or shorter than the diameter of the well. The diameter of the beads can be related to the diameter of a cell (e.g., a single cell entrapped by a well of the substrate). In some embodiments, the diameter of the bead can be, or be about, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200%, 250%, 300%, or a number or a range between any two of these values, longer or shorter than the diameter of the cell. In some embodiments, the diameter of the beads can be at least, or be at most, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200%, 250%, or 300% longer or shorter than the diameter of the cell.

**[0454]** A bead can be attached to and/or embedded in a substrate. A bead can be attached to and/or embedded in a gel, hydrogel, polymer and/or matrix. The spatial position of a bead within a substrate (e.g., gel, matrix, scaffold, or polymer) can be identified using the spatial label present on the barcode on the bead which can serve as a location address.

[0455] Examples of beads can include, but are not limited to, streptavidin beads, agarose beads, magnetic beads, Dynabeads®, MACS® microbeads, antibody conjugated beads (e.g., anti-immunoglobulin microbeads), protein A conjugated beads, protein G conjugated beads, protein A/G conjugated beads, protein L conjugated beads, oligo(dT) conjugated beads, silica beads, silica-like beads, anti-biotin microbeads, anti-fluorochrome microbeads, and BcMag™ Carboxyl-Terminated Magnetic Beads.

[0456] A bead can be associated with (e.g., impregnated with) quantum dots or fluorescent dyes to make it fluorescent in one fluorescence optical channel or multiple optical channels. A bead can be associated with iron oxide or chromium oxide to make it paramagnetic or ferromagnetic. Beads can be identifiable. For example, a bead can be imaged using a camera. A bead can have a detectable code associated with the bead. For example, a bead can comprise a barcode. A bead can change size, for example, due to swelling in an organic or inorganic solution. A bead can be hydrophobic. A bead can be hydrophilic. A bead can be biocompatible.

[0457] A solid support (e.g., a bead) can be visualized. The solid support can comprise a visualizing tag (e.g., fluorescent dye). A solid support (e.g., a bead) can be etched with an identifier (e.g., a number). The identifier can be visualized through imaging the beads.

[0458] A solid support can comprise an insoluble, semi-soluble, or insoluble material. A solid support can be referred to as “functionalized” when it includes a linker, a scaffold, a building block, or other reactive moiety attached thereto, whereas a solid support may be “nonfunctionalized” when it lack such a reactive moiety attached thereto. The solid support can be employed free in solution, such as in a microtiter well format; in a flow-through format, such as in a column; or in a dipstick.

[0459] The solid support can comprise a membrane, paper, plastic, coated surface, flat surface, glass, slide, chip, or any combination thereof. A solid support can take the form of resins, gels, microspheres, or other geometric configurations. A solid support can comprise silica chips, microparticles, nanoparticles, plates, arrays, capillaries, flat supports such as glass fiber filters, glass surfaces, metal surfaces (steel, gold silver, aluminum, silicon and copper), glass supports, plastic supports, silicon supports, chips, filters, membranes, microwell plates, slides, plastic materials including multiwell plates or membranes (e.g.,

formed of polyethylene, polypropylene, polyamide, polyvinylidenedifluoride), and/or wafers, combs, pins or needles (e.g., arrays of pins suitable for combinatorial synthesis or analysis) or beads in an array of pits or nanoliter wells of flat surfaces such as wafers (e.g., silicon wafers), wafers with pits with or without filter bottoms.

[0460] The solid support can comprise a polymer matrix (e.g., gel, hydrogel). The polymer matrix may be able to permeate intracellular space (e.g., around organelles). The polymer matrix may be able to be pumped throughout the circulatory system.

#### Substrates and Microwell Array

[0461] As used herein, a substrate can refer to a type of solid support. A substrate can refer to a solid support that can comprise barcodes or stochastic barcodes of the disclosure. A substrate can, for example, comprise a plurality of microwells. For example, a substrate can be a well array comprising two or more microwells. In some embodiments, a microwell can comprise a small reaction chamber of defined volume. In some embodiments, a microwell can entrap one or more cells. In some embodiments, a microwell can entrap only one cell. In some embodiments, a microwell can entrap one or more solid supports. In some embodiments, a microwell can entrap only one solid support. In some embodiments, a microwell entraps a single cell and a single solid support (e.g., a bead). A microwell can comprise barcode reagents of the disclosure.

#### Methods of Barcoding

[0462] The disclosure provides for methods for estimating the number of distinct targets at distinct locations in a physical sample (e.g., tissue, organ, tumor, cell). The methods can comprise placing barcodes (e.g., stochastic barcodes) in close proximity with the sample, lysing the sample, associating distinct targets with the barcodes, amplifying the targets and/or digitally counting the targets. The method can further comprise analyzing and/or visualizing the information obtained from the spatial labels on the barcodes. In some embodiments, a method comprises visualizing the plurality of targets in the sample. Mapping the plurality of targets onto the map of the sample can include generating a two dimensional map or a three dimensional map of the sample. The two dimensional map and the three dimensional map can be generated prior to or after barcoding (e.g., stochastically

barcoding) the plurality of targets in the sample. Visualizing the plurality of targets in the sample can include mapping the plurality of targets onto a map of the sample. Mapping the plurality of targets onto the map of the sample can include generating a two dimensional map or a three dimensional map of the sample. The two dimensional map and the three dimensional map can be generated prior to or after barcoding the plurality of targets in the sample. In some embodiments, the two dimensional map and the three dimensional map can be generated before or after lysing the sample. Lysing the sample before or after generating the two dimensional map or the three dimensional map can include heating the sample, contacting the sample with a detergent, changing the pH of the sample, or any combination thereof.

[0463] In some embodiments, barcoding the plurality of targets comprises hybridizing a plurality of barcodes with a plurality of targets to create barcoded targets (e.g., stochastically barcoded targets). Barcoding the plurality of targets can comprise generating an indexed library of the barcoded targets. Generating an indexed library of the barcoded targets can be performed with a solid support comprising the plurality of barcodes (e.g., stochastic barcodes).

Contacting a Sample and a Barcode

[0464] The disclosure provides for methods for contacting a sample (e.g., cells) to a substrate of the disclosure. A sample comprising, for example, a cell, organ, or tissue thin section, can be contacted to barcodes (e.g., stochastic barcodes). The cells can be contacted, for example, by gravity flow wherein the cells can settle and create a monolayer. The sample can be a tissue thin section. The thin section can be placed on the substrate. The sample can be one-dimensional (e.g., forms a planar surface). The sample (e.g., cells) can be spread across the substrate, for example, by growing/culturing the cells on the substrate.

[0465] When barcodes are in close proximity to targets, the targets can hybridize to the barcode. The barcodes can be contacted at a non-depletable ratio such that each distinct target can associate with a distinct barcode of the disclosure. To ensure efficient association between the target and the barcode, the targets can be cross-linked to barcode.

Cell Lysis

[0466] Following the distribution of cells and barcodes, the cells can be lysed to liberate the target molecules. Cell lysis can be accomplished by any of a variety of means,

for example, by chemical or biochemical means, by osmotic shock, or by means of thermal lysis, mechanical lysis, or optical lysis. Cells can be lysed by addition of a cell lysis buffer comprising a detergent (e.g., SDS, Li dodecyl sulfate, Triton X-100, Tween-20, or NP-40), an organic solvent (e.g., methanol or acetone), or digestive enzymes (e.g., proteinase K, pepsin, or trypsin), or any combination thereof. To increase the association of a target and a barcode, the rate of the diffusion of the target molecules can be altered by for example, reducing the temperature and/or increasing the viscosity of the lysate.

[0467] In some embodiments, the sample can be lysed using a filter paper. The filter paper can be soaked with a lysis buffer on top of the filter paper. The filter paper can be applied to the sample with pressure which can facilitate lysis of the sample and hybridization of the targets of the sample to the substrate.

[0468] In some embodiments, lysis can be performed by mechanical lysis, heat lysis, optical lysis, and/or chemical lysis. Chemical lysis can include the use of digestive enzymes such as proteinase K, pepsin, and trypsin. Lysis can be performed by the addition of a lysis buffer to the substrate. A lysis buffer can comprise Tris HCl. A lysis buffer can comprise at least about 0.01, 0.05, 0.1, 0.5, or 1 M or more Tris HCl. A lysis buffer can comprise at most about 0.01, 0.05, 0.1, 0.5, or 1 M or more Tris HCL. A lysis buffer can comprise about 0.1 M Tris HCl. The pH of the lysis buffer can be at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more. The pH of the lysis buffer can be at most about 1, 2, 3, 4, 5, 6, 7, 8, 9,10, or more. In some embodiments, the pH of the lysis buffer is about 7.5. The lysis buffer can comprise a salt (e.g., LiCl). The concentration of salt in the lysis buffer can be at least about 0.1, 0.5, or 1 M or more. The concentration of salt in the lysis buffer can be at most about 0.1, 0.5, or 1 M or more. In some embodiments, the concentration of salt in the lysis buffer is about 0.5M. The lysis buffer can comprise a detergent (e.g., SDS, Li dodecyl sulfate, triton X, tween, NP-40). The concentration of the detergent in the lysis buffer can be at least about 0.0001%, 0.0005%, 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, or 7%, or more. The concentration of the detergent in the lysis buffer can be at most about 0.0001%, 0.0005%, 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, or 7%, or more. In some embodiments, the concentration of the detergent in the lysis buffer is about 1% Li dodecyl sulfate. The time used in the method for lysis can be dependent on the amount of detergent used. In some embodiments, the more detergent used,



the less time needed for lysis. The lysis buffer can comprise a chelating agent (e.g., EDTA, EGTA). The concentration of a chelating agent in the lysis buffer can be at least about 1, 5, 10, 15, 20, 25, or 30 mM or more. The concentration of a chelating agent in the lysis buffer can be at most about 1, 5, 10, 15, 20, 25, or 30mM or more. In some embodiments, the concentration of chelating agent in the lysis buffer is about 10 mM. The lysis buffer can comprise a reducing reagent (e.g., beta-mercaptoethanol, DTT). The concentration of the reducing reagent in the lysis buffer can be at least about 1, 5, 10, 15, or 20 mM or more. The concentration of the reducing reagent in the lysis buffer can be at most about 1, 5, 10, 15, or 20 mM or more. In some embodiments, the concentration of reducing reagent in the lysis buffer is about 5 mM. In some embodiments, a lysis buffer can comprise about 0.1M TrisHCl, about pH 7.5, about 0.5M LiCl, about 1% lithium dodecyl sulfate, about 10mM EDTA, and about 5mM DTT.

[0469] Lysis can be performed at a temperature of about 4, 10, 15, 20, 25, or 30 °C. Lysis can be performed for about 1, 5, 10, 15, or 20 or more minutes. A lysed cell can comprise at least about 100000, 200000, 300000, 400000, 500000, 600000, or 700000 or more target nucleic acid molecules. A lysed cell can comprise at most about 100000, 200000, 300000, 400000, 500000, 600000, or 700000 or more target nucleic acid molecules.

Attachment of Barcodes to Target Nucleic Acid Molecules

[0470] Following lysis of the cells and release of nucleic acid molecules therefrom, the nucleic acid molecules can randomly associate with the barcodes of the co-localized solid support. Association can comprise hybridization of a barcode's target recognition region to a complementary portion of the target nucleic acid molecule (e.g., oligo(dT) of the barcode can interact with a poly(A) tail of a target). The assay conditions used for hybridization (e.g., buffer pH, ionic strength, temperature, etc.) can be chosen to promote formation of specific, stable hybrids. In some embodiments, the nucleic acid molecules released from the lysed cells can associate with the plurality of probes on the substrate (e.g., hybridize with the probes on the substrate). When the probes comprise oligo(dT), mRNA molecules can hybridize to the probes and be reverse transcribed. The oligo(dT) portion of the oligonucleotide can act as a primer for first strand synthesis of the cDNA molecule. For example, in a non-limiting example of barcoding illustrated in FIG. 2, at block 216, mRNA molecules can hybridize to barcodes on beads. For example, single-

stranded nucleotide fragments can hybridize to the target-binding regions of barcodes.

[0471] Attachment can further comprise ligation of a barcode's target recognition region and a portion of the target nucleic acid molecule. For example, the target binding region can comprise a nucleic acid sequence that can be capable of specific hybridization to a restriction site overhang (e.g., an EcoRI sticky-end overhang). The assay procedure can further comprise treating the target nucleic acids with a restriction enzyme (e.g., EcoRI) to create a restriction site overhang. The barcode can then be ligated to any nucleic acid molecule comprising a sequence complementary to the restriction site overhang. A ligase (e.g., T4 DNA ligase) can be used to join the two fragments.

[0472] For example, in a non-limiting example of barcoding illustrated in FIG. 2, at block 220, the labeled targets from a plurality of cells (or a plurality of samples) (e.g., target-barcode molecules) can be subsequently pooled, for example, into a tube. The labeled targets can be pooled by, for example, retrieving the barcodes and/or the beads to which the target-barcode molecules are attached.

[0473] The retrieval of solid support-based collections of attached target-barcode molecules can be implemented by use of magnetic beads and an externally-applied magnetic field. Once the target-barcode molecules have been pooled, all further processing can proceed in a single reaction vessel. Further processing can include, for example, reverse transcription reactions, amplification reactions, cleavage reactions, dissociation reactions, and/or nucleic acid extension reactions. Further processing reactions can be performed within the microwells, that is, without first pooling the labeled target nucleic acid molecules from a plurality of cells.

#### Reverse Transcription

[0474] The disclosure provides for a method to create a target-barcode conjugate using reverse transcription (e.g., at block 224 of FIG. 2). The target-barcode conjugate can comprise the barcode and a complementary sequence of all or a portion of the target nucleic acid (i.e., a barcoded cDNA molecule, such as a stochastically barcoded cDNA molecule). Reverse transcription of the associated RNA molecule can occur by the addition of a reverse transcription primer along with the reverse transcriptase. The reverse transcription primer can be an oligo(dT) primer, a random hexanucleotide primer, or a target-specific oligonucleotide primer. Oligo(dT) primers can be, or can be about, 12–18 nucleotides in

length and bind to the endogenous poly(A) tail at the 3' end of mammalian mRNA. Random hexanucleotide primers can bind to mRNA at a variety of complementary sites. Target-specific oligonucleotide primers typically selectively prime the mRNA of interest.

**[0475]** In some embodiments, reverse transcription of the labeled-RNA molecule can occur by the addition of a reverse transcription primer. In some embodiments, the reverse transcription primer is an oligo(dT) primer, random hexanucleotide primer, or a target-specific oligonucleotide primer. Generally, oligo(dT) primers are 12–18 nucleotides in length and bind to the endogenous poly(A) tail at the 3' end of mammalian mRNA. Random hexanucleotide primers can bind to mRNA at a variety of complementary sites. Target-specific oligonucleotide primers typically selectively prime the mRNA of interest.

**[0476]** Reverse transcription can occur repeatedly to produce multiple labeled-cDNA molecules. The methods disclosed herein can comprise conducting at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 reverse transcription reactions. The method can comprise conducting at least about 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 reverse transcription reactions.

#### Amplification

**[0477]** One or more nucleic acid amplification reactions (e.g., at block 228 of FIG. 2) can be performed to create multiple copies of the labeled target nucleic acid molecules. Amplification can be performed in a multiplexed manner, wherein multiple target nucleic acid sequences are amplified simultaneously. The amplification reaction can be used to add sequencing adaptors to the nucleic acid molecules. The amplification reactions can comprise amplifying at least a portion of a sample label, if present. The amplification reactions can comprise amplifying at least a portion of the cellular label and/or barcode sequence (e.g., a molecular label). The amplification reactions can comprise amplifying at least a portion of a sample tag, a cell label, a spatial label, a barcode sequence (e.g., a molecular label), a target nucleic acid, or a combination thereof. The amplification reactions can comprise amplifying 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 100%, or a range or a number between any two of these values, of the plurality of nucleic acids. The method can further comprise conducting one or more cDNA synthesis reactions to produce one or more cDNA copies of target-barcode molecules comprising a sample label,

a cell label, a spatial label, and/or a barcode sequence (e.g., a a molecular label).

[0478] In some embodiments, amplification can be performed using a polymerase chain reaction (PCR). As used herein, PCR can refer to a reaction for the in vitro amplification of specific DNA sequences by the simultaneous primer extension of complementary strands of DNA. As used herein, PCR can encompass derivative forms of the reaction, including but not limited to, RT-PCR, real-time PCR, nested PCR, quantitative PCR, multiplexed PCR, digital PCR, and assembly PCR.

[0479] Amplification of the labeled nucleic acids can comprise non-PCR based methods. Examples of non-PCR based methods include, but are not limited to, multiple displacement amplification (MDA), transcription-mediated amplification (TMA), nucleic acid sequence-based amplification (NASBA), strand displacement amplification (SDA), real-time SDA, rolling circle amplification, or circle-to-circle amplification. Other non-PCR-based amplification methods include multiple cycles of DNA-dependent RNA polymerase-driven RNA transcription amplification or RNA-directed DNA synthesis and transcription to amplify DNA or RNA targets, a ligase chain reaction (LCR), and a Q $\beta$  replicase (Q $\beta$ ) method, use of palindromic probes, strand displacement amplification, oligonucleotide-driven amplification using a restriction endonuclease, an amplification method in which a primer is hybridized to a nucleic acid sequence and the resulting duplex is cleaved prior to the extension reaction and amplification, strand displacement amplification using a nucleic acid polymerase lacking 5' exonuclease activity, rolling circle amplification, and ramification extension amplification (RAM). In some embodiments, the amplification does not produce circularized transcripts.

[0480] In some embodiments, the methods disclosed herein further comprise conducting a polymerase chain reaction on the labeled nucleic acid (e.g., labeled-RNA, labeled-DNA, labeled-cDNA) to produce a labeled amplicon (e.g., a stochastically labeledamplicon). The labeled amplicon can be double-stranded molecule. The double-stranded molecule can comprise a double-stranded RNA molecule, a double-stranded DNA molecule, or a RNA molecule hybridized to a DNA molecule. One or both of the strands of the double-stranded molecule can comprise a sample label, a spatial label, a cell label, and/or a barcode sequence (e.g., a molecular label). The labeled amplicon can be a single-stranded molecule. The single-stranded molecule can comprise DNA, RNA, or a combination thereof.

The nucleic acids of the disclosure can comprise synthetic or altered nucleic acids.

[0481] Amplification can comprise use of one or more non-natural nucleotides. Non-natural nucleotides can comprise photolabile or triggerable nucleotides. Examples of non-natural nucleotides can include, but are not limited to, peptide nucleic acid (PNA), morpholino and locked nucleic acid (LNA), as well as glycol nucleic acid (GNA) and threose nucleic acid (TNA). Non-natural nucleotides can be added to one or more cycles of an amplification reaction. The addition of the non-natural nucleotides can be used to identify products as specific cycles or time points in the amplification reaction.

[0482] Conducting the one or more amplification reactions can comprise the use of one or more primers. The one or more primers can comprise, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 or more nucleotides. The one or more primers can comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 or more nucleotides. The one or more primers can comprise less than 12-15 nucleotides. The one or more primers can anneal to at least a portion of the plurality of labeled targets (e.g., stochastically labeled targets). The one or more primers can anneal to the 3' end or 5' end of the plurality of labeled targets. The one or more primers can anneal to an internal region of the plurality of labeled targets. The internal region can be at least about 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 650, 700, 750, 800, 850, 900 or 1000 nucleotides from the 3' ends the plurality of labeled targets. The one or more primers can comprise a fixed panel of primers. The one or more primers can comprise at least one or more custom primers. The one or more primers can comprise at least one or more control primers. The one or more primers can comprise at least one or more gene-specific primers.

[0483] The one or more primers can comprise a universal primer. The universal primer can anneal to a universal primer binding site. The one or more custom primers can anneal to a first sample label, a second sample label, a spatial label, a cell label, a barcode sequence (e.g., a molecular label), a target, or any combination thereof. The one or more primers can comprise a universal primer and a custom primer. The custom primer can be designed to amplify one or more targets. The targets can comprise a subset of the total nucleic acids in one or more samples. The targets can comprise a subset of the total labeled

targets in one or more samples. The one or more primers can comprise at least 96 or more custom primers. The one or more primers can comprise at least 960 or more custom primers. The one or more primers can comprise at least 9600 or more custom primers. The one or more custom primers can anneal to two or more different labeled nucleic acids. The two or more different labeled nucleic acids can correspond to one or more genes.

[0484] Any amplification scheme can be used in the methods of the present disclosure. For example, in one scheme, the first round PCR can amplify molecules attached to the bead using a gene specific primer and a primer against the universal Illumina sequencing primer 1 sequence. The second round of PCR can amplify the first PCR products using a nested gene specific primer flanked by Illumina sequencing primer 2 sequence, and a primer against the universal Illumina sequencing primer 1 sequence. The third round of PCR adds P5 and P7 and sample index to turn PCR products into an Illumina sequencing library. Sequencing using 150 bp x 2 sequencing can reveal the cell label and barcode sequence (e.g., molecular label) on read 1, the gene on read 2, and the sample index on index 1 read.

[0485] In some embodiments, nucleic acids can be removed from the substrate using chemical cleavage. For example, a chemical group or a modified base present in a nucleic acid can be used to facilitate its removal from a solid support. For example, an enzyme can be used to remove a nucleic acid from a substrate. For example, a nucleic acid can be removed from a substrate through a restriction endonuclease digestion. For example, treatment of a nucleic acid containing a dUTP or ddUTP with uracil-d-glycosylase (UDG) can be used to remove a nucleic acid from a substrate. For example, a nucleic acid can be removed from a substrate using an enzyme that performs nucleotide excision, such as a base excision repair enzyme, such as an apurinic/apyrimidinic (AP) endonuclease. In some embodiments, a nucleic acid can be removed from a substrate using a photocleavable group and light. In some embodiments, a cleavable linker can be used to remove a nucleic acid from the substrate. For example, the cleavable linker can comprise at least one of biotin/avidin, biotin/streptavidin, biotin/neutravidin, Ig-protein A, a photo-labile linker, acid or base labile linker group, or an aptamer.

[0486] When the probes are gene-specific, the molecules can hybridize to the probes and be reverse transcribed and/or amplified. In some embodiments, after the nucleic acid has been synthesized (e.g., reverse transcribed), it can be amplified. Amplification can

be performed in a multiplex manner, wherein multiple target nucleic acid sequences are amplified simultaneously. Amplification can add sequencing adaptors to the nucleic acid.

[0487] In some embodiments, amplification can be performed on the substrate, for example, with bridge amplification. cDNAs can be homopolymer tailed in order to generate a compatible end for bridge amplification using oligo(dT) probes on the substrate. In bridge amplification, the primer that is complementary to the 3' end of the template nucleic acid can be the first primer of each pair that is covalently attached to the solid particle. When a sample containing the template nucleic acid is contacted with the particle and a single thermal cycle is performed, the template molecule can be annealed to the first primer and the first primer is elongated in the forward direction by addition of nucleotides to form a duplex molecule consisting of the template molecule and a newly formed DNA strand that is complementary to the template. In the heating step of the next cycle, the duplex molecule can be denatured, releasing the template molecule from the particle and leaving the complementary DNA strand attached to the particle through the first primer. In the annealing stage of the annealing and elongation step that follows, the complementary strand can hybridize to the second primer, which is complementary to a segment of the complementary strand at a location removed from the first primer. This hybridization can cause the complementary strand to form a bridge between the first and second primers secured to the first primer by a covalent bond and to the second primer by hybridization. In the elongation stage, the second primer can be elongated in the reverse direction by the addition of nucleotides in the same reaction mixture, thereby converting the bridge to a double-stranded bridge. The next cycle then begins, and the double-stranded bridge can be denatured to yield two single-stranded nucleic acid molecules, each having one end attached to the particle surface via the first and second primers, respectively, with the other end of each unattached. In the annealing and elongation step of this second cycle, each strand can hybridize to a further complementary primer, previously unused, on the same particle, to form new single-strand bridges. The two previously unused primers that are now hybridized elongate to convert the two new bridges to double-strand bridges.

[0488] The amplification reactions can comprise amplifying at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, or 100% of the plurality of nucleic acids.

[0489] Amplification of the labeled nucleic acids can comprise PCR-based methods or non-PCR based methods. Amplification of the labeled nucleic acids can comprise exponential amplification of the labeled nucleic acids. Amplification of the labeled nucleic acids can comprise linear amplification of the labeled nucleic acids. Amplification can be performed by polymerase chain reaction (PCR). PCR can refer to a reaction for the in vitro amplification of specific DNA sequences by the simultaneous primer extension of complementary strands of DNA. PCR can encompass derivative forms of the reaction, including but not limited to, RT-PCR, real-time PCR, nested PCR, quantitative PCR, multiplexed PCR, digital PCR, suppression PCR, semi-suppressive PCR and assembly PCR.

[0490] In some embodiments, amplification of the labeled nucleic acids comprises non-PCR based methods. Examples of non-PCR based methods include, but are not limited to, multiple displacement amplification (MDA), transcription-mediated amplification (TMA), nucleic acid sequence-based amplification (NASBA), strand displacement amplification (SDA), real-time SDA, rolling circle amplification, or circle-to-circle amplification. Other non-PCR-based amplification methods include multiple cycles of DNA-dependent RNA polymerase-driven RNA transcription amplification or RNA-directed DNA synthesis and transcription to amplify DNA or RNA targets, a ligase chain reaction (LCR), a Q $\beta$  replicase (Q $\beta$ ), use of palindromic probes, strand displacement amplification, oligonucleotide-driven amplification using a restriction endonuclease, an amplification method in which a primer is hybridized to a nucleic acid sequence and the resulting duplex is cleaved prior to the extension reaction and amplification, strand displacement amplification using a nucleic acid polymerase lacking 5' exonuclease activity, rolling circle amplification, and/or ramification extension amplification (RAM).

[0491] In some embodiments, the methods disclosed herein further comprise conducting a nested polymerase chain reaction on the amplified amplicon (e.g., target). The amplicon can be double-stranded molecule. The double-stranded molecule can comprise a double-stranded RNA molecule, a double-stranded DNA molecule, or a RNA molecule hybridized to a DNA molecule. One or both of the strands of the double-stranded molecule can comprise a sample tag or molecular identifier label. Alternatively, the amplicon can be a single-stranded molecule. The single-stranded molecule can comprise DNA, RNA, or a combination thereof. The nucleic acids of the present invention can comprise synthetic or



altered nucleic acids.

[0492] In some embodiments, the method comprises repeatedly amplifying the labeled nucleic acid to produce multiple amplicons. The methods disclosed herein can comprise conducting at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 amplification reactions. Alternatively, the method comprises conducting at least about 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amplification reactions.

[0493] Amplification can further comprise adding one or more control nucleic acids to one or more samples comprising a plurality of nucleic acids. Amplification can further comprise adding one or more control nucleic acids to a plurality of nucleic acids. The control nucleic acids can comprise a control label.

[0494] Amplification can comprise use of one or more non-natural nucleotides. Non-natural nucleotides can comprise photolabile and/or triggerable nucleotides. Examples of non-natural nucleotides include, but are not limited to, peptide nucleic acid (PNA), morpholino and locked nucleic acid (LNA), as well as glycol nucleic acid (GNA) and threose nucleic acid (TNA). Non-natural nucleotides can be added to one or more cycles of an amplification reaction. The addition of the non-natural nucleotides can be used to identify products as specific cycles or time points in the amplification reaction.

[0495] Conducting the one or more amplification reactions can comprise the use of one or more primers. The one or more primers can comprise one or more oligonucleotides. The one or more oligonucleotides can comprise at least about 7-9 nucleotides. The one or more oligonucleotides can comprise less than 12-15 nucleotides. The one or more primers can anneal to at least a portion of the plurality of labeled nucleic acids. The one or more primers can anneal to the 3' end and/or 5' end of the plurality of labeled nucleic acids. The one or more primers can anneal to an internal region of the plurality of labeled nucleic acids. The internal region can be at least about 50, 100, 150, 200, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 650, 700, 750, 800, 850, 900 or 1000 nucleotides from the 3' ends the plurality of labeled nucleic acids. The one or more primers can comprise a fixed panel of primers. The one or more primers can comprise at least one or more custom primers. The one or more primers can comprise at least one or more control primers. The one or more

primers can comprise at least one or more housekeeping gene primers. The one or more primers can comprise a universal primer. The universal primer can anneal to a universal primer binding site. The one or more custom primers can anneal to the first sample tag, the second sample tag, the molecular identifier label, the nucleic acid or a product thereof. The one or more primers can comprise a universal primer and a custom primer. The custom primer can be designed to amplify one or more target nucleic acids. The target nucleic acids can comprise a subset of the total nucleic acids in one or more samples. In some embodiments, the primers are the probes attached to the array of the disclosure.

[0496] In some embodiments, barcoding (e.g., stochastically barcoding) the plurality of targets in the sample further comprises generating an indexed library of the barcoded targets (e.g., stochastically barcoded targets) or barcoded fragments of the targets. The barcode sequences of different barcodes (e.g., the molecular labels of different stochastic barcodes) can be different from one another. Generating an indexed library of the barcoded targets includes generating a plurality of indexed polynucleotides from the plurality of targets in the sample. For example, for an indexed library of the barcoded targets comprising a first indexed target and a second indexed target, the label region of the first indexed polynucleotide can differ from the label region of the second indexed polynucleotide by, by about, by at least, or by at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, or a number or a range between any two of these values, nucleotides. In some embodiments, generating an indexed library of the barcoded targets includes contacting a plurality of targets, for example mRNA molecules, with a plurality of oligonucleotides including a poly(T) region and a label region; and conducting a first strand synthesis using a reverse transcriptase to produce single-strand labeled cDNA molecules each comprising a cDNA region and a label region, wherein the plurality of targets includes at least two mRNA molecules of different sequences and the plurality of oligonucleotides includes at least two oligonucleotides of different sequences. Generating an indexed library of the barcoded targets can further comprise amplifying the single-strand labeled cDNA molecules to produce double-strand labeled cDNA molecules; and conducting nested PCR on the double-strand labeled cDNA molecules to produce labeled amplicons. In some embodiments, the method can include generating an adaptor-labeled amplicon.

[0497] Barcoding (e.g., stochastic barcoding) can include using nucleic acid

barcodes or tags to label individual nucleic acid (e.g., DNA or RNA) molecules. In some embodiments, it involves adding DNA barcodes or tags to cDNA molecules as they are generated from mRNA. Nested PCR can be performed to minimize PCR amplification bias. Adaptors can be added for sequencing using, for example, next generation sequencing (NGS). The sequencing results can be used to determine cell labels, molecular labels, and sequences of nucleotide fragments of the one or more copies of the targets, for example at block 232 of FIG. 2.

[0498] FIG. 3 is a schematic illustration showing a non-limiting exemplary process of generating an indexed library of the barcoded targets (e.g., stochastically barcoded targets), such as barcoded mRNAs or fragments thereof. As shown in step 1, the reverse transcription process can encode each mRNA molecule with a unique molecular label, a cell label, and a universal PCR site. In particular, RNA molecules 302 can be reverse transcribed to produce labeled cDNA molecules 304, including a cDNA region 306, by hybridization (e.g., stochastic hybridization) of a set of barcodes (e.g., stochastic barcodes) 310 to the poly(A) tail region 308 of the RNA molecules 302. Each of the barcodes 310 can comprise a target-binding region, for example a poly(dT) region 312, a label region 314 (e.g., a barcode sequence or a molecule), and a universal PCR region 316.

[0499] In some embodiments, the cell label can include 3 to 20 nucleotides. In some embodiments, the molecular label can include 3 to 20 nucleotides. In some embodiments, each of the plurality of stochastic barcodes further comprises one or more of a universal label and a cell label, wherein universal labels are the same for the plurality of stochastic barcodes on the solid support and cell labels are the same for the plurality of stochastic barcodes on the solid support. In some embodiments, the universal label can include 3 to 20 nucleotides. In some embodiments, the cell label comprises 3 to 20 nucleotides.

[0500] In some embodiments, the label region 314 can include a barcode sequence or a molecular label 318 and a cell label 320. In some embodiments, the label region 314 can include one or more of a universal label, a dimension label, and a cell label. The barcode sequence or molecular label 318 can be, can be about, can be at least, or can be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or a number or a range between any of these values, of nucleotides in length. The cell label 320 can be, can be

about, can be at least, or can be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or a number or a range between any of these values, of nucleotides in length. The universal label can be, can be about, can be at least, or can be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or a number or a range between any of these values, of nucleotides in length. Universal labels can be the same for the plurality of stochastic barcodes on the solid support and cell labels are the same for the plurality of stochastic barcodes on the solid support. The dimension label can be, can be about, can be at least, or can be at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or a number or a range between any of these values, of nucleotides in length.

**[0501]** In some embodiments, the label region 314 can comprise, comprise about, comprise at least, or comprise at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any of these values, different labels, such as a barcode sequence or a molecular label 318 and a cell label 320. Each label can be, can be about, can be at least, or can be at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or a number or a range between any of these values, of nucleotides in length. A set of barcodes or stochastic barcodes 310 can contain, contain about, contain at least, or can be at most, 10, 20, 40, 50, 70, 80, 90,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ ,  $10^{11}$ ,  $10^{12}$ ,  $10^{13}$ ,  $10^{14}$ ,  $10^{15}$ ,  $10^{20}$ , or a number or a range between any of these values, barcodes or stochastic barcodes 310. And the set of barcodes or stochastic barcodes 310 can, for example, each contain a unique label region 314. The labeled cDNA molecules 304 can be purified to remove excess barcodes or stochastic barcodes 310. Purification can comprise Ampure bead purification.

**[0502]** As shown in step 2, products from the reverse transcription process in step 1 can be pooled into 1 tube and PCR amplified with a 1<sup>st</sup> PCR primer pool and a 1<sup>st</sup> universal PCR primer. Pooling is possible because of the unique label region 314. In particular, the labeled cDNA molecules 304 can be amplified to produce nested PCR labeled amplicons 322. Amplification can comprise multiplex PCR amplification. Amplification can comprise a multiplex PCR amplification with 96 multiplex primers in a single reaction volume. In some embodiments, multiplex PCR amplification can utilize, utilize about, utilize at least, or utilize at most, 10, 20, 40, 50, 70, 80, 90,  $10^2$ ,  $10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ ,  $10^{10}$ ,  $10^{11}$ ,  $10^{12}$ ,  $10^{13}$ ,  $10^{14}$ ,  $10^{15}$ ,  $10^{20}$ , or a number or a range between any of these values, multiplex

primers in a single reaction volume. Amplification can comprise using a 1<sup>st</sup> PCR primer pool 324 comprising custom primers 326A-C targeting specific genes and a universal primer 328. The custom primers 326 can hybridize to a region within the cDNA portion 306' of the labeled cDNA molecule 304. The universal primer 328 can hybridize to the universal PCR region 316 of the labeled cDNA molecule 304.

[0503] As shown in step 3 of FIG. 3, products from PCR amplification in step 2 can be amplified with a nested PCR primers pool and a 2<sup>nd</sup> universal PCR primer. Nested PCR can minimize PCR amplification bias. In particular, the nested PCR labeled amplicons 322 can be further amplified by nested PCR. The nested PCR can comprise multiplex PCR with nested PCR primers pool 330 of nested PCR primers 332a-c and a 2<sup>nd</sup> universal PCR primer 328' in a single reaction volume. The nested PCR primer pool 328 can contain, contain about, contain at least, or contain at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any of these values, different nested PCR primers 330. The nested PCR primers 332 can contain an adaptor 334 and hybridize to a region within the cDNA portion 306'' of the labeled amplicon 322. The universal primer 328' can contain an adaptor 336 and hybridize to the universal PCR region 316 of the labeled amplicon 322. Thus, step 3 produces adaptor-labeled amplicon 338. In some embodiments, nested PCR primers 332 and the 2<sup>nd</sup> universal PCR primer 328' may not contain the adaptors 334 and 336. The adaptors 334 and 336 can instead be ligated to the products of nested PCR to produce adaptor-labeled amplicon 338.

[0504] As shown in step 4, PCR products from step 3 can be PCR amplified for sequencing using library amplification primers. In particular, the adaptors 334 and 336 can be used to conduct one or more additional assays on the adaptor-labeled amplicon 338. The adaptors 334 and 336 can be hybridized to primers 340 and 342. The one or more primers 340 and 342 can be PCR amplification primers. The one or more primers 340 and 342 can be sequencing primers. The one or more adaptors 334 and 336 can be used for further amplification of the adaptor-labeled amplicons 338. The one or more adaptors 334 and 336 can be used for sequencing the adaptor-labeled amplicon 338. The primer 342 can contain a plate index 344 so that amplicons generated using the same set of barcodes or stochastic barcodes 310 can be sequenced in one sequencing reaction using next generation sequencing (NGS).

Compositions Comprising Cellular Component Binding Reagents Associated with Oligonucleotides

[0505] Some embodiments disclosed herein provide a plurality of compositions each comprising a cellular component binding reagent (such as a protein binding reagent) that is conjugated with an oligonucleotide, wherein the oligonucleotide comprises a unique identifier for the cellular component binding reagent that it is conjugated with. In some embodiments, the cellular component binding reagent is capable of specifically binding to a cellular component target. For example, a binding target of the cellular component binding reagent can be, or comprise, a carbohydrate, a lipid, a protein, an extracellular protein, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, an intracellular protein, or any combination thereof. In some embodiments, the cellular component binding reagent (e.g., a protein binding reagent) is capable of specifically binding to an antigen target or a protein target. In some embodiments, each of the oligonucleotides can comprise a barcode, such as a stochastic barcode. A barcode can comprise a barcode sequence (e.g., a molecular label), a cell label, a sample label, or any combination thereof. In some embodiments, each of the oligonucleotides can comprise a linker. In some embodiments, each of the oligonucleotides can comprise a binding site for an oligonucleotide probe, such as a poly(A) tail. For example, the poly(A) tail can be, e.g., unanchored to a solid support or anchored to a solid support. The poly(A) tail can be from about 10 to 50 nucleotides in length. In some embodiments, the poly(A) tail can be 18 nucleotides in length. The oligonucleotides can comprise deoxyribonucleotides, ribonucleotides, or both.

[0506] The unique identifiers can be, for example, a nucleotide sequence having any suitable length, for example, from about 4 nucleotides to about 200 nucleotides. In some embodiments, the unique identifier is a nucleotide sequence of 25 nucleotides to about 45 nucleotides in length. In some embodiments, the unique identifier can have a length that is, is about, is less than, is greater than, 4 nucleotides, 5 nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, 10 nucleotides, 15 nucleotides, 20 nucleotides, 25 nucleotides, 30 nucleotides, 35 nucleotides, 40 nucleotides, 45 nucleotides, 50 nucleotides, 55 nucleotides, 60 nucleotides, 70 nucleotides, 80 nucleotides, 90 nucleotides, 100

nucleotides, 200 nucleotides, or a range that is between any two of the above values.

[0507] In some embodiments, the unique identifiers are selected from a diverse set of unique identifiers. The diverse set of unique identifiers can comprise, or comprise about, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 5000, or a number or a range between any two of these values, different unique identifiers. The diverse set of unique identifiers can comprise at least, or comprise at most, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, or 5000, different unique identifiers. In some embodiments, the set of unique identifiers is designed to have minimal sequence homology to the DNA or RNA sequences of the sample to be analyzed. In some embodiments, the sequences of the set of unique identifiers are different from each other, or the complement thereof, by, or by about, 1 nucleotide, 2 nucleotides, 3 nucleotides, 4 nucleotides, 5 nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, 10 nucleotides, or a number or a range between any two of these values. In some embodiments, the sequences of the set of unique identifiers are different from each other, or the complement thereof, by at least, or by at most, 1 nucleotide, 2 nucleotides, 3 nucleotides, 4 nucleotides, 5 nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, 10 nucleotides. In some embodiments, the sequences of the set of unique identifiers are different from each other, or the complement thereof, by at least 3%, at least 5%, at least 8%, at least 10%, at least 15%, at least 20%, or more.

[0508] In some embodiments, the unique identifiers can comprise a binding site for a primer, such as universal primer. In some embodiments, the unique identifiers can comprise at least two binding sites for a primer, such as a universal primer. In some embodiments, the unique identifiers can comprise at least three binding sites for a primer, such as a universal primer. The primers can be used for amplification of the unique identifiers, for example, by PCR amplification. In some embodiments, the primers can be used for nested PCR reactions.

[0509] Any suitable cellular component binding reagents are contemplated in this disclosure, such as protein binding reagents, antibodies or fragments thereof, aptamers, small molecules, ligands, peptides, oligonucleotides, etc., or any combination thereof. In some embodiments, the cellular component binding reagents can be polyclonal antibodies, monoclonal antibodies, recombinant antibodies, single chain antibody (sc-Ab), or fragments

thereof, such as Fab, Fv, etc. In some embodiments, the plurality of cellular component binding reagents can comprise, or comprise about, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 5000, or a number or a range between any two of these values, different cellular component reagents. In some embodiments, the plurality of cellular component binding reagents can comprise at least, or comprise at most, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, or 5000, different cellular component reagents.

**[0510]** The oligonucleotide can be conjugated with the cellular component binding reagent through various mechanism. In some embodiments, the oligonucleotide can be conjugated with the cellular component binding reagent covalently. In some embodiment, the oligonucleotide can be conjugated with the cellular component binding reagent non-covalently. In some embodiments, the oligonucleotide is conjugated with the cellular component binding reagent through a linker. The linker can be, for example, cleavable or detachable from the cellular component binding reagent and/or the oligonucleotide. In some embodiments, the linker can comprise a chemical group that reversibly attaches the oligonucleotide to the cellular component binding reagents. The chemical group can be conjugated to the linker, for example, through an amine group. In some embodiments, the linker can comprise a chemical group that forms a stable bond with another chemical group conjugated to the cellular component binding reagent. For example, the chemical group can be a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, etc. In some embodiments, the chemical group can be conjugated to the cellular component binding reagent through a primary amine on an amino acid, such as lysine, or the N-terminus. Commercially available conjugation kits, such as the Protein-Oligo Conjugation Kit (Solulink, Inc., San Diego, California), the Thunder-Link® oligo conjugation system (Innova Biosciences, Cambridge, United Kingdom), etc., can be used to conjugate the oligonucleotide to the cellular component binding reagent.

**[0511]** The oligonucleotide can be conjugated to any suitable site of the cellular component binding reagent (e.g., a protein binding reagent), as long as it does not interfere with the specific binding between the cellular component binding reagent and its cellular component target. In some embodiments, the cellular component binding reagent is a protein, such as an antibody. In some embodiments, the cellular component binding reagent



is not an antibody. In some embodiments, the oligonucleotide can be conjugated to the antibody anywhere other than the antigen-binding site, for example, the Fc region, the C<sub>H</sub>1 domain, the C<sub>H</sub>2 domain, the C<sub>H</sub>3 domain, the C<sub>L</sub> domain, etc. Methods of conjugating oligonucleotides to cellular component binding reagents (e.g., antibodies) have been previously disclosed, for example, in U.S. Patent. No. 6,531,283, the content of which is hereby expressly incorporated by reference in its entirety. Stoichiometry of oligonucleotide to cellular component binding reagent can be varied. To increase the sensitivity of detecting the cellular component binding reagent specific oligonucleotide in sequencing, it may be advantageous to increase the ratio of oligonucleotide to cellular component binding reagent during conjugation. In some embodiments, each cellular component binding reagent can be conjugated with a single oligonucleotide molecule. In some embodiments, each cellular component binding reagent can be conjugated with more than one oligonucleotide molecule, for example, at least, or at most, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, or a number or a range between any two of these values, oligonucleotide molecules wherein each of the oligonucleotide molecule comprises the same, or different, unique identifiers. In some embodiments, each cellular component binding reagent can be conjugated with more than one oligonucleotide molecule, for example, at least, or at most, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, oligonucleotide molecules, wherein each of the oligonucleotide molecule comprises the same, or different, unique identifiers.

[0512] In some embodiments, the plurality of cellular component binding reagents are capable of specifically binding to a plurality of cellular component targets in a sample, such as a single cell, a plurality of cells, a tissue sample, a tumor sample, a blood sample, or the like. In some embodiments, the plurality of cellular component targets comprises a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, an antibody, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. In some embodiments, the plurality of cellular component targets can comprise intracellular cellular components. In some embodiments, the plurality of cellular component targets can comprise intracellular cellular components. In some embodiments, the plurality of cellular components can be, or be about, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or a number or a range between any two of these values, of all the cellular components (e.g., proteins) in a cell

or an organism. In some embodiments, the plurality of cellular components can be at least, or be at most, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99%, of all the cellular components (e.g., proteins) in a cell or an organism. In some embodiments, the plurality of cellular component targets can comprise, or comprise about, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, 10000, or a number or a range between any two of these values, different cellular component targets. In some embodiments, the plurality of cellular component targets can comprise at least, or comprise at most, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, 10000, different cellular component targets.

**[0513]** FIG. 4 shows a schematic illustration of an exemplary cellular component binding reagent, e.g., an antibody, that is conjugated with an oligonucleotide comprising a unique identifier sequence for the antibody. An oligonucleotide-conjugated with a cellular component binding reagent, an oligonucleotide for conjugation with a cellular component binding reagent, or an oligonucleotide previously conjugated with a cellular component binding reagent can be referred to herein as an antibody oligonucleotide (abbreviated as a binding reagent oligonucleotide). An oligonucleotide-conjugated with an antibody, an oligonucleotide for conjugation with an antibody, or an oligonucleotide previously conjugated with an antibody can be referred to herein as an antibody oligonucleotide (abbreviated as an “AbOligo” or “AbO”). The oligonucleotide can also comprise additional components, including but not limited to, one or more linker, one or more unique identifier for the antibody, optionally one or more barcode sequences (e.g., molecular labels), and a poly(A) tail. In some embodiments, the oligonucleotide can comprise, from 5' to 3', a linker, a unique identifier, a barcode sequence (e.g., a molecular label), and a poly(A) tail. An antibody oligonucleotide can be an mRNA mimic.

**[0514]** FIG. 5 shows a schematic illustration of an exemplary cellular component binding reagent, e.g., an antibody, that is conjugated with an oligonucleotide comprising a unique identifier sequence for the antibody. The cellular component binding reagent can be capable of specifically binding to at least one cellular component target, such as an antigen target or a protein target. A binding reagent oligonucleotide (e.g., a sample indexing oligonucleotide, or an antibody oligonucleotide) can comprise a sequence (e.g., a sample indexing sequence) for performing the methods of the disclosure. For example, a sample indexing oligonucleotide can comprise a sample indexing sequence for identifying sample

origin of one or more cells of a sample. Indexing sequences (e.g., sample indexing sequences) of at least two compositions comprising two cellular component binding reagents (e.g., sample indexing compositions) of the plurality of compositions comprising cellular component binding reagents can comprise different sequences. In some embodiments, the binding reagent oligonucleotide is not homologous to genomic sequences of a species. The binding reagent oligonucleotide can be configured to be detachable or non-detachable from the cellular component binding reagent.

**[0515]** The oligonucleotide conjugated to a cellular component binding reagent can, for example, comprise a barcode sequence (e.g., a molecular label sequence), a poly(A) tail, or a combination thereof. An oligonucleotide conjugated to a cellular component binding reagent can be an mRNA mimic. In some embodiments, the sample indexing oligonucleotide comprises a sequence complementary to a capture sequence of at least one barcode of the plurality of barcodes. A target binding region of the barcode can comprise the capture sequence. The target binding region can, for example, comprise a poly(dT) region. In some embodiments, the sequence of the sample indexing oligonucleotide complementary to the capture sequence of the barcode can comprise a poly(A) tail. The sample indexing oligonucleotide can comprise a molecular label.

**[0516]** In some embodiments, the binding reagent oligonucleotide (e.g., the sample oligonucleotide) comprises a nucleotide sequence of, or a nucleotide sequence of about, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 128, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, or a number or a range between any two of these values, nucleotides in length. In some embodiments, the binding reagent oligonucleotide comprises a nucleotide sequence of at least, or of at most, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 128, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590,

600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, or 1000, nucleotides in length.

**[0517]** In some embodiments, the cellular component binding reagent comprises an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The binding reagent oligonucleotide can be conjugated to the cellular component binding reagent, for example, through a linker. The binding reagent oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly, or irreversibly, attached to the molecule of the cellular component binding reagent. The chemical group can be selected from the group consisting of a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, and any combination thereof.

**[0518]** In some embodiments, the cellular component binding reagent can bind to ADAM10, CD156c, ANO6, ATP1B2, ATP1B3, BSG, CD147, CD109, CD230, CD29, CD298, ATP1B3, CD44, CD45, CD47, CD51, CD59, CD63, CD97, CD98, SLC3A2, CLDND1, HLA-ABC, ICAM1, ITFG3, MPZL1, NA K ATPase alpha1, ATP1A1, NPTN, PMCA ATPase, ATP2B1, SLC1A5, SLC29A1, SLC2A1, SLC44A2, or any combination thereof.

**[0519]** In some embodiments, the protein target is, or comprises, an extracellular protein, an intracellular protein, or any combination thereof. In some embodiments, the antigen or protein target is, or comprises, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, or any combination thereof. The antigen or protein target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The protein target can be selected from a group comprising a number of protein targets. The number of antigen taragett or protein targets can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or a number or a range between any two of these values. The number of protein targets can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000.

**[0520]** The cellular component binding reagent (e.g., a protein binding reagent)

can be associated with two or more binding reagent oligonucleotide (e.g., sample indexing oligonucleotides) with an identical sequence. The cellular component binding reagent can be associated with two or more binding reagent oligonucleotides with different sequences. The number of binding reagent oligonucleotides associated with the cellular component binding reagent can be different in different implementations. In some embodiments, the number of binding reagent oligonucleotides, whether having an identical sequence, or different sequences, can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of binding reagent oligonucleotides can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000.

**[0521]** The plurality of compositions comprising cellular component binding reagents (e.g., the plurality of sample indexing compositions) can comprise one or more additional cellular component binding reagents not conjugated with the binding reagent oligonucleotide (such as sample indexing oligonucleotide), which is also referred to herein as the binding reagent oligonucleotide-free cellular component binding reagent (such as sample indexing oligonucleotide-free cellular component binding reagent). The number of additional cellular component binding reagents in the plurality of compositions can be different in different implementations. In some embodiments, the number of additional cellular component binding reagents can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or a number or a range between any two of these values. In some embodiments, the number of additional cellular component binding reagents can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100. The cellular component binding reagent and any of the additional cellular component binding reagents can be identical, in some embodiments.

**[0522]** In some embodiments, a mixture comprising cellular component binding reagent(s) that is conjugated with one or more binding reagent oligonucleotides (e.g., sample indexing oligonucleotides) and cellular component binding reagent(s) that is not conjugated with binding reagent oligonucleotides is provided. The mixture can be used in some embodiments of the methods disclosed herein, for example, to contact the sample(s) and/or cell(s). The ratio of (1) the number of a cellular component binding reagent conjugated with

a binding reagent oligonucleotide and (2) the number of another cellular component binding reagent (e.g., the same cellular component binding reagent) not conjugated with the binding reagent oligonucleotide (e.g., sample indexing oligonucleotide) or other binding reagent oligonucleotide(s) in the mixture can be different in different implementations. In some embodiments, the ratio can be, or be about, 1:1, 1:1.1, 1:1.2, 1:1.3, 1:1.4, 1:1.5, 1:1.6, 1:1.7, 1:1.8, 1:1.9, 1:2, 1:2.5, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:21, 1:22, 1:23, 1:24, 1:25, 1:26, 1:27, 1:28, 1:29, 1:30, 1:31, 1:32, 1:33, 1:34, 1:35, 1:36, 1:37, 1:38, 1:39, 1:40, 1:41, 1:42, 1:43, 1:44, 1:45, 1:46, 1:47, 1:48, 1:49, 1:50, 1:51, 1:52, 1:53, 1:54, 1:55, 1:56, 1:57, 1:58, 1:59, 1:60, 1:61, 1:62, 1:63, 1:64, 1:65, 1:66, 1:67, 1:68, 1:69, 1:70, 1:71, 1:72, 1:73, 1:74, 1:75, 1:76, 1:77, 1:78, 1:79, 1:80, 1:81, 1:82, 1:83, 1:84, 1:85, 1:86, 1:87, 1:88, 1:89, 1:90, 1:91, 1:92, 1:93, 1:94, 1:95, 1:96, 1:97, 1:98, 1:99, 1:100, 1:200, 1:300, 1:400, 1:500, 1:600, 1:700, 1:800, 1:900, 1:1000, 1:2000, 1:3000, 1:4000, 1:5000, 1:6000, 1:7000, 1:8000, 1:9000, 1:10000, or a number or a range between any two of the values. In some embodiments, the ratio can be at least, or be at most, 1:1, 1:1.1, 1:1.2, 1:1.3, 1:1.4, 1:1.5, 1:1.6, 1:1.7, 1:1.8, 1:1.9, 1:2, 1:2.5, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:21, 1:22, 1:23, 1:24, 1:25, 1:26, 1:27, 1:28, 1:29, 1:30, 1:31, 1:32, 1:33, 1:34, 1:35, 1:36, 1:37, 1:38, 1:39, 1:40, 1:41, 1:42, 1:43, 1:44, 1:45, 1:46, 1:47, 1:48, 1:49, 1:50, 1:51, 1:52, 1:53, 1:54, 1:55, 1:56, 1:57, 1:58, 1:59, 1:60, 1:61, 1:62, 1:63, 1:64, 1:65, 1:66, 1:67, 1:68, 1:69, 1:70, 1:71, 1:72, 1:73, 1:74, 1:75, 1:76, 1:77, 1:78, 1:79, 1:80, 1:81, 1:82, 1:83, 1:84, 1:85, 1:86, 1:87, 1:88, 1:89, 1:90, 1:91, 1:92, 1:93, 1:94, 1:95, 1:96, 1:97, 1:98, 1:99, 1:100, 1:200, 1:300, 1:400, 1:500, 1:600, 1:700, 1:800, 1:900, 1:1000, 1:2000, 1:3000, 1:4000, 1:5000, 1:6000, 1:7000, 1:8000, 1:9000, or 1:10000.

**[0523]** In some embodiments, the ratio can be, or be about, 1:1, 1.1:1, 1.2:1, 1.3:1, 1.4:1, 1.5:1, 1.6:1, 1.7:1, 1.8:1, 1.9:1, 2:1, 2.5:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 21:1, 22:1, 23:1, 24:1, 25:1, 26:1, 27:1, 28:1, 29:1, 30:1, 31:1, 32:1, 33:1, 34:1, 35:1, 36:1, 37:1, 38:1, 39:1, 40:1, 41:1, 42:1, 43:1, 44:1, 45:1, 46:1, 47:1, 48:1, 49:1, 50:1, 51:1, 52:1, 53:1, 54:1, 55:1, 56:1, 57:1, 58:1, 59:1, 60:1, 61:1, 62:1, 63:1, 64:1, 65:1, 66:1, 67:1, 68:1, 69:1, 70:1, 71:1, 72:1, 73:1, 74:1, 75:1, 76:1, 77:1, 78:1, 79:1, 80:1, 81:1, 82:1, 83:1, 84:1, 85:1, 86:1, 87:1, 88:1, 89:1, 90:1, 91:1, 92:1, 93:1, 94:1, 95:1, 96:1, 97:1, 98:1, 99:1, 100:1, 200:1, 300:1, 400:1, 500:1, 600:1,

700:1, 800:1, 900:1, 1000:1, 2000:1, 3000:1, 4000:1, 5000:1, 6000:1, 7000:1, 8000:1, 9000:1, 10000:1, or a number or a range between any two of the values. In some embodiments, the ratio can be at least, or be at most, 1:1, 1.1:1, 1.2:1, 1.3:1, 1.4:1, 1.5:1, 1.6:1, 1.7:1, 1.8:1, 1.9:1, 2:1, 2.5:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1, 21:1, 22:1, 23:1, 24:1, 25:1, 26:1, 27:1, 28:1, 29:1, 30:1, 31:1, 32:1, 33:1, 34:1, 35:1, 36:1, 37:1, 38:1, 39:1, 40:1, 41:1, 42:1, 43:1, 44:1, 45:1, 46:1, 47:1, 48:1, 49:1, 50:1, 51:1, 52:1, 53:1, 54:1, 55:1, 56:1, 57:1, 58:1, 59:1, 60:1, 61:1, 62:1, 63:1, 64:1, 65:1, 66:1, 67:1, 68:1, 69:1, 70:1, 71:1, 72:1, 73:1, 74:1, 75:1, 76:1, 77:1, 78:1, 79:1, 80:1, 81:1, 82:1, 83:1, 84:1, 85:1, 86:1, 87:1, 88:1, 89:1, 90:1, 91:1, 92:1, 93:1, 94:1, 95:1, 96:1, 97:1, 98:1, 99:1, 100:1, 200:1, 300:1, 400:1, 500:1, 600:1, 700:1, 800:1, 900:1, 1000:1, 2000:1, 3000:1, 4000:1, 5000:1, 6000:1, 7000:1, 8000:1, 9000:1, or 10000:1.

[0524] A cellular component binding reagent can be conjugated with a binding reagent oligonucleotide (e.g., a sample indexing oligonucleotide), or not. In some embodiments, the percentage of the cellular component binding reagent conjugated with a binding reagent oligonucleotide (e.g., a sample indexing oligonucleotide) in a mixture comprising the cellular component binding reagent that is conjugated with the binding reagent oligonucleotide and the cellular component binding reagent(s) that is not conjugated with the binding reagent oligonucleotide can be, or be about, 0.000000001%, 0.00000001%, 0.0000001%, 0.000001%, 0.00001%, 0.0001%, 0.001%, 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of the cellular component binding reagent conjugated with a sample indexing oligonucleotide in a mixture can be at least, or be at most, 0.0000000001%, 0.000000001%, 0.00000001%, 0.0000001%, 0.000001%, 0.0001%, 0.001%, 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%,

41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%.

[0525] In some embodiments, the percentage of the cellular component binding reagent not conjugated with a binding reagent oligonucleotide (e.g., a sample indexing oligonucleotide) in a mixture comprising a cellular component binding reagent conjugated with a binding reagent oligonucleotide (e.g., a sample indexing oligonucleotide) and the cellular component binding reagent that is not conjugated with the sample indexing oligonucleotide can be, or be about, 0.000000001%, 0.00000001%, 0.0000001%, 0.000001%, 0.00001%, 0.0001%, 0.001%, 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of the cellular component binding reagent not conjugated with a binding reagent oligonucleotide in a mixture can be at least, or be at most, 0.0000000001%, 0.000000001%, 0.00000001%, 0.0000001%, 0.000001%, 0.0001%, 0.001%, 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%.

#### Cellular Component Cocktails

[0526] In some embodiments, a cocktail of cellular component binding reagents (e.g., an antibody cocktail) can be used to increase labeling sensitivity in the methods



disclosed herein. Without being bound by any particular theory, it is believed that this may be because cellular component expression or protein expression can vary between cell types and cell states, making finding a universal cellular component binding reagent or antibody that labels all cell types challenging. For example, cocktail of cellular component binding reagents can be used to allow for more sensitive and efficient labeling of more sample types. The cocktail of cellular component binding reagents can include two or more different types of cellular component binding reagents, for example a wider range of cellular component binding reagents or antibodies. Cellular component binding reagents that label different cellular component targets can be pooled together to create a cocktail that sufficiently labels all cell types, or one or more cell types of interest.

**[0527]** In some embodiments, each of the plurality of compositions (e.g., sample indexing compositions) comprises a cellular component binding reagent. In some embodiments, a composition of the plurality of compositions comprises two or more cellular component binding reagents, wherein each of the two or more cellular component binding reagents is associated with a binding reagent oligonucleotide (e.g., a sample indexing oligonucleotide), wherein at least one of the two or more cellular component binding reagents is capable of specifically binding to at least one of the one or more cellular component targets. The sequences of the binding reagent oligonucleotides associated with the two or more cellular component binding reagents can be identical. The sequences of the binding reagent oligonucleotides associated with the two or more cellular component binding reagents can comprise different sequences. Each of the plurality of compositions can comprise the two or more cellular component binding reagents.

**[0528]** The number of different types of cellular component binding reagents (e.g., a CD147 antibody and a CD47 antibody) in a composition can be different in different implementations. A composition with two or more different types of cellular component binding reagents can be referred to herein as a cellular component binding reagent cocktail (e.g., a sample indexing composition cocktail). The number of different types of cellular component binding reagents in a cocktail can vary. In some embodiments, the number of different types of cellular component binding reagents in cocktail can be, or be about 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, or a number or a range between any two of these values. In some

embodiments, the number of different types of cellular component binding reagents in cocktail can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, or 100000. The different types of cellular component binding reagents can be conjugated to binding reagent oligonucleotides with the same or different sequences (e.g., sample indexing sequences).

#### Methods of Quantitative Analysis of Cellular Component Targets

[0529] In some embodiments, the methods disclosed herein can also be used for quantitative analysis of a plurality of cellular component targets (for example, protein targets) in a sample using the compositions disclosed herein and oligonucleotide probes that can associate a barcode sequence (e.g., a molecular label sequence) to the oligonucleotides of the cellular component binding reagents (e.g., protein binding reagents). The oligonucleotides of the cellular component binding reagents can be, or comprise, an antibody oligonucleotide, a sample indexing oligonucleotide, a cell identification oligonucleotide, a control particle oligonucleotide, a control oligonucleotide, an interaction determination oligonucleotide, etc. In some embodiments, the sample can be a single cell, a plurality of cells, a tissue sample, a tumor sample, a blood sample, or the like. In some embodiments, the sample can comprise a mixture of cell types, such as normal cells, tumor cells, blood cells, B cells, T cells, maternal cells, fetal cells, etc., or a mixture of cells from different subjects.

[0530] In some embodiments, the sample can comprise a plurality of single cells separated into individual compartments, such as microwells in a microwell array.

[0531] In some embodiments, the binding target of the plurality of cellular component target (i.e., the cellular component target) can be, or comprise, a carbohydrate, a lipid, a protein, an extracellular protein, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an integrin, an intracellular protein, or any combination thereof. In some embodiments, the cellular component target is a protein target. In some embodiments, the plurality of cellular component targets comprises a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, an antibody, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. In some embodiments, the plurality of cellular component targets can comprise intracellular cellular components. In some embodiments, the plurality of cellular

components can be at least 1%, at least 2%, at least 3%, at least 4%, at least 5%, at least 6%, at least 7%, at least 8%, at least 9%, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 98%, at least 99%, or more, of all the encoded cellular components in an organism. In some embodiments, the plurality of cellular component targets can comprise at least 2, at least 3, at least 4, at least 5, at least 10, at least 20, at least 30, at least 40, at least 50, at least 100, at least 1000, at least 10000, or more different cellular component targets.

[0532] In some embodiments, the plurality of cellular component binding reagents is contacted with the sample for specific binding with the plurality of cellular component targets. Unbound cellular component binding reagents can be removed, for example, by washing. In embodiments where the sample comprises cells, any cellular component binding reagents not specifically bound to the cells can be removed.

[0533] In some instances, cells from a population of cells can be separated (e.g., isolated) into wells of a substrate of the disclosure. The population of cells can be diluted prior to separating. The population of cells can be diluted such that at least 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% or 100%, of wells of the substrate receive a single cell. The population of cells can be diluted such that at most 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%, of wells of the substrate receive a single cell. The population of cells can be diluted such that the number of cells in the diluted population is, or is at least, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%, of the number of wells on the substrate. The population of cells can be diluted such that the number of cells in the diluted population is, or is at least, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%, of the number of wells on the substrate. In some instances, the population of cells is diluted such that the number of cell is about 10% of the number of wells in the substrate.

[0534] Distribution of single cells into wells of the substrate can follow a Poisson distribution. For example, there can be at least a 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, or 10%, or more probability that a well of the substrate has more than one cell. There can be at least a 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, or 10%, or more

probability that a well of the substrate has more than one cell. Distribution of single cells into wells of the substrate can be random. Distribution of single cells into wells of the substrate can be non-random. The cells can be separated such that a well of the substrate receives only one cell.

[0535] In some embodiments, the cellular component binding reagents can be additionally conjugated with fluorescent molecules to enable flow sorting of cells into individual compartments.

[0536] In some embodiments, the methods disclosed herein provide contacting a plurality of compositions with the sample for specific binding with the plurality of cellular component targets. It would be appreciated that the conditions used may allow specific binding of the cellular component binding reagents, e.g., antibodies, to the cellular component targets. Following the contacting step, unbound compositions can be removed. For example, in embodiments where the sample comprises cells, and the compositions specifically bind to cellular component targets are cell-surface cellular components, such as cell-surface proteins, unbound compositions can be removed by washing the cells with buffer such that only compositions that specifically bind to the cellular component targets remain with the cells.

[0537] In some embodiments, the methods disclosed herein can comprise associating an oligonucleotide (e.g., a barcode, or a stochastic barcode), including a barcode sequence (such as a molecular label), a cell label, a sample label, etc., or any combination thereof, to the plurality of oligonucleotides associated with the cellular component binding reagents. For example, a plurality of oligonucleotide probes comprising a barcode can be used to hybridize to the plurality of oligonucleotides of the compositions.

[0538] In some embodiments, the plurality of oligonucleotide probes can be immobilized on solid supports. The solid supports can be free floating, e.g., beads in a solution. The solid supports can be embedded in a semi-solid or solid array. In some embodiments, the plurality of oligonucleotide probes may not be immobilized on solid supports. When the plurality of oligonucleotide probes are in close proximity to the plurality associated with oligonucleotides of the cellular component binding reagents, the plurality of oligonucleotides of the cellular component binding reagents can hybridize to the oligonucleotide probes. The oligonucleotide probes can be contacted at a non-depletable

ratio such that each distinct oligonucleotide of the cellular component binding reagents can associate with oligonucleotide probes having different barcode sequences (e.g., molecular labels) of the disclosure.

[0539] In some embodiments, the methods disclosed herein provide detaching the oligonucleotides from the cellular component binding reagents that are specifically bound to the cellular component targets. Detachment can be performed in a variety of ways to separate the chemical group from the cellular component binding reagent, such as UV photocleaving, chemical treatment (e.g., dithiothreitol treatment), heating, enzyme treatment, or any combination thereof. Detaching the oligonucleotide from the cellular component binding reagent can be performed either before, after, or during the step of hybridizing the plurality of oligonucleotide probes to the plurality of oligonucleotides of the compositions.

#### Methods of Simultaneous Quantitative Analysis of Cellular Component and Nucleic Acid Targets

[0540] In some embodiments, the methods disclosed herein can also be used for simultaneous quantitative analysis of a plurality of cellular component targets (e.g., protein targets) and a plurality of nucleic acid target molecules in a sample using the compositions disclosed herein and oligonucleotide probes that can associate a barcode sequence (e.g., a molecular label sequence) to both the oligonucleotides of the cellular component binding reagents and nucleic acid target molecules. Other methods of simultaneous quantitative analysis of a plurality of cellular component targets and a plurality of nucleic acid target molecules are described in U.S. Patent Application Number 15/715028, filed on September 25, 2017; the content of which is incorporated herein by reference in its entirety. In some embodiments, the sample can be a single cell, a plurality of cells, a tissue sample, a tumor sample, a blood sample, or the like. In some embodiments, the sample can comprise a mixture of cell types, such as normal cells, tumor cells, blood cells, B cells, T cells, maternal cells, fetal cells, or a mixture of cells from different subjects.

[0541] In some embodiments, the sample can comprise a plurality of single cells separated into individual compartments, such as microwells in a microwell array.

[0542] In some embodiments, the plurality of cellular component targets comprises a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, an

antibody, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. In some embodiments, the plurality of cellular component targets can comprise intracellular cellular components. In some embodiments, the plurality of cellular components can be, or be about, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or a number or a range between any two of these values, of all the cellular components, such as expressed proteins, in an organism, or one or more cell of the organism. In some embodiments, the plurality of cellular components can be at least, or be at most, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99%, of all the cellular components, such as proteins could be expressed, in an organism, or one or more cells of the organism. In some embodiments, the plurality of cellular component targets can comprise, or comprise about, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, 10000, or a number or a range between any two of these values, different cellular component targets. In some embodiments, the plurality of cellular component targets can comprise at least, or comprise at most, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, or 10000, different cellular component targets.

[0543] In some embodiments, the plurality of cellular component binding reagents is contacted with the sample for specific binding with the plurality of cellular component targets. Unbound cellular component binding reagents can be removed, for example, by washing. In embodiments where the sample comprises cells, any cellular component binding reagents not specifically bound to the cells can be removed.

[0544] In some instances, cells from a population of cells can be separated (e.g., isolated) into wells of a substrate of the disclosure. The population of cells can be diluted prior to separating. The population of cells can be diluted such that at least 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of wells of the substrate receive a single cell. The population of cells can be diluted such that at most 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of wells of the substrate receive a single cell. The population of cells can be diluted such that the number of cells in the diluted population is, or is at least, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the number of wells on the

substrate. The population of cells can be diluted such that the number of cells in the diluted population is, or is at least, 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% of the number of wells on the substrate. In some instances, the population of cells is diluted such that the number of cell is about 10% of the number of wells in the substrate.

**[0545]** Distribution of single cells into wells of the substrate can follow a Poisson distribution. For example, there can be at least a 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, or 10%, or more probability that a well of the substrate has more than one cell. There can be at least a 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, or 10%, or more probability that a well of the substrate has more than one cell. Distribution of single cells into wells of the substrate can be random. Distribution of single cells into wells of the substrate can be non-random. The cells can be separated such that a well of the substrate receives only one cell.

**[0546]** In some embodiments, the cellular component binding reagents can be additionally conjugated with fluorescent molecules to enable flow sorting of cells into individual compartments.

**[0547]** In some embodiments, the methods disclosed herein provide contacting a plurality of compositions with the sample for specific binding with the plurality of cellular component targets. It would be appreciated that the conditions used may allow specific binding of the cellular component binding reagents, e.g., antibodies, to the cellular component targets. Following the contacting step, unbound compositions can be removed. For example, in embodiments where the sample comprises cells, and the compositions specifically bind to cellular component targets are on the cell surface, such as cell-surface proteins, unbound compositions can be removed by washing the cells with buffer such that only compositions that specifically bind to the cellular component targets remain with the cells.

**[0548]** In some embodiments, the methods disclosed herein can provide releasing the plurality of nucleic acid target molecules from the sample, e.g., cells. For example, the cells can be lysed to release the plurality of nucleic acid target molecules. Cell lysis may be accomplished by any of a variety of means, for example, by chemical treatment, osmotic shock, thermal treatment, mechanical treatment, optical treatment, or any combination

thereof. Cells may be lysed by addition of a cell lysis buffer comprising a detergent (e.g., SDS, Li dodecyl sulfate, Triton X-100, Tween-20, or NP-40), an organic solvent (e.g., methanol or acetone), or digestive enzymes (e.g., proteinase K, pepsin, or trypsin), or any combination thereof.

**[0549]** It would be appreciated by one of ordinary skill in the art that the plurality of nucleic acid molecules can comprise a variety of nucleic acid molecules. In some embodiments, the plurality of nucleic acid molecules can comprise, DNA molecules, RNA molecules, genomic DNA molecules, mRNA molecules, rRNA molecules, siRNA molecules, or a combination thereof, and can be double-stranded or single-stranded. In some embodiments, the plurality of nucleic acid molecules comprise, or comprise about, 100, 1000, 10000, 20000, 30000, 40000, 50000, 100000, 1000000, or a number or a range between any two of these values, species. In some embodiments, the plurality of nucleic acid molecules comprise at least, or comprise at most, 100, 1000, 10000, 20000, 30000, 40000, 50000, 100000, or 1000000, species. In some embodiments, the plurality of nucleic acid molecules can be from a sample, such as a single cell, or a plurality of cells. In some embodiments, the plurality of nucleic acid molecules can be pooled from a plurality of samples, such as a plurality of single cells.

**[0550]** In some embodiments, the methods disclosed herein can comprise associating a barcode (e.g., a stochastic barcode), which can include a barcode sequence (such as a molecular label), a cell label, a sample label, etc., or any combination thereof, to the plurality of nucleic acid target molecules and the plurality of oligonucleotides of the cellular component binding reagents. For example, a plurality of oligonucleotide probes comprising a stochastic barcode can be used to hybridize to the plurality of nucleic acid target molecules and the plurality of oligonucleotides of the compositions.

**[0551]** In some embodiments, the plurality of oligonucleotide probes can be immobilized on solid supports. The solid supports can be free floating, e.g., beads in a solution. The solid supports can be embedded in a semi-solid or solid array. In some embodiments, the plurality of oligonucleotide probes may not be immobilized on solid supports. When the plurality of oligonucleotide probes are in close proximity to the plurality of nucleic acid target molecules and the plurality of oligonucleotides of the cellular component binding reagents, the plurality of nucleic acid target molecules and the plurality



of oligonucleotides of the cellular component binding reagents can hybridize to the oligonucleotide probes. The oligonucleotide probes can be contacted at a non-depletable ratio such that each distinct nucleic acid target molecules and oligonucleotides of the cellular component binding reagents can associate with oligonucleotide probes having different barcode sequences (e.g., molecular labels) of the disclosure.

[0552] In some embodiments, the methods disclosed herein provide detaching the oligonucleotides from the cellular component binding reagents that are specifically bound to the cellular component targets. Detachment can be performed in a variety of ways to separate the chemical group from the cellular component binding reagent, such as UV photocleaving, chemical treatment (e.g., dithiothreitol treatment), heating, enzyme treatment, or any combination thereof. Detaching the oligonucleotide from the cellular component binding reagent can be performed either before, after, or during the step of hybridizing the plurality of oligonucleotide probes to the plurality of nucleic acid target molecules and the plurality of oligonucleotides of the compositions.

#### Simultaneous Quantitative Analysis of Protein and Nucleic Acid Targets

[0553] In some embodiments, the methods disclosed herein also can be used for simultaneous quantitative analysis of multiple types of target molecules, for example protein and nucleic acid targets. For example, the target molecules can be, or comprise, cellular components. FIG. 6 shows a schematic illustration of an exemplary method of simultaneous quantitative analysis of both nucleic acid targets and other cellular component targets (e.g., proteins) in single cells. In some embodiments, a plurality of compositions 605, 605b, 605c, etc., each comprising a cellular component binding reagent, such as an antibody, is provided. Different cellular component binding reagents, such as antibodies, which bind to different cellular component targets are conjugated with different unique identifiers. Next, the cellular component binding reagents can be incubated with a sample containing a plurality of cells 610. The different cellular component binding reagents can specifically bind to cellular components on the cell surface, such as a cell marker, a B-cell receptor, a T-cell receptor, an antibody, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. Unbound cellular component binding reagents can be removed, e.g., by washing the cells with a buffer. The cells with the cellular component binding reagents can

be then separated into a plurality of compartments, such as a microwell array, wherein a single compartment 615 is sized to fit a single cell and a single bead 620. Each bead can comprise a plurality of oligonucleotide probes, which can comprise a cell label that is common to all oligonucleotide probes on a bead, and barcode sequences (e.g., molecular label sequences). In some embodiments, each oligonucleotide probe can comprise a target binding region, for example, a poly(dT) sequence. The oligonucleotides 625 conjugated to the cellular component binding reagent can be detached from the cellular component binding reagent using chemical, optical or other means. The cell can be lysed 635 to release nucleic acids within the cell, such as genomic DNA or cellular mRNA 630. Cellular mRNA 630, oligonucleotides 625 or both can be captured by the oligonucleotide probes on bead 620, for example, by hybridizing to the poly(dT) sequence. A reverse transcriptase can be used to extend the oligonucleotide probes hybridized to the cellular mRNA 630 and the oligonucleotides 625 using the cellular mRNA 630 and the oligonucleotides 625 as templates. The extension products produced by the reverse transcriptase can be subject to amplification and sequencing. Sequencing reads can be subject to demultiplexing of sequences or identifies of cell labels, barcodes (e.g., molecular labels), genes, cellular component binding reagent specific oligonucleotides (e.g., antibody specific oligonucleotides), etc., which can give rise to a digital representation of cellular components and gene expression of each single cell in the sample.

#### Association of Barcodes

[0554] The oligonucleotides associated with the cellular component binding reagents (e.g., antigen binding reagents or protein binding reagents) and/or the nucleic acid molecules may randomly associate with the oligonucleotide probes. The oligonucleotides associated with the cellular component binding reagents, referred to herein as binding reagent oligonucleotides, can be, or comprise oligonucleotides of the disclosure, such as an antibody oligonucleotide, a sample indexing oligonucleotide, a cell identification oligonucleotide, a control particle oligonucleotide, a control oligonucleotide, an interaction determination oligonucleotide, etc. Association can, for example, comprise hybridization of an oligonucleotide probe's target binding region to a complementary portion of the target nucleic acid molecule and/or the oligonucleotides of the protein binding reagents. For

example, a oligo(dT) region of a barcode (e.g., a stochastic barcode) can interact with a poly(A) tail of a target nucleic acid molecule and/or a poly(A) tail of an oligonucleotide of a protein binding reagent. The assay conditions used for hybridization (e.g., buffer pH, ionic strength, temperature, etc.) can be chosen to promote formation of specific, stable hybrids.

[0555] The disclosure provides for methods of associating a molecular label with a target nucleic acid and/or an oligonucleotide associated with a cellular component binding reagent using reverse transcription. As a reverse transcriptase can use both RNA and DNA as template. For example, the oligonucleotide originally conjugated on the cellular component binding reagent can be either RNA or DNA bases, or both. A binding reagent oligonucleotide can be copied and linked (e.g., covalently linked) to a cell label and a barcode sequence (e.g., a molecular label) in addition to the sequence, or a portion thereof, of the binding reagent sequence. As another example, an mRNA molecule can be copied and linked (e.g., covalently linked) to a cell label and a barcode sequence (e.g., a molecular label) in addition to the sequence of the mRNA molecule, or a portion thereof.

[0556] In some embodiments, molecular labels can be added by ligation of an oligonucleotide probe target binding region and a portion of the target nucleic acid molecule and/or the oligonucleotides associated with (e.g., currently, or previously, associated with) with cellular component binding reagents. For example, the target binding region may comprise a nucleic acid sequence that can be capable of specific hybridization to a restriction site overhang (e.g., an EcoRI sticky-end overhang). The methods can further comprise treating the target nucleic acids and/or the oligonucleotides associated with cellular component binding reagents with a restriction enzyme (e.g., EcoRI) to create a restriction site overhang. A ligase (e.g., T4 DNA ligase) may be used to join the two fragments.

#### Determining the Number or Presence of Unique Molecular Label Sequences

[0557] In some embodiments, the methods disclosed herein comprise determining the number or presence of unique molecular label sequences for each unique identifier, each nucleic acid target molecule, and/or each binding reagent oligonucleotides (e.g., antibody oligonucleotides). For example, the sequencing reads can be used to determine the number of unique molecular label sequences for each unique identifier, each nucleic acid target molecule, and/or each binding reagent oligonucleotide. As another example, the sequencing

reads can be used to determine the presence or absence of a molecular label sequence (such as a molecular label sequence associated with a target, a binding reagent oligonucleotide, an antibody oligonucleotide, a sample indexing oligonucleotide, a cell identification oligonucleotide, a control particle oligonucleotide, a control oligonucleotide, an interaction determination oligonucleotide, etc. in the sequencing reads).

[0558] In some embodiments, the number of unique molecular label sequences for each unique identifier, each nucleic acid target molecule, and/or each binding reagent oligonucleotide indicates the quantity of each cellular component target (e.g., an antigen target or a protein target) and/or each nucleic acid target molecule in the sample. In some embodiments, the quantity of a cellular component target and the quantity of its corresponding nucleic acid target molecules, e.g., mRNA molecules, can be compared to each other. In some embodiments, the ratio of the quantity of a cellular component target and the quantity of its corresponding nucleic acid target molecules, e.g., mRNA molecules, can be calculated. The cellular component targets can be, for example, cell surface protein markers. In some embodiments, the ratio between the protein level of a cell surface protein marker and the level of the mRNA of the cell surface protein marker is low.

[0559] The methods disclosed herein can be used for a variety of applications. For example, the methods disclosed herein can be used for proteome and/or transcriptome analysis of a sample. In some embodiments, the methods disclosed herein can be used to identify a cellular component target and/or a nucleic acid target, i.e., a biomarker, in a sample. In some embodiments, the cellular component target and the nucleic acid target correspond to each other, i.e., the nucleic acid target encodes the cellular component target. In some embodiments, the methods disclosed herein can be used to identify cellular component targets that have a desired ratio between the quantity of the cellular component target and the quantity of its corresponding nucleic acid target molecule in a sample, e.g., mRNA molecule. In some embodiments, the ratio is, or is about, 0.001, 0.01, 0.1, 1, 10, 100, 1000, or a number or a range between any two of the above values. In some embodiments, the ratio is at least, or is at most, 0.001, 0.01, 0.1, 1, 10, 100, or 1000. In some embodiments, the methods disclosed herein can be used to identify cellular component targets in a sample that the quantity of its corresponding nucleic acid target molecule in the sample is, or is about, 1000, 100, 10, 5, 2, 1, 0, or a number or a range between any two of these values. In

some embodiments, the methods disclosed herein can be used to identify cellular component targets in a sample that the quantity of its corresponding nucleic acid target molecule in the sample is more than, or less than, 1000, 100, 10, 5, 2, 1, or 0.

#### Compositions and Kits

[0560] Some embodiments disclosed herein provide kits and compositions for simultaneous quantitative analysis of a plurality of cellular components (e.g., proteins) and/or a plurality of nucleic acid target molecules in a sample. The kits and compositions can, in some embodiments, comprise a plurality of cellular component binding reagents (e.g., a plurality of protein binding reagents) each conjugated with an oligonucleotide, wherein the oligonucleotide comprises a unique identifier for the cellular component binding reagent, and a plurality of oligonucleotide probes, wherein each of the plurality of oligonucleotide probes comprises a target binding region, a barcode sequence (e.g., a molecular label sequence), wherein the barcode sequence is from a diverse set of unique barcode sequences. In some embodiments, each of the oligonucleotides can comprise a molecular label, a cell label, a sample label, or any combination thereof. In some embodiments, each of the oligonucleotides can comprise a linker. In some embodiments, each of the oligonucleotides can comprise a binding site for an oligonucleotide probe, such as a poly(A) tail. For example, the poly(A) tail can be, e.g., oligodA<sub>18</sub> (unanchored to a solid support) or oligoA<sub>18</sub>V (anchored to a solid support). The oligonucleotides can comprise DNA residues, RNA residues, or both.

[0561] Disclosed herein includes a plurality of sample indexing compositions. Each of the plurality of sample indexing compositions can comprise two or more cellular component binding reagents. Each of the two or more cellular component binding reagents can be associated with a sample indexing oligonucleotide. At least one of the two or more cellular component binding reagents can be capable of specifically binding to at least one cellular component target. The sample indexing oligonucleotide can comprise a sample indexing sequence for identifying sample origin of one or more cells of a sample. Sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions can comprise different sequences.

[0562] Disclosed herein include kits comprising control particle compositions. In

some embodiments, the control particle composition comprises a plurality of control particle oligonucleotides associated with a control particle, wherein each of the plurality of control particle oligonucleotides comprises a control barcode sequence and a poly(dA) region. At least two of the plurality of control particle oligonucleotides can comprise different control barcode sequences. The control particle oligonucleotide can comprise a molecular label sequence. The control particle oligonucleotide can comprise a binding site for a universal primer. Also disclosed herein include kits for sequencing control. In some embodiments, the kit comprises: a control particle composition comprising a plurality of control particle oligonucleotides associated with a control particle, wherein each of the plurality of control particle oligonucleotides comprises a control barcode sequence and a poly(dA) region.

[0563] Disclosed herein include kits comprising a plurality of control compositions for sequencing control. In some embodiments, each of the plurality of control compositions comprises a cellular component binding reagent associated with a control oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of a plurality of binding targets, and wherein the control oligonucleotide comprises a control barcode sequence and a pseudo-target region comprising a sequence substantially complementary to the target-binding region of at least one of the plurality of barcodes. In some embodiments, the pseudo-target region comprises a poly(dA) region.

[0564] Disclosed herein include kits comprising sample indexing compositions for cell identification. In some embodiments. Each of two sample indexing compositions comprises a cellular component binding reagent (e.g., a protein binding reagent) associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of one or more cellular component targets (e.g., one or more protein targets), wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences. In some embodiments, the sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a universal primer, or a combination thereof.

[0565] Disclosed herein includes kits for cell identification. In some embodiments, the kit comprises: two or more sample indexing compositions. Each of the

two or more sample indexing compositions can comprise a cellular component binding reagent (e.g., an antigen binding reagent) associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences. In some embodiments, the sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a universal primer, or a combination thereof. Disclosed herein includes kits for multiplet identification. In some embodiments, the kit comprises two sample indexing compositions. Each of two sample indexing compositions can comprise a cellular component binding reagent (e.g., an antigen binding reagent) associated with a sample indexing oligonucleotide, wherein the antigen binding reagent is capable of specifically binding to at least one of one or more cellular component targets (e.g., antigen targets), wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences.

[0566] Disclosed herein include kits for identifying interactions between cellular components, for example protein-protein interactions. In some embodiments, the kit comprises: a first pair of interaction determination compositions, wherein each of the first pair of interaction determination compositions comprises a protein binding reagent associated with an interaction determination oligonucleotide, wherein the protein binding reagent of one of the first pair of interaction determination compositions is capable of specifically binding to a first protein target and a protein binding reagent of the other of the first pair of interaction determination compositions is capable of specifically binding to the second protein target, wherein the interaction determination oligonucleotide comprises an interaction determination sequence and a bridge oligonucleotide hybridization region, and wherein the interaction determination sequences of the first pair of interaction determination compositions comprise different sequences; and a plurality of bridge oligonucleotides each comprising two hybridization regions capable of specifically binding to the bridge oligonucleotide hybridization regions of the first pair of interaction determination compositions.

[0567] The unique identifiers (or oligonucleotides associated with cellular component binding reagents, such as binding reagent oligonucleotides, antibody

oligonucleotides, sample indexing oligonucleotides, cell identification oligonucleotides, control particle oligonucleotides, control oligonucleotides, or interaction determination oligonucleotides) can have any suitable length, for example, from about 25 nucleotides to about 45 nucleotides long. In some embodiments, the unique identifier can have a length that is, is about, is less than, is greater than, 4 nucleotides, 5 nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, 10 nucleotides, 15 nucleotides, 20 nucleotides, 25 nucleotides, 30 nucleotides, 35 nucleotides, 40 nucleotides, 45 nucleotides, 50 nucleotides, 55 nucleotides, 60 nucleotides, 70 nucleotides, 80 nucleotides, 90 nucleotides, 100 nucleotides, 200 nucleotides, or a range that is between any two of the above values.

**[0568]** In some embodiments, the unique identifiers are selected from a diverse set of unique identifiers. The diverse set of unique identifiers can comprise, or comprise about, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 5000, or a number or a range between any two of these values, different unique identifiers. The diverse set of unique identifiers can comprise at least, or comprise at most, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, or 5000, different unique identifiers. In some embodiments, the set of unique identifiers is designed to have minimal sequence homology to the DNA or RNA sequences of the sample to be analyzed. In some embodiments, the sequences of the set of unique identifiers are different from each other, or the complement thereof, by, or by about, 1 nucleotide, 2 nucleotides, 3 nucleotides, 4 nucleotides, 5 nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, 10 nucleotides, or a number or a range between any two of these values. In some embodiments, the sequences of the set of unique identifiers are different from each other, or the complement thereof, by at least, or by at most, 1 nucleotide, 2 nucleotides, 3 nucleotides, 4 nucleotides, 5 nucleotides, 6 nucleotides, 7 nucleotides, 8 nucleotides, 9 nucleotides, or 10 nucleotides.

**[0569]** In some embodiments, the unique identifiers can comprise a binding site for a primer, such as universal primer. In some embodiments, the unique identifiers can comprise at least two binding sites for a primer, such as a universal primer. In some embodiments, the unique identifiers can comprise at least three binding sites for a primer, such as a universal primer. The primers can be used for amplification of the unique identifiers, for example, by PCR amplification. In some embodiments, the primers can be



used for nested PCR reactions.

**[0570]** Any suitable cellular component binding reagents are contemplated in this disclosure, such as any protein binding reagents (e.g., antibodies or fragments thereof, aptamers, small molecules, ligands, peptides, oligonucleotides, etc., or any combination thereof). In some embodiments, the cellular component binding reagents can be polyclonal antibodies, monoclonal antibodies, recombinant antibodies, single-chain antibody (scAb), or fragments thereof, such as Fab, Fv, etc. In some embodiments, the plurality of protein binding reagents can comprise, or comprise about, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 5000, or a number or a range between any two of these values, different protein binding reagents. In some embodiments, the plurality of protein binding reagents can comprise at least, or comprise at most, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, or 5000, different protein binding reagents.

**[0571]** In some embodiments, the oligonucleotide is conjugated with the cellular component binding reagent through a linker. In some embodiments, the oligonucleotide can be conjugated with the protein binding reagent covalently. In some embodiment, the oligonucleotide can be conjugated with the protein binding reagent non-covalently. In some embodiments, the linker can comprise a chemical group that reversibly or irreversibly attached the oligonucleotide to the protein binding reagents. The chemical group can be conjugated to the linker, for example, through an amine group. In some embodiments, the linker can comprise a chemical group that forms a stable bond with another chemical group conjugated to the protein binding reagent. For example, the chemical group can be a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, etc. In some embodiments, the chemical group can be conjugated to the protein binding reagent through a primary amine on an amino acid, such as lysine, or the N-terminus. The oligonucleotide can be conjugated to any suitable site of the protein binding reagent, as long as it does not interfere with the specific binding between the protein binding reagent and its protein target. In embodiments where the protein binding reagent is an antibody, the oligonucleotide can be conjugated to the antibody anywhere other than the antigen-binding site, for example, the Fc region, the C<sub>H</sub>1 domain, the C<sub>H</sub>2 domain, the C<sub>H</sub>3 domain, the C<sub>L</sub> domain, etc. In some embodiments, each protein binding reagent can be conjugated with a single oligonucleotide

molecule. In some embodiments, each protein binding reagent can be conjugated with, or with about, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, or a number or a range between any two of these values, oligonucleotide molecules, wherein each of the oligonucleotide molecule comprises the same unique identifier. In some embodiments, each protein binding reagent can be conjugated with more than one oligonucleotide molecule, for example, at least, or at most, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, or 1000, oligonucleotide molecules, wherein each of the oligonucleotide molecule comprises the same unique identifier.

[0572] In some embodiments, the plurality of cellular component binding reagents (e.g., protein binding reagents) are capable of specifically binding to a plurality of cellular component targets (e.g., protein targets) in a sample. The sample can be, or comprise, a single cell, a plurality of cells, a tissue sample, a tumor sample, a blood sample, or the like. In some embodiments, the plurality of cellular component targets comprises a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, an antibody, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. In some embodiments, the plurality of cellular component targets can comprise intracellular proteins. In some embodiments, the plurality of cellular component targets can comprise intracellular proteins. In some embodiments, the plurality of cellular component targets can be, or be about, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or a number or a range between any two of these values of all cellular component targets (e.g., proteins expressed or could be expressed) in an organism. In some embodiments, the plurality of cellular component targets can be at least, or be at most, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99%, of all cellular component targets (e.g., proteins expressed or could be expressed) in an organism. In some embodiments, the plurality of cellular component targets can comprise, or comprise about, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, 10000, or a number or a range between any two of these values, different cellular component targets. In some embodiments, the plurality of cellular component targets can comprise at least, or comprise at most, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 1000, or 10000, different cellular component targets.

#### Sample Indexing Using Oligonucleotide-Conjugated Cellular Component Binding Reagent

[0573] Disclosed herein include methods for sample identification. In some embodiments, the method comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more cellular component targets, wherein each of the plurality of sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; removing unbound sample indexing compositions of the plurality of sample indexing compositions; barcoding (e.g., stochastically barcoding) the sample indexing oligonucleotides using a plurality of barcodes (e.g., stochastic barcodes) to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides.

[0574] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0575] An oligonucleotide-conjugated with an antibody, an oligonucleotide for conjugation with an antibody, or an oligonucleotide previously conjugated with an antibody is referred to herein as an antibody oligonucleotide (“AbOligo”). Antibody oligonucleotides in the context of sample indexing are referred to herein as sample indexing oligonucleotides.

An antibody conjugated with an antibody oligonucleotide is referred to herein as a hot antibody or an oligonucleotide antibody. An antibody not conjugated with an antibody oligonucleotide is referred to herein as a cold antibody or an oligonucleotide free antibody. An oligonucleotide-conjugated with a binding reagent (e.g., a protein binding reagent), an oligonucleotide for conjugation with a binding reagent, or an oligonucleotide previously conjugated with a binding reagent is referred to herein as a reagent oligonucleotide. Reagent oligonucleotides in the context of sample indexing are referred to herein as sample indexing oligonucleotides. A binding reagent conjugated with an antibody oligonucleotide is referred to herein as a hot binding reagent or an oligonucleotide binding reagent. A binding reagent not conjugated with an antibody oligonucleotide is referred to herein as a cold binding reagent or an oligonucleotide free binding reagent.

[0576] FIG. 7 shows a schematic illustration of an exemplary workflow using oligonucleotide-conjugated cellular component binding reagents for sample indexing. In some embodiments, a plurality of compositions 705a, 705b, etc., each comprising a binding reagent is provided. The binding reagent can be a protein binding reagent, such as an antibody. The cellular component binding reagent can comprise an antibody, a tetramer, an aptamers, a protein scaffold, or a combination thereof. The binding reagents of the plurality of compositions 705a, 705b can bind to an identical cellular component target. For example, the binding reagents of the plurality of compositions 705a, 705b can be identical (except for the sample indexing oligonucleotides associated with the binding reagents).

[0577] Different compositions can include binding reagents conjugated with sample indexing oligonucleotides with different sample indexing sequences. The number of different compositions can be different in different implementations. In some embodiments, the number of different compositions can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or a number or a range between any two of these values. In some embodiments, the number of different compositions can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000.

[0578] In some embodiments, the sample indexing oligonucleotides of binding reagents in one composition can include an identical sample indexing sequence. The sample

indexing oligonucleotides of binding reagents in one composition may not be identical. In some embodiments, the percentage of sample indexing oligonucleotides of binding reagents in one composition with an identical sample indexing sequence can be, or be about, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.9%, or a number or a range between any two of these values. In some embodiments, the percentage of sample indexing oligonucleotides of binding reagents in one composition with an identical sample indexing sequence can be at least, or be at most, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.9%.

**[0579]** The compositions 705a and 705b can be used to label samples of different samples. For example, the sample indexing oligonucleotides of the cellular component binding reagent in the composition 705a can have one sample indexing sequence and can be used to label cells 710a, shown as black circles, in a sample 707a, such as a sample of a patient. The sample indexing oligonucleotides of the cellular component binding reagents in the composition 705b can have another sample indexing sequence and can be used to label cells 710b, shown as hatched circles, in a sample 707b, such as a sample of another patient or another sample of the same patient. The cellular component binding reagents can specifically bind to cellular component targets or proteins on the cell surface, such as a cell marker, a B-cell receptor, a T-cell receptor, an antibody, a major histocompatibility complex, a tumor antigen, a receptor, or any combination thereof. Unbound cellular component binding reagents can be removed, e.g., by washing the cells with a buffer.

**[0580]** The cells with the cellular component binding reagents can be then separated into a plurality of compartments, such as a microwell array, wherein a single compartment 715a, 715b is sized to fit a single cell 710a and a single bead 720a or a single cell 710b and a single bead 720b. Each bead 720a, 720b can comprise a plurality of oligonucleotide probes, which can comprise a cell label that is common to all oligonucleotide probes on a bead, and molecular label sequences. In some embodiments, each

oligonucleotide probe can comprise a target binding region, for example, a poly(dT) sequence. The sample indexing oligonucleotides 725a conjugated to the cellular component binding reagent of the composition 705a can be configured to be detachable or non-detachable from the cellular component binding reagent. The sample indexing oligonucleotides 725a conjugated to the cellular component binding reagent of the composition 705a can be detached from the cellular component binding reagent using chemical, optical or other means. The sample indexing oligonucleotides 725b conjugated to the cellular component binding reagent of the composition 705b can be configured to be detachable or non-detachable from the cellular component binding reagent. The sample indexing oligonucleotides 725b conjugated to the cellular component binding reagent of the composition 705b can be detached from the cellular component binding reagent using chemical, optical or other means.

[0581] The cell 710a can be lysed to release nucleic acids within the cell 710a, such as genomic DNA or cellular mRNA 730a. The lysed cell 735a is shown as a dotted circle. Cellular mRNA 730a, sample indexing oligonucleotides 725a, or both can be captured by the oligonucleotide probes on bead 720a, for example, by hybridizing to the poly(dT) sequence. A reverse transcriptase can be used to extend the oligonucleotide probes hybridized to the cellular mRNA 730a and the oligonucleotides 725a using the cellular mRNA 730a and the oligonucleotides 725a as templates. The extension products produced by the reverse transcriptase can be subject to amplification and sequencing.

[0582] Similarly, the cell 710b can be lysed to release nucleic acids within the cell 710b, such as genomic DNA or cellular mRNA 730b. The lysed cell 735b is shown as a dotted circle. Cellular mRNA 730b, sample indexing oligonucleotides 725b, or both can be captured by the oligonucleotide probes on bead 720b, for example, by hybridizing to the poly(dT) sequence. A reverse transcriptase can be used to extend the oligonucleotide probes hybridized to the cellular mRNA 730b and the oligonucleotides 725b using the cellular mRNA 730b and the oligonucleotides 725b as templates. The extension products produced by the reverse transcriptase can be subject to amplification and sequencing.

[0583] Sequencing reads can be subject to demultiplexing of cell labels, molecular labels, gene identities, and sample identities (e.g., in terms of sample indexing sequences of sample indexing oligonucleotides 725a and 725b). Demultiplexing of cell

labels, molecular labels, and gene identities can give rise to a digital representation of gene expression of each single cell in the sample. Demultiplexing of cell labels, molecular labels, and sample identities, using sample indexing sequences of sample indexing oligonucleotides, can be used to determine a sample origin.

[0584] In some embodiments, cellular component binding reagents against cellular component binding reagents on the cell surface can be conjugated to a library of unique sample indexing oligonucleotides to allow cells to retain sample identity. For example, antibodies against cell surface markers can be conjugated to a library of unique sample indexing oligonucleotides to allow cells to retain sample identity. This will enable multiple samples to be loaded onto the same Rhapsody<sup>TM</sup> cartridge as information pertaining sample source is retained throughout library preparation and sequencing. Sample indexing can allow multiple samples to be run together in a single experiment, simplifying and shortening experiment time, and eliminating batch effect.

[0585] Disclosed herein include methods for sample identification. In some embodiments, the method comprise: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more cellular component targets, wherein each of the plurality of sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; removing unbound sample indexing compositions of the plurality of sample indexing compositions. The method can include barcoding (e.g., stochastically barcoding) the sample indexing oligonucleotides using a plurality of barcodes (e.g., stochastic barcodes) to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides.

**[0586]** In some embodiments, the method for sample identification comprises: contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing compositions, wherein each of the one or more cells comprises one or more cellular component targets, wherein each of the plurality of sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular component targets, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of at least two sample indexing compositions of the plurality of sample indexing compositions comprise different sequences; removing unbound sample indexing compositions of the plurality of sample indexing compositions; and identifying sample origin of at least one cell of the one or more cells based on the sample indexing sequence of at least one sample indexing oligonucleotide of the plurality of sample indexing compositions.

**[0587]** In some embodiments, identifying the sample origin of the at least one cell comprises: barcoding (e.g., stochastically barcoding) sample indexing oligonucleotides of the plurality of sample indexing compositions using a plurality of barcodes (e.g., stochastic barcodes) to create a plurality of barcoded sample indexing oligonucleotides; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying the sample origin of the cell based on the sample indexing sequence of at least one barcoded sample indexing oligonucleotide of the plurality of barcoded sample indexing oligonucleotides. In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides comprises stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

**[0588]** In some embodiments, identifying the sample origin of the at least one cell can comprise identifying the presence or absence of the sample indexing sequence of at least one sample indexing oligonucleotide of the plurality of sample indexing compositions. Identifying the presence or absence of the sample indexing sequence can comprise: replicating the at least one sample indexing oligonucleotide to generate a plurality of



replicated sample indexing oligonucleotides; obtaining sequencing data of the plurality of replicated sample indexing oligonucleotides; and identifying the sample origin of the cell based on the sample indexing sequence of a replicated sample indexing oligonucleotide of the plurality of sample indexing oligonucleotides that correspond to the least one barcoded sample indexing oligonucleotide in the sequencing data.

[0589] In some embodiments, replicating the at least one sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides comprises: prior to replicating the at least one barcoded sample indexing oligonucleotide, ligating a replicating adaptor to the at least one barcoded sample indexing oligonucleotide. Replicating the at least one barcoded sample indexing oligonucleotide can comprise replicating the at least one barcoded sample indexing oligonucleotide using the replicating adaptor ligated to the at least one barcoded sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides.

[0590] In some embodiments, replicating the at least one sample indexing oligonucleotide to generate the plurality of replicated sample indexing oligonucleotides comprises: prior to replicating the at least one barcoded sample indexing oligonucleotide, contacting a capture probe with the at least one sample indexing oligonucleotide to generate a capture probe hybridized to the sample indexing oligonucleotide; and extending the capture probe hybridized to the sample indexing oligonucleotide to generate a sample indexing oligonucleotide associated with the capture probe. Replicating the at least one sample indexing oligonucleotide can comprise replicating the sample indexing oligonucleotide associated with the capture probe to generate the plurality of replicated sample indexing oligonucleotides.

#### Daisy-Chaining Oligonucleotides

[0591] Disclosed herein include systems, methods, kits, and compositions for extending barcoded oligonucleotides (such as a sample indexing oligonucleotide, a control particle oligonucleotide, a control oligonucleotide, or an oligonucleotide conjugated to an antibody for determining protein expression), by PCR amplification of barcoded oligonucleotides using primers with long overhangs (e.g., 160 or more nucleotides in length). Such primers can be, for example, 200 or more nucleotides in length, such as Ultramer®

DNA Oligonucleotides, Megamer® Single-Stranded DNA Fragments, or single-stranded DNA fragments of gBlocks® Gene Fragments (Integrated DNA Technologies, Inc. (Coralville, Iowa)). The barcoded oligonucleotides can contain a common sequence (e.g., a common sequence that is 40 nucleotides in length) for the long primer to bind and extend during PCR. In some embodiments, the long overhang can include a secondary barcode sequence (e.g., a molecular label sequence) to add complexity to the sample. Thus, the length of barcoded oligonucleotides can be increased from, for example, 200-300 nucleotides, to more than 400 nucleotides. Such longer barcoded oligonucleotides can survive ever increasing size selection during DNA purification using, for example, Solid Phase Reversible Immobilization (SPRI)\_magnetic beads (Applied Biological Materials Inc. (Richmond, British Columbia, Canada)).

[0592] In some embodiments, oligonucleotides associated with (e.g., attached to) cellular component binding reagents (e.g., antibody molecules) can be barcoded (e.g., stochastically barcoded), after dissociation from the cellular component binding reagents, to generate the barcoded oligonucleotides. The barcoded oligonucleotides can be further lengthened using a primer with a long overhang. Thus, the lengths of the oligonucleotides associated with cellular component binding reagents and the lengths of barcoded oligonucleotides, which may be limited by DNA synthesis technologies or experimental design considerations, do not limit the subsequent workflow (e.g., size selection during DNA purification). One non-limiting exemplary experimental design consideration is the effects of long oligonucleotides attached to cellular component binding reagents on the reagents' binding specificity and/or efficacy. PCR daisy-chaining of barcoded oligonucleotides can overcome or address limitations of DNA synthesis technologies while maintaining antibody specificity and/or efficacy. The methods of the disclosure for lengthening barcoded oligonucleotides with a common sequence using a primer with a long overhang is referred to herein as daisy-chaining the barcoded oligonucleotides using daisy-chaining amplification primers to generate daisy-chaining elongated amplicons.

[0593] Disclosed herein include methods for extending sample indexing oligonucleotides (or other barcoded oligonucleotides, such as control particle oligonucleotides). In some embodiments, after contacting one or more cells from each of a plurality of samples with a sample indexing composition of a plurality of sample indexing

compositions, the sample indexing oligonucleotides can be barcoded using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides (See FIG. 7). The plurality of barcoded sample indexing oligonucleotides can be amplified using a plurality of primers with long overhangs (referred to herein as daisy-chaining amplification primers) to create a plurality of longer amplicons (referred to herein as daisy-chaining elongated amplicons). Such longer amplicons can be, for example, substantially longer (e.g., 160 nucleotides longer) than the barcoded sample indexing oligonucleotides.

**[0594]** FIGS. 8A8 and 8A2 show a schematic illustration of an exemplary workflow of daisy-chaining a barcoded oligonucleotide. In some embodiments, a sample indexing composition 805 can include a cellular component binding reagent, such as an antibody, that is associated with a sample indexing oligonucleotide 825. The sample indexing oligonucleotide 825 can include a sample indexing sequence 825s, a common region 825d for binding to a daisy-chaining amplification primer (referred to herein as a daisy-chaining amplification primer-binding region or sequence 825d), and a poly(A) tail 825a. In some embodiments, the common region 825d for binding to a daisy-chaining amplification primer can be the same for all or some of sample indexing oligonucleotides 825. In some embodiments, the daisy-chaining amplification primer-binding region 825d of different sample indexing oligonucleotides 825 are different. A sample indexing oligonucleotide 825 can be an mRNA mimic. The sample indexing sequences 825s of at least two sample indexing compositions 805 of the plurality of sample indexing compositions can comprise different sequences.

**[0595]** The daisy-chaining amplification primer-binding region 825d can be different in different implementations, ranging from 10 to 200 nucleotides in length. In some embodiments, daisy-chaining amplification primer-binding region 825d can be, or be about, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, or a number or a range between any two of these values, nucleotides in length. In some embodiments, the daisy-chaining amplification primer-binding region 825d can be at least, or be at most, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,

30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides in length.

[0596] The sample indexing oligonucleotides 825 can be barcoded using a plurality of barcodes 815 (e.g., barcodes 815 associated with a particle, such as a bead 81) to create a plurality of barcoded sample indexing oligonucleotides 840. In some embodiments, a barcode 815 can include a poly(dT) region 815t for binding to a sample indexing oligonucleotide 825, optionally a molecular label 815m (e.g., for determining the number of occurrences of the sample indexing oligonucleotides), a cell label 815c, and a universal label 815u. A barcoded sample indexing oligonucleotide 840 can include a reverse complement 825d\_rt of the daisy-chaining amplification primer-binding region 825d.

[0597] Barcoded sample indexing oligonucleotides 840 can be amplified using forward and reverse primers 845, 850, including a plurality of daisy-chaining amplification primers 845, to create a plurality of daisy-chaining elongated amplicons 860. A daisy-chaining amplification primer 845 can comprise the sequence 845d, a complement, a reverse complement, or a combination thereof, of the daisy-chaining amplification primer-binding region 825d (referred to herein as barcoded sample indexing oligonucleotide-binding region 845d). A daisy-chaining amplification primer-binding region 825d and a barcoded sample indexing oligonucleotide-binding region 845d can be the same lengths, or different lengths. The barcoded sample indexing oligonucleotide-binding region 845d of a daisy-chaining amplification primer 845 can bind to its reverse complementary sequence 825d\_rt on a barcoded sample indexing oligonucleotide 840. A daisy-chaining amplification primer 845 can include an overhang region 845o (the overhang region 845o is an overhang of the daisy-chaining amplification primer 845 when the daisy-chaining amplification primer 845 binds to a barcoded sample indexing oligonucleotide 850). The plurality of daisy-chaining elongated amplicons 860, or a portion thereof, can be sequenced for sample identification (or protein expression profiling, sequencing control, etc.). As illustrated in FIG. 8A1-8A2, a sample indexing oligonucleotide 825 can be 200 nucleotides in length, a barcoded sample indexing oligonucleotides 840 can be 240 nucleotides in length, and a daisy-chaining elongated amplicon 860 can be 400 nucleotides in length.

[0598] The barcoded sample indexing oligonucleotide-binding region 845d can have different lengths in different implementations, for example, ranging from 10 to 200 nucleotides in length. In some embodiments, the barcoded sample indexing oligonucleotide-binding region 845d can be, or be about, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, or a number or a range between any two of these values, nucleotides in length. In some embodiments, the barcoded sample indexing oligonucleotide-binding region 845d can be at least, or be at most, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides in length.

[0599] In some embodiments, at least two daisy-chaining amplification primers 845 can comprise barcoded sample indexing oligonucleotide-binding regions 845d with an identical sequence. Two or more daisy-chaining amplification primers 845 can comprise barcoded sample indexing oligonucleotide-binding regions 845d with different sequences. Some or all of daisy-chaining amplification primers 845 can comprise barcoded sample indexing oligonucleotide-binding regions 845d with an identical sequence. The barcoded sample indexing oligonucleotide-binding regions 845d can comprise the daisy-chaining amplification primer binding sequence 825d, a complement thereof, a reverse complement thereof, or a combination thereof. Daisy-chaining amplification primer binding sequences 825d of at least two sample indexing compositions 805 can comprise an identical sequence. Daisy-chaining amplification primer binding sequences 825d of at least two sample indexing compositions 805 can comprise different sequences.

[0600] In some embodiments, the overhang region 845o can comprise a barcode sequence (referred to herein as a daisy-chaining amplification primer barcode sequence), such as a molecular label. For example, two daisy-chaining amplification primers 845o can comprise overhang regions 845o with an identical daisy-chaining amplification primer

barcode sequence. As another example, two daisy-chaining amplification primers 845 can comprise overhang regions 845o with two different daisy-chaining amplification primer barcode sequences. Overhang regions 845o of daisy-chaining amplification primers 845 can comprise different daisy-chaining amplification primer barcode sequences.

**[0601]** The overhang region 845 can have different lengths in different implementations, for example, ranging from 50 to 1000 nucleotides in length. In some embodiments, the overhang region 845 can be, or be about, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 or a number or a range between any two of these values, nucleotides in length. In some embodiments, the overhang region 845 can be at least, or be at most, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, or 1000.

**[0602]** The daisy-chaining elongated amplicon 860 can have different lengths in different implementations, for example, ranging from 50 to 1000 nucleotides in length. In some embodiments, the daisy-chaining elongated amplicon 860 can be, or be about, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 or a number or a range between any two of these values, nucleotides in length. In some embodiments, the daisy-chaining elongated amplicon 860 can be at least, or be at most, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390,

400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, or 1000.

**[0603]** FIGS. 8B1 and 8B2 show a schematic illustration of another exemplary workflow of daisy-chaining a barcoded oligonucleotide. In this workflow, the reverse complementary sequence of a switch oligonucleotide 860 can be optionally added to a barcoded sample indexing oligonucleotide 840. During reverse transcription, upon reaching the 5' end of the sample indexing oligonucleotide 824, the terminal transferase activity of an enzyme (e.g., a reverse transcriptase, such as a Moloney murine leukemia virus (MMLV)) adds a few additional nucleotides (mostly deoxycytidine) to the 3' end of the newly synthesized strand 840. These bases can function as an anchoring site of the template switch (TS) oligonucleotide 860. Upon base pairing between the template switch oligonucleotide 860 and the appended deoxycytidine stretch, the enzyme "switches" template strands, from sample indexing oligonucleotide 825 to the template switch oligonucleotide 860, and continues replication to the 5' end of the TS oligonucleotide 860. Thus, the resulting barcoded sample indexing oligonucleotide 840 contains a reverse complement of the sample indexing oligonucleotide 825 and the template switch oligonucleotide 860.

**[0604]** Barcoded sample indexing oligonucleotides 840 with a reverse complement sequence of the template switch oligonucleotide 860 can be amplified using forward and reverse primers 845, 850 to create daisy-chaining elongated amplicons 860 with the reverse complement of the template switch oligonucleotide 860. One of the two primers can bind to the Read 1 sequence 845r1, a universal sequence for amplification, which can be equivalent to the universal label 845u in FIGS. 8A1-8A2. The second primer, a daisy-chaining amplification primers 845, can bind to the reverse complement sequence 825d\_rt of the daisy-chaining amplification primer-binding region 825d. Accordingly, a long tail can be added to the barcoded sample indexing oligonucleotide 860. After enzymatic fragmentation, end repair, A-tailing, and optionally sample index PCR, the resulting amplicons can include the P5 sequence 865p5, P7 sequence 865p7, Read 1 sequence 815r1, and Read 2 sequence 865r2 for next generation sequencing (e.g., using bridge amplification). As illustrated in FIGS. 8A1-8A2, a sample indexing oligonucleotide 825 can be 200 nucleotides in length, a

barcoded sample indexing oligonucleotides 840 can be approximately 260 nucleotides in length, and a daisy-chaining elongated amplicon 860 can be more than 406 nucleotides in length.

#### Single Cell Sequencing Control Particles

[0605] Disclosed herein includes control particle compositions that can be used for, for example, single cell sequencing control. In some embodiments, the control particle composition comprises a plurality of control particle oligonucleotides associated with a control particle. The control particle associated with the plurality of control particle oligonucleotides is referred to herein also as a functionalized control particle. FIG. 9A is a non-limiting exemplary schematic illustration of a particle functionalized with a plurality of oligonucleotides. FIG. 9A shows that the control particle oligonucleotide associated with the control particle can comprise a control barcode sequence and a poly(dA) region, mimicking a mRNA poly(A) tail. The control particle oligonucleotide can comprise a molecular label sequence, a binding site for a universal primer, or a combination thereof. The control particle oligonucleotides associated with the control particles can be the same or different from one another. In some embodiments, at least two of the control particle oligonucleotides associated with the control particle have different control barcode sequence. In some embodiments, a plurality of a first control particle oligonucleotides and a plurality of a second control oligonucleotides are associated with the control particle, wherein the first and the second particle oligonucleotides have different control barcode sequence.

[0606] A bead, such as the CompBead<sup>TM</sup> Plus (BD (Franklin Lake, New Jersey)) can be functionalized with antibodies conjugated with oligonucleotides. CompBeads Plus are about 7.5 microns in size, which is similar to the size of an immune cell. When functionalized with antibodies conjugated with oligonucleotides, CompBead Plus can be used as control cells or control particles for single cell workflows. The AbO functionalized bead can be used with any single cell workflow as a single cell sequencing control.

#### Control Particle Oligonucleotide

[0607] The length of the control barcode sequence can be different in different implementations. In some embodiments, the control barcode sequence is, or is about, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110,



120, 128, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, or a number or a range between any two of these values, nucleotides in length. In some embodiments, the control barcode sequence is at least, or is at most, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 128, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, or 1000 nucleotides in length.

[0608] The length of the control particle oligonucleotide can be different in different implementations. In some embodiments, the control particle oligonucleotide is, or is about, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 128, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, or a number or a range between any two of these values, nucleotides in length. In some embodiments, the control particle oligonucleotide is at least, or is at most, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 128, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, or 1000 nucleotides in length.

[0609] In some embodiments, the number of the plurality of control particle

oligonucleotides can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 1000000, 10000000, 100000000, 1000000000, or a number or a range between any two of these values. In some embodiments, the number of the plurality of control particle oligonucleotides can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 1000000, 10000000, 100000000, or 1000000000.

**[0610]** The plurality of control particle oligonucleotides can comprise the same or different control barcode sequences. For example, at least two of the plurality of control particle oligonucleotides can comprise different control barcode sequences. In some embodiments, the control barcode sequences of at least or at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 1000000, 10000000, 100000000, or 1000000000 of the plurality of control particle oligonucleotides can be identical. In some embodiments, the control barcode sequences of, or of about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 1000000, 10000000, 100000000, or a number or a range between any two of these values, of the plurality of control particle oligonucleotides can be identical.

**[0611]** The control barcode sequences of at least or at most 0.000000001%, 0.00000001%, 0.0000001%, 0.00001%, 0.0001%, 0.001%, 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values of the plurality of

control particle oligonucleotides can be identical. The control barcode sequences of, or of about, 0.000000001%, 0.00000001%, 0.0000001%, 0.000001%, 0.00001%, 0.0001%, 0.001%, 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, or a number or a range between any two of these values, of the plurality of control particle oligonucleotides can be identical.

**[0612]** In some embodiments, the control barcode sequence is not homologous to genomic sequences of a species. The control barcode sequence can be homologous to genomic sequences of a species. The species can be a non-mammalian species. The non-mammalian species can be a phage species. The phage species can be T7 phage, a PhiX phage, or a combination thereof.

#### Control Particle

**[0613]** In some embodiments, at least one of the plurality of control particle oligonucleotides is associated with the control particle through a linker. The at least one of the plurality of control particle oligonucleotides can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the at least one of the plurality of control particle oligonucleotides. The chemical group can comprise a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

**[0614]** The diameter of the control particle can be, or be about, 1-1000 micrometers, such as 10-100 micrometer or 7.5 micrometer. In some embodiments, the diameter of the control particle can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 micrometers, or a number or a range between any two of these values. In some embodiments, the diameter of the control particle can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 micrometers.

**[0615]** In some embodiments, the plurality of control particle oligonucleotides is

immobilized on the control particle. The plurality of control particle oligonucleotides can be partially immobilized on the control particle. The plurality of control particle oligonucleotides can be enclosed in the control particle. The plurality of control particle oligonucleotides can be partially enclosed in the control particle. The control particle can be disruptable.

[0616] In some embodiments, the control particle can be a bead. The bead can be, or comprise, a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a conjugated bead, a protein A conjugated bead, a protein G conjugated bead, a protein A/G conjugated bead, a protein L conjugated bead, an oligo(dT) conjugated bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, or any combination thereof. The control particle can comprise a material of polydimethylsiloxane (PDMS), polystyrene, glass, polypropylene, agarose, gelatin, hydrogel, paramagnetic, ceramic, plastic, glass, methylstyrene, acrylic polymer, titanium, latex, Sepharose, cellulose, nylon, silicone, or any combination thereof. The control particle can comprise a disruptable hydrogel particle.

Protein Binding Reagent

[0617] In some embodiments, the control particle is associated with a plurality of first protein binding reagents, and at least one of the plurality of first protein binding reagents is associated with one of the plurality of control particle oligonucleotides. FIG. 9B shows a non-limiting exemplary particle coated with a number of antibodies functionalized with oligonucleotides. The first protein binding reagent can comprise a first antibody (e.g., a primary antibody, or a secondary antibody). The control particle oligonucleotide can be conjugated to the first protein binding reagent through a linker. The first control particle oligonucleotide can comprise the linker. The linker can comprise a chemical group. The chemical group can be reversibly or irreversibly attached to the first protein binding reagent. The chemical group can comprise a UV photocleavable group, a disulfide bond, a streptavidin, a biotin, an amine, a disulfide linkage, or any combination thereof.

[0618] In some embodiments, the control particle is associated with a plurality of second protein binding reagents. At least one of the plurality of second protein binding reagents can be associated with one of the plurality of control particle oligonucleotides. FIG. 9C shows a non-limiting exemplary particle coated with a plurality of first antibodies

functionalized with oligonucleotides and a plurality of second antibodies not functionalized with oligonucleotides. The antibodies on the control particle can be titrated with ratios of hot antibodies (e.g., associated with control particle oligonucleotide) and cold antibodies (e.g., not associated with control particle oligonucleotides) to alter the amount of sequencing reads obtained from a control particle. The first antibodies and the second antibodies can be the same or different.

[0619] FIG. 9D is a non-limiting exemplary schematic illustration of a particle functionalized with a plurality of first control particle oligonucleotides, a plurality of second control particle oligonucleotides conjugated to a plurality of second antibodies and a plurality of third control particle oligonucleotides with relative high, medium, and low abundance. The plurality of first control particle oligonucleotides can be conjugated to a plurality of first protein binding reagents. The plurality of second control particle oligonucleotides can be conjugated to a plurality of second protein binding reagents. The plurality of third control particle oligonucleotides can be conjugated to a plurality of third protein binding reagents.

[0620] The relative abundance of the first, second, and third control particle oligonucleotides can mimic mRNAs with different expression levels. The control particle oligonucleotide associated with the first protein binding reagent and the control particle oligonucleotide associated with the second protein binding reagent can comprise different control barcode sequences. Different protein binding reagents, such as antibodies, and the different control particle oligonucleotides on the control particle can be titrated to generate a standard curve. The first protein binding reagents, the second protein binding reagents, or the third protein binding reagents can be identical or different protein binding reagents.

[0621] In some embodiments, the ratio of the number of the plurality of first control particle oligonucleotides and the number of the plurality of second (or third) control particle oligonucleotides can be, or be about, 1:1, 1:1.1, 1:1.2, 1:1.3, 1:1.4, 1:1.5, 1:1.6, 1:1.7, 1:1.8, 1:1.9, 1:2, 1:2.5, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:21, 1:22, 1:23, 1:24, 1:25, 1:26, 1:27, 1:28, 1:29, 1:30, 1:31, 1:32, 1:33, 1:34, 1:35, 1:36, 1:37, 1:38, 1:39, 1:40, 1:41, 1:42, 1:43, 1:44, 1:45, 1:46, 1:47, 1:48, 1:49, 1:50, 1:51, 1:52, 1:53, 1:54, 1:55, 1:56, 1:57, 1:58, 1:59, 1:60, 1:61, 1:62, 1:63, 1:64, 1:65, 1:66, 1:67, 1:68, 1:69, 1:70, 1:71, 1:72, 1:73, 1:74, 1:75, 1:76, 1:77, 1:78, 1:79, 1:80, 1:81, 1:82, 1:83, 1:84, 1:85, 1:86, 1:87, 1:88, 1:89, 1:90, 1:91, 1:92, 1:93, 1:94, 1:95,

1:96, 1:97, 1:98, 1:99, 1:100, or a number or a range between any two of these numbers. In some embodiments, the ratio of the number of the plurality of first control particle oligonucleotides and the number of the plurality of second (or third) control particle oligonucleotides can be at least, or be at most, 1:1, 1:1.1, 1:1.2, 1:1.3, 1:1.4, 1:1.5, 1:1.6, 1:1.7, 1:1.8, 1:1.9, 1:2, 1:2.5, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:21, 1:22, 1:23, 1:24, 1:25, 1:26, 1:27, 1:28, 1:29, 1:30, 1:31, 1:32, 1:33, 1:34, 1:35, 1:36, 1:37, 1:38, 1:39, 1:40, 1:41, 1:42, 1:43, 1:44, 1:45, 1:46, 1:47, 1:48, 1:49, 1:50, 1:51, 1:52, 1:53, 1:54, 1:55, 1:56, 1:57, 1:58, 1:59, 1:60, 1:61, 1:62, 1:63, 1:64, 1:65, 1:66, 1:67, 1:68, 1:69, 1:70, 1:71, 1:72, 1:73, 1:74, 1:75, 1:76, 1:77, 1:78, 1:79, 1:80, 1:81, 1:82, 1:83, 1:84, 1:85, 1:86, 1:87, 1:88, 1:89, 1:90, 1:91, 1:92, 1:93, 1:94, 1:95, 1:96, 1:97, 1:98, 1:99, or 1:100.

**[0622]** In some embodiments, the first protein binding reagent can be associated with a partner binding reagent (e.g., a secondary antibody), and the first protein binding reagent is associated with the control particle using the partner binding reagent. The partner binding reagent can comprise a partner antibody. The partner antibody can comprise an anti-cat antibody, an anti-chicken antibody, an anti-cow antibody, an anti-dog antibody, an anti-donkey antibody, an anti-goat antibody, an anti-guinea pig antibody, an anti-hamster antibody, an anti-horse antibody, an anti-human antibody, an anti-llama antibody, an anti-monkey antibody, an anti-mouse antibody, an anti-pig antibody, an anti-rabbit antibody, an anti-rat antibody, an anti-sheep antibody, or a combination thereof. The partner antibody can comprise an immunoglobulin G (IgG), a F(ab') fragment, a F(ab')<sub>2</sub> fragment, a combination thereof, or a fragment thereof.

**[0623]** In some embodiments, the first protein binding reagent is associated with two or more of the plurality of control particle oligonucleotides with an identical control barcode sequence. The first protein binding reagent can be associated with two or more of the plurality of control particle oligonucleotides with different control barcode sequences. In some embodiments, at least one of the plurality of first protein binding reagents is not associated with any of the plurality of control particle oligonucleotides. The first protein binding reagent associated with the control particle oligonucleotide and the first protein binding reagent not associated with any control particle oligonucleotide can be identical protein binding reagents.

Control Barcode Diversity

[0624] The plurality of control particle oligonucleotides associated with one control particle can comprise a number of control particle oligonucleotides with different control barcode sequences. The number of control barcode sequences that control particle oligonucleotides have can be different in different implementation. In some embodiments, the number of control barcode sequences that the control particle oligonucleotides have can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, 1000000, or a number or a range between any two of these values. In some embodiments, the number of control barcode sequences that the control particle oligonucleotides have can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, or 1000000.

[0625] In some embodiments, the number of control particle oligonucleotides with the same control particle oligonucleotide sequence can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, 1000000, or a number or a range between any two of these values. In some embodiments, the number of control particle oligonucleotides with the same control particle oligonucleotide sequence can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, or 1000000.

[0626] In some embodiments, the plurality of control particle oligonucleotides comprises a plurality of first control particle oligonucleotides each comprising a first control barcode sequence, and a plurality of second control particle oligonucleotides each comprising a second control barcode sequence, and the first control barcode sequence and the second control barcode sequence have different sequences. The number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be about the same. The number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be different. The number of the plurality of first control particle oligonucleotides can be at least 2 times, 10 times, 100 times, or more greater than the number of the plurality of second control particle oligonucleotides. In some embodiments, the ratio of the number of

the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be, or be about, 1:1, 1:1.1, 1:1.2, 1:1.3, 1:1.4, 1:1.5, 1:1.6, 1:1.7, 1:1.8, 1:1.9, 1:2, 1:2.5, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:21, 1:22, 1:23, 1:24, 1:25, 1:26, 1:27, 1:28, 1:29, 1:30, 1:31, 1:32, 1:33, 1:34, 1:35, 1:36, 1:37, 1:38, 1:39, 1:40, 1:41, 1:42, 1:43, 1:44, 1:45, 1:46, 1:47, 1:48, 1:49, 1:50, 1:51, 1:52, 1:53, 1:54, 1:55, 1:56, 1:57, 1:58, 1:59, 1:60, 1:61, 1:62, 1:63, 1:64, 1:65, 1:66, 1:67, 1:68, 1:69, 1:70, 1:71, 1:72, 1:73, 1:74, 1:75, 1:76, 1:77, 1:78, 1:79, 1:80, 1:81, 1:82, 1:83, 1:84, 1:85, 1:86, 1:87, 1:88, 1:89, 1:90, 1:91, 1:92, 1:93, 1:94, 1:95, 1:96, 1:97, 1:98, 1:99, 1:100, 1:200, 1:300, 1:400, 1:500, 1:600, 1:700, 1:800, 1:900, 1:1000, 1:2000, 1:3000, 1:4000, 1:5000, 1:6000, 1:7000, 1:8000, 1:9000, 1:10000, or a number or a range between any two of the values. In some embodiments, the ratio of the number of the plurality of first control particle oligonucleotides and the number of the plurality of second control particle oligonucleotides can be at least, or be at most, 1:1, 1:1.1, 1:1.2, 1:1.3, 1:1.4, 1:1.5, 1:1.6, 1:1.7, 1:1.8, 1:1.9, 1:2, 1:2.5, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8, 1:9, 1:10, 1:11, 1:12, 1:13, 1:14, 1:15, 1:16, 1:17, 1:18, 1:19, 1:20, 1:21, 1:22, 1:23, 1:24, 1:25, 1:26, 1:27, 1:28, 1:29, 1:30, 1:31, 1:32, 1:33, 1:34, 1:35, 1:36, 1:37, 1:38, 1:39, 1:40, 1:41, 1:42, 1:43, 1:44, 1:45, 1:46, 1:47, 1:48, 1:49, 1:50, 1:51, 1:52, 1:53, 1:54, 1:55, 1:56, 1:57, 1:58, 1:59, 1:60, 1:61, 1:62, 1:63, 1:64, 1:65, 1:66, 1:67, 1:68, 1:69, 1:70, 1:71, 1:72, 1:73, 1:74, 1:75, 1:76, 1:77, 1:78, 1:79, 1:80, 1:81, 1:82, 1:83, 1:84, 1:85, 1:86, 1:87, 1:88, 1:89, 1:90, 1:91, 1:92, 1:93, 1:94, 1:95, 1:96, 1:97, 1:98, 1:99, 1:100, 1:200, 1:300, 1:400, 1:500, 1:600, 1:700, 1:800, 1:900, 1:1000, 1:2000, 1:3000, 1:4000, 1:5000, 1:6000, 1:7000, 1:8000, 1:9000, or 1:10000.

#### Detectable Moiety

[0627] In some embodiments, the control particle is associated with a detectable moiety, for example an optical moiety, such as a fluorophore or a chromophore. The control particle oligonucleotide can be associated with a detectable moiety, for example an optical moiety. In some embodiments, the first protein binding reagent can be associated with an optical moiety (FIG. 9E). The second protein binding reagent can be associated with an optical moiety. A control particle associated with an optical moiety (e.g., a bead fluorescently tagged) can also be used for imaging and flow cytometry.

[0628] The detectable moiety can be selected from a group of spectrally-distinct



detectable moieties. Spectrally-distinct detectable moieties include detectable moieties with distinguishable emission spectra even if their emission spectral may overlap. Non-limiting examples of detectable moieties include Xanthene derivatives: fluorescein, rhodamine, Oregon green, eosin, and Texas red; Cyanine derivatives: cyanine, indocarbocyanine, oxacarbocyanine, thiocarbocyanine, and merocyanine; Squaraine derivatives and ring-substituted squaraines, including Seta, SeTau, and Square dyes; Naphthalene derivatives (dansyl and prodan derivatives); Coumarin derivatives; oxadiazole derivatives: pyridyloxazole, nitrobenzoxadiazole and benzoxadiazole; Anthracene derivatives: anthraquinones, including DRAQ5, DRAQ7 and CyTRAK Orange; Pyrene derivatives: cascade blue; Oxazine derivatives: Nile red, Nile blue, cresyl violet, oxazine 170; Acridine derivatives: proflavin, acridine orange, acridine yellow; Arylmethine derivatives: auramine, crystal violet, malachite green; and Tetrapyrrole derivatives: porphrin, phthalocyanine, bilirubin. Other non-limiting examples of detectable moieties include Hydroxycoumarin, Aminocoumarin, Methoxycoumarin, Cascade Blue, Pacific Blue, Pacific Orange, Lucifer yellow, NBD, R-Phycoerythrin (PE), PE-Cy5 conjugates, PE-Cy7 conjugates, Red 613, PerCP, TruRed, FluorX, Fluorescein, BODIPY-FL, Cy2, Cy3, Cy3B, Cy3.5, Cy5, Cy5.5, Cy7, TRITC, X-Rhodamine, Lissamine Rhodamine B, Texas Red, Allophycocyanin (APC), APC-Cy7 conjugates, Hoechst 33342, DAPI, Hoechst 33258, SYTOX Blue, Chromomycin A3, Mithramycin, YOYO-1, Ethidium Bromide, Acridine Orange, SYTOX Green, TOTO-1, TO-PRO-1, TO-PRO: Cyanine Monomer, Thiazole Orange, CyTRAK Orange, Propidium Iodide (PI), LDS 751, 7-AAD, SYTOX Orange, TOTO-3, TO-PRO-3, DRAQ5, DRAQ7, Indo-1, Fluo-3, Fluo-4, DCFH, DHR, and SNARF.

[0629] The excitation wavelength of the detectable moieties can vary, for example be, or be about, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 nanometers, or a number or a range between any two of these values. The emission wavelength of the detectable moieties can also vary, for example be, or be about, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120,

130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 nanometers, or a number or a range between any two of these values.

**[0630]** The molecular weights of the detectable moieties can vary, for example be, or be about, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 Daltons (Da), or a number or a range between any two of these values. The molecular weights of the detectable moieties can also vary, for example be, or be about, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 kilo Daltons (kDa), or a number or a range between any two of these values.

**[0631]** The group of spectrally distinct detectable moieties can, for example, include five different fluorophores, five different chromophores, a combination of five fluorophores and chromophores, a combination of four different fluorophores and a non-fluorophore, a combination of four chromophores and a non-chromophore, or a combination of four fluorophores and chromophores and a non-fluorophore non-chromophore. In some embodiments, the detectable moieties can be one of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or a number or a range between any two of these values, of spectrally-distinct moieties.

Control Particle Workflow

[0632] The AbO functionalized bead can be used with any single cell workflow as a single cell sequencing control. Single cell workflows can utilize microwell arrays or microwell cartridges (e.g., BD Rhapsody™) or microfluidics devices (e.g., 10X Genomics (San Francisco, CA), Drop-seq (McCarroll Lab, Harvard Medical School (Cambridge, Massachusetts); Macosko et al., Cell, 2015 May 21 16:5:1202, the content of which is incorporated herein by reference in its entirety), or Abseq (Mission Bio (San Francisco, CA); Shahi et al., Sci Rep. 2017 Mar 14;7:44447, the content of which is hereby incorporated by reference in its entirety) in combination with solid or semisolid particles associated with stochastic barcodes (e.g., BD Rhapsody, or Drop-seq) or disruptable hydrogel particles enclosing releasable stochastic barcodes (e.g., 10X Genomics, or Abseq). The functionalized bead can be a control for determining efficiency of single cell workflows, analogous to external RNA control consortiums (ERCCs) being used for bulk RNAseq or microarray studies.

[0633] Disclosed herein include methods for determining the numbers of targets using a plurality of control particle oligonucleotides. The methods for determining the number of targets (e.g., gene expression) can be used with other methods disclosed herein. For example, a workflow can be used for determining protein expression and gene expression using a plurality of control particle oligonucleotides. In some embodiments, the method comprises: barcoding (e.g., stochastically barcoding) a plurality of targets of a cell of a plurality of cells and a plurality of control particle oligonucleotides using a plurality of barcodes (e.g., stochastic barcodes) to create a plurality of barcoded targets (e.g., stochastically barcoded targets) and a plurality of barcoded control particle oligonucleotides (e.g., stochastically barcoded control particle oligonucleotides), wherein each of the plurality of barcodes comprises a cell label sequence, a barcode sequence (e.g., a molecular label), and a target-binding region, wherein at least two barcodes of the plurality of barcodes comprise different barcode sequences (e.g., molecular labels), and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence, wherein a control particle composition comprises a control particle associated with the plurality of control particle oligonucleotides, wherein each of the plurality of control particle oligonucleotides comprises a control barcode sequence and a pseudo-target region comprising a sequence substantially

complementary to the target-binding region of at least one of the plurality of barcodes. The method can comprise: obtaining sequencing data of the plurality of barcoded targets and the plurality of barcoded control particle oligonucleotides; counting the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the plurality of control particle oligonucleotides with the control barcode sequence in the sequencing data. The method can comprise: for at least one target of the plurality of targets: counting the number of barcode sequences with distinct sequences associated with the target in the sequencing data; and estimating the number of the target, wherein the number of the target estimated correlates with the number of barcodes sequences (e.g., molecular labels) with distinct sequences associated with the target counted and the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the control barcode sequence. In some embodiments, the pseudo-target region comprises a poly(dA) region. The pseudo-target region can comprise a subsequence of the target.

[0634] In some embodiments, the number of the target estimated can correlate with the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the target counted, the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the control barcode sequence, and the number of the plurality of control particle oligonucleotides comprising the control barcode sequence. The number of the target estimated can correlate with the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the target counted, and a ratio of the number of the plurality of control particle oligonucleotides comprising the control barcode sequence and the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the control barcode sequence.

[0635] For example, if the control particle has 100 control particle oligonucleotides with a particular control barcode sequence and the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the control barcode sequence (e.g., the number of control particle oligonucleotides with the control barcode sequence that survive the library preparation process) is 80, then the efficiency of the library preparation (e.g., reverse transcription, amplification, etc.) is 80%. Thus, data from different library preparations can be compared by normalizing using the library preparation efficiency.

[0636] As another example, the control particle can comprise five control particle

oligonucleotides with a particular control barcode sequencing mimicking a low expression gene. If the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the control barcode sequence is five, and a low expression gene is not detected, then a conclusion that the low expression gene is not expressed (or the cell has fewer than five mRNAs of the gene) can be made. However, if the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the control barcode sequence is zero, and a low expression gene is not detected, then a conclusion that the low expression gene is not expressed cannot be made.

[0637] Capture efficiency can be determined for control particle oligonucleotides with different abundance. Normalization can be performed based on the capture efficiency of control particle oligonucleotides with two or more control barcode sequences. In some embodiments, counting the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the plurality of control particle oligonucleotides with the control barcode sequence in the sequencing data comprises: counting the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the first control barcode sequence in the sequencing data; and counting the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the second control barcode sequence in the sequencing data. The number of the target estimated can correlate with the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the target counted, the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the first control barcode sequence, and the number of barcode sequences (e.g., molecular labels) with distinct sequences associated with the second control barcode sequence.

[0638] In some embodiments, the method comprises releasing the at least one of the plurality of control particle oligonucleotides from the control particle prior to stochastically barcoding the plurality of targets and the control particle and the plurality of control particle oligonucleotides.

[0639] In some embodiments, barcoding (e.g., stochastically barcoding) the plurality of targets and the plurality of control particle oligonucleotides using the plurality of barcodes (e.g., stochastic barcodes) comprises: contacting the plurality of barcodes with targets of the plurality of targets and control particle oligonucleotides of the plurality of

control particle oligonucleotides to generate barcodes hybridized to the targets and the control particle oligonucleotides; and extending the barcodes hybridized to the targets and the control particle oligonucleotides to generate the plurality of barcoded targets and the plurality of barcoded control particle oligonucleotides (e.g., the plurality of stochastically barcoded targets and the plurality of stochastically barcoded control particle oligonucleotides). Extending the barcodes can comprise extending the barcodes using a DNA polymerase, a reverse transcriptase, or a combination thereof.

[0640] In some embodiments, the method comprises amplifying the plurality of stochastically barcoded targets and the plurality of stochastically barcoded control particle oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of stochastically barcoded targets and the plurality of stochastically barcoded control particle oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the barcode sequence (e.g., molecular label) and at least a portion of the control particle oligonucleotide or at least a portion of the barcode sequence (e.g., molecular label) and at least a portion of the control particle oligonucleotide. Obtaining the sequencing data can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing the at least a portion of the barcode sequence (e.g., molecular label) and the at least a portion of the control particle oligonucleotide, or the at least a portion of the barcode sequence (e.g., molecular label) and the at least a portion of the control particle oligonucleotide.

*Microwell Cartridge or Array Workflow*

[0641] FIG. 10 is a schematic illustration of an exemplary workflow of using particles functionalized with oligonucleotides for single cell sequencing control. In some embodiments, a control particle composition comprises a plurality of control particle oligonucleotides associated with a control particle 1004. For example, a control particle 1040 can be associated with a control particle oligonucleotide 1025 conjugated to an antibody 1005 bound to the control particle 1040. A plurality of control particles 1040 functionalized with control particle oligonucleotides 1025 can be spiked into a plurality of cells at, for example, 5%. Control particles 1040 can be treated as “cells” in the subsequent workflow. The control particles 1040 can also be referred to as control cells or control cell particles. Cells 1010 and the control particles 1040 can be then separated into a plurality of

compartments, such as wells of a microwell array, wherein a single compartment 1015a, 1015b is sized to fit a single cell or control particle and a single bead 1020a, 1020b. Beads 1020a, 1020b can be loaded into the compartments 1015a, 1015b

[0642] Each bead can comprise a plurality of oligonucleotide probes, which can comprise a cell label that is common to all oligonucleotide probes on a bead, and barcode sequences (e.g., molecular label sequences). In some embodiments, each oligonucleotide probe can comprise a target binding region, for example, a poly(dT) sequence. The oligonucleotides 1025 conjugated to the antibody 1005 can be detached from the antibody 1005 using chemical, optical or other means. The cell 1010 can be lysed 1035 to release nucleic acids within the cell, such as genomic DNA or cellular mRNA 1030. Cellular mRNAs 1030 and control particle oligonucleotides 1025 can be captured by the oligonucleotide probes on beads 1020a, 1020b respectively, for example, by hybridizing to the poly(dT) sequence. Beads can be retrieved and the captured cellular mRNAs 1030 and control particle oligonucleotides 1025 (e.g., corresponding to around 5000 cells in total) can be pooled.

[0643] A reverse transcriptase can be used to extend the oligonucleotide probes hybridized to the cellular mRNA 1030 and the control particle oligonucleotides 1025 using the cellular mRNA 1030 and the oligonucleotides 1025 as templates. The extension products produced by the reverse transcriptase can be subject to amplification and sequencing. Sequencing reads can be subject to demultiplexing of a cell label, a barcode sequence (e.g., a molecular label), gene identity, control particle oligonucleotide sequence, etc., which can be used to determine single cell gene expression profiles and quantity efficiency of the entire or part of the workflow (e.g., cell capture efficiency). For example, the number of control particles captured can be determined based on the number of cell labels associated with the control barcode sequence in the data. The efficiency of the workflow can be a ratio of the number of control particles captured and the number of control particles spiked in.

#### Microfluidics Workflow

[0644] FIG. 10 is a schematic illustration of another exemplary workflow of using particles functionalized with oligonucleotides for single cell sequencing control. A plurality of control particles 1040 functionalized with control particle oligonucleotides 1025 can be spiked into a plurality of cells at, for example, 5%. Control particles 1040 can be treated as

“cells” in the subsequent workflow. The control particles 1040 can also be referred to as control cells or control cell particles. Cells 1010 and the control particles 1040 can be then separated using a microfluidics device into a plurality of compartments, such as droplets 1045a, 1045b. Each droplet 1045a, 1045b can include one cell 1010 or one control particle 1040 and a hydrogel bead 1020a, 1020b.

[0645] Each bead 1020a, 1020b can comprise a plurality of oligonucleotide probes, which can comprise a cell label that is common to all oligonucleotide probes on a bead, and barcode sequences (e.g., molecular label sequences). In some embodiments, each oligonucleotide probe can comprise a target binding region, for example, a poly(dT) sequence. The bead 1020a, 1020b can include reagents for the subsequent workflow (e.g., reverse transcription). The oligonucleotides 1025 conjugated to the antibody 1005 can be detached from the antibody 1005 using chemical, optical or other means. The cell 1010 can be lysed 1035 to release nucleic acids within the cell, such as genomic DNA or cellular mRNA 1030. Cellular mRNAs 630 and control particle oligonucleotides 1025 can be captured by the oligonucleotide probes released from beads 1020a, 1020b respectively, for example, by hybridizing to the poly(dT) sequence. A reverse transcriptase can be used to extend the oligonucleotide probes hybridized to the cellular mRNA 1030 and the oligonucleotides 1025 using the cellular mRNA 1030 and the oligonucleotides 1025 as templates.

[0646] After breaking up the droplets 1045a, 1045b, the extension products produced by the reverse transcriptase can be pooled and subject to amplification and sequencing. Sequencing reads can be subject to demultiplexing of cell label, molecular label, gene identity, control particle oligonucleotide sequence, etc. to determine single cell gene expression profiles and quantity efficiency of the entire or part of the workflow.

#### Control Oligonucleotides for Determining Single Cell Sequencing Efficiency

[0647] Also disclosed herein include methods, kits and systems for determining single cell sequencing efficiency. Such methods, kits and systems can be used in, or in combination with, any suitable methods, kits and systems disclosed herein, such as the methods, kits and systems for measuring cellular component expression level (for example protein expression level) using cellular component binding reagents associated with



oligonucleotides.

[0648] In some embodiments, by labeling single cells with antibodies conjugated with oligonucleotides (e.g., with a universal antibody or biomarker antibody) and generating next generation sequencing libraries with them, the signals from the oligonucleotides in NGS reads can be used to determine single cell NGS efficiency. This can then be used as a QC step or an evaluation tool for efficacy for different single cell sequencing platforms. For example, the control oligonucleotides can be used in any suitable methods, kits and systems disclosed herein, for example the methods, kits and systems for measuring cellular component expression level (for example protein expression level) using cellular component binding reagents associated with oligonucleotides.

[0649] Antibodies conjugated with oligonucleotides (referred to herein as “AbOs”) can be used with any single cell workflow as a single cell sequencing control. Single cell workflows can utilize microwell arrays or microwell cartridges (e.g., BD Rhapsody™) or microfluidics devices (e.g., 10X Genomics (San Francisco, CA), Drop-seq (McCarroll Lab, Harvard Medical School (Cambridge, Massachusetts); Macosko et al., *Cell*, 2015 May 21 16;5:1202, the content of which is incorporated herein by reference in its entirety), or Abseq (Mission Bio (San Francisco, CA); Shahi et al., *Sci Rep.* 2017 Mar 14;7:44447, the content of which is hereby incorporated by reference in its entirety) in combination with solid or semisolid particles associated with barcodes, such as stochastic barcodes (e.g., BD Rhapsody, or Drop-seq), or disruptable hydrogel particles enclosing releasable barcodes, such as stochastic barcodes (e.g., 10X Genomics, or Abseq). AbOs can be a control for determining efficiency of single cell workflows. For example, the single cell sequencing platform from 10X Genomics performs single cell capture using emulsions to encapsulate single cells in droplets. Because these droplets cannot be easily visualized, capture efficiency of single cells cannot be easily determined. The use of AbOs upstream of such single cell sequencing workflow allows users to evaluate the single cell capture efficiency after sequencing and rate of doublet formation.

[0650] FIG. 11 shows a schematic illustration of an exemplary workflow of using control oligonucleotide-conjugated antibodies for determining single cell sequencing efficiency. In some embodiments, one or more cells (e.g., 5000) 1100 can be stained with an antibody 1105 conjugated with a control oligonucleotide 1125 prior to being loading onto a

microwell 1115 of a microwell cartridge or array. Cells 1110 can be then separated into a plurality of compartments, such as wells of a microwell array, wherein a single compartment 1115 is sized to fit a single cell and a single bead 1120.

[0651] The bead can, for example, comprise a plurality of oligonucleotide probes, which can comprise a cell label that is common to all oligonucleotide probes on a bead, and barcode sequences (e.g., molecular label sequences). In some embodiments, each oligonucleotide probe can comprise a target binding region, for example, a poly(dT) sequence. The oligonucleotides 1125 conjugated to the antibody 1105 can be detached from the antibody 1105 using chemical, optical or other means. The cell 1110 can be lysed 1135 to release nucleic acids within the cell, such as genomic DNA or cellular mRNA 1130. Cellular mRNAs 1130 and control oligonucleotides 1125 can be captured by the oligonucleotide probes on a bead 1120, for example, by hybridizing to the poly(dT) sequence. Beads can be retrieved and the captured cellular mRNAs 1130 (e.g., corresponding to around 5000 cells in total) can be pooled.

[0652] A reverse transcriptase can be used to extend the oligonucleotide probes hybridized to the cellular mRNA 1130 and the oligonucleotides 1125 using the cellular mRNA 1130 and the oligonucleotides 1125 as templates. The extension products produced by the reverse transcriptase can be subject to amplification and sequencing. Sequencing reads can be subject to demultiplexing of a cell label, a barcode sequence (e.g., a molecular label), gene identity, control oligonucleotide sequence, etc. to determine single cell gene expression profiles and quantity efficiency of the entire or part of the workflow (e.g., cell capture efficiency). The number of cells that are captured and go through the library preparation successfully (e.g., fewer than 5000 cells) can be determined.

[0653] FIG. 12 shows another schematic illustration of an exemplary workflow of using control oligonucleotide-conjugated antibodies for determining single cell sequencing efficiency. In FIG. 12, droplets 1245a, 1245b containing a single cell 1210a, 1210b and a single particle 1220a, 1220b can be formed using a microfluidic device. The single cells 1210a, 1210b can be bound to antibodies 1205a, 1205b conjugated with control oligonucleotides 1225a, 1225b. After cell lysis and reverse transcription in droplets 1245a, 1245b, droplets can be broken up and the content pooled for library preparation. The number of cells that are captured and go through the library preparation successfully can be

determined.

[0654] FIGS. 13A-13C are plots showing that control oligonucleotides can be used for cell counting. FIGS. 13A-13B show that control oligonucleotides of AbOs can be used as a control for cell counting. The falling points of the mRNA counts plot and the control oligonucleotide counts plot can coincide if 100% capture and library preparation efficiency is achieved. FIG. 13C shows that using conventional mRNA-only cell label calling may miss transcriptionally low cells in the mist of transcriptionally high cells. This method may call cutoff at  $n$  cell labels. This may occur when quiescent T cells within a large population of activated T cells, where activated T cells can have several fold higher in RNA transcription. This may also occur when in a targeted panel (e.g., cancer panel), non-targeted cells (non-cancer cells) with low expression of targeted genes are going to be dropped off. However, since protein expression is much higher, transcriptionally low cells still have higher chance to be called. This method may call cutoff at  $m$  cell labels, where  $m > n$ .

[0655] Disclosed herein include methods for sequencing control. For example, the methods can be used for determining single cell sequencing efficiency. The methods for determining single cell sequencing efficiency can be used with other methods disclosed herein. For example, the method for used for single cell sequencing efficiency can be used with the method for determining protein expression. As another example, a workflow can be used for determining single cell sequencing efficiency, protein expression, and/or gene expression.

[0656] In some embodiments, the method comprises: contacting one or more cells of a plurality of cells with a control composition of a plurality of control compositions, wherein a cell of the plurality of cells comprises a plurality of targets and a plurality of protein targets, wherein each of the plurality of control compositions comprises a protein binding reagent associated with a control oligonucleotide, wherein the protein binding reagent is capable of specifically binding to at least one of the plurality of protein targets, and wherein the control oligonucleotide comprises a control barcode sequence and a pseudo-target region comprising a sequence substantially complementary to the target-binding region of at least one of the plurality of barcodes; barcoding the control oligonucleotides using a plurality of barcodes to create a plurality of barcoded control oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a barcode sequence (e.g., a

molecular label sequence), and a target-binding region, wherein the barcode sequences (e.g., the molecular label sequences) of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded control oligonucleotides; determining at least one characteristic (e.g., the number of cells that are captured and go through the library preparation successfully) of the one or more cells using at least one characteristic of the plurality of barcoded control oligonucleotides in the sequencing data. In some embodiments, the pseudo-target region comprises a poly(dA) region.

[0657] In some embodiments, determining the at least one characteristic of the one or more cells comprises: determining the number of cell label sequences with distinct sequences associated with the plurality of barcoded control oligonucleotides in the sequencing data; and determining the number of the one or more cells using the number of cell label sequences with distinct sequences associated with the plurality of barcoded control oligonucleotides. The method can comprise: determining single cell capture efficiency based the number of the one or more cells determined. The method can comprise: comprising determining single cell capture efficiency based on the ratio of the number of the one or more cells determined and the number of the plurality of cells.

[0658] In some embodiments, determining the at least one characteristic of the one or more cells using the characteristics of the plurality of barcoded control oligonucleotides in the sequencing data comprises: for each cell label in the sequencing data, determining the number of barcode sequences (e.g., molecular label sequences) with distinct sequences associated with the cell label and the control barcode sequence; and determining the number of the one or more cells using the number of barcode sequences with distinct sequences associated with the cell label and the control barcode sequence. Determining the number of barcode sequences with distinct sequences associated with the cell label and the control barcode sequence can comprise: for each cell label in the sequencing data, determining the number of barcode sequences with the highest number of distinct sequences associated with the cell label and the control barcode sequence. Determining the number of the one or more cells using the number of barcode sequences with distinct sequences associated with the cell label and the control barcode sequence can comprise: generating a

plot of the number of barcode sequences with the highest number of distinct sequences with the number of cell labels in the sequencing data associated with the number of barcode sequences with the highest number of distinct sequences; and determining a cutoff in the plot as the number of the one or more cells.

[0659] In some embodiments, the method comprises releasing the control oligonucleotide from the cellular component binding reagent (e.g., the protein binding reagent) prior to barcoding the control oligonucleotides. In some embodiments, the method comprises removing unbound control compositions of the plurality of control compositions. Removing the unbound control compositions can comprise washing the one or more cells of the plurality of cells with a washing buffer. Removing the unbound unbound control compositions can comprise selecting cells bound to at least one cellular component binding reagent of the control composition using flow cytometry.

[0660] In some embodiments, barcoding the control oligonucleotides comprises: barcoding the control oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded control oligonucleotides. In some embodiments, barcoding the plurality of control oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with control oligonucleotides of the plurality of control compositions to generate barcodes hybridized to the control oligonucleotides; and extending the stochastic barcodes hybridized to the control oligonucleotides to generate the plurality of barcoded control oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase, a reverse transcriptase, or a combination thereof. In some embodiments, the method comprises amplifying the plurality of barcoded control oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded control oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of the stochastic barcode sequence (e.g., a molecular label sequence) and at least a portion of the control oligonucleotide. In some embodiments, obtaining the sequencing data comprises obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing the at least a portion of the molecular label sequence and the at least a portion of the control oligonucleotide.

Cell Overloading and Multiplet Identification

[0661] Also disclosed herein include methods, kits and systems for identifying cell overloading and multiplet. Such methods, kits and systems can be used in, or in combination with, any suitable methods, kits and systems disclosed herein, for example the methods, kits and systems for measuring cellular component expression level (such as protein expression level) using cellular component binding reagents associated with oligonucleotides.

[0662] Using current cell-loading technology, when about 20000 cells are loaded into a microwell cartridge or array with ~60000 microwells, the number of microwells or droplets with two or more cells (referred to as doublets or multiplets) can be minimal. However, when the number of cells loaded increases, the number of microwells or droplets with multiple cells can increase significantly. For example, when about 50000 cells are loaded into about 60000 microwells of a microwell cartridge or array, the percentage of microwells with multiple cells can be quite high, such as 11-14%. Such loading of high number of cells into microwells can be referred to as cell overloading. However, if the cells are divided into a number of groups (e.g., 5), and cells in each group are labeled with sample indexing oligonucleotides with distinct sample indexing sequences, a cell label (e.g., a cell label of a barcode, such as a stochastic barcode) associated with two or more sample indexing sequences can be identified in sequencing data and removed from subsequent processing. In some embodiments, the cells are divided into a large number of groups (e.g., 10000), and cells in each group are labeled with sample indexing oligonucleotides with distinct sample indexing sequences, a sample label associated with two or more sample indexing sequences can be identified in sequencing data and removed from subsequent processing. In some embodiments, different cells are labeled with cell identification oligonucleotides with distinct cell identification sequences, a cell identification sequence associated with two or more cell identification oligonucleotides can be identified in sequencing data and removed from subsequent processing. Such higher number of cells can be loaded into microwells relative to the number of microwells in a microwell cartridge or array.

[0663] Disclosed herein includes methods for sample identification. In some embodiments, the method comprises: contacting a first plurality of cells and a second

plurality of cells with two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprise one or more cellular components, wherein each of the two sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular components, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of barcoded sample indexing oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a barcode sequence (e.g., a molecular label sequence), and a target-binding region, wherein the barcode sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying one or more cell label sequences that is each associated with two or more sample indexing sequences in the sequencing data obtained; and removing the sequencing data associated with the one or more cell label sequences that is each associated with two or more sample indexing sequences from the sequencing data obtained and/or excluding the sequencing data associated with the one or more cell label sequences that is each associated with two or more sample indexing sequences from subsequent analysis (e.g., single cell mRNA profiling, or whole transcriptome analysis). In some embodiments, the sample indexing oligonucleotide comprises a barcode sequence (e.g., a molecular label sequence), a binding site for a universal primer, or a combination thereof.

[0664] For example, the method can be used to load 50000 or more cells (compared to 10000-20000 cells) using sample indexing. Sample indexing can use oligonucleotide-conjugated cellular component binding reagents (e.g., antibodies) or cellular component binding reagents against a cellular component (e.g., a universal protein marker) to label cells from different samples with a unique sample index. When two or more cells from different samples, two or more cells from different populations of cells of a sample, or two or more cells of a sample, are captured in the same microwell or droplet, the combined “cell”

(or contents of the two or more cells) can be associated with sample indexing oligonucleotides with different sample indexing sequences (or cell identification oligonucleotides with different cell identification sequences). The number of different populations of cells can be different in different implementations. In some embodiments, the number of different populations can be, or be about, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or a number or a range between any two of these values. In some embodiments, the number of different populations can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100. The number, or the average number, of cells in each population can be different in different implementations. In some embodiments, the number, or the average number, of cells in each population can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or a number or a range between any two of these values. In some embodiments, the number, or the average number, of cells in each population can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100. When the number, or the average number, of cells in each population is sufficiently small (e.g., equal to, or fewer than, 50, 25, 10, 5, 4, 3, 2, or 1 cells per population), the sample indexing composition for cell overloading and multiplet identification can be referred to as cell identification compositions.

[0665] Cells of a sample can be divided into multiple populations by aliquoting the cells of the sample into the multiple populations. A “cell” associated with more than one sample indexing sequence in the sequencing data can be identified as a “multiplet” based on two or more sample indexing sequences associated with one cell label sequence (e.g., a cell label sequence of a barcode, such as a stochastic barcode) in the sequencing data. The sequencing data of a combined “cell” is also referred to herein as a multiplet. A multiplet can be a doublet, a triplet, a quartet, a quintet, a sextet, a septet, an octet, a nonet, or any combination thereof. A multiplet can be any  $n$ -plet. In some embodiments,  $n$  is, or is about, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or a range between any two of these values. In some embodiments,  $n$  is at least, or is at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20.

[0666] When determining expression profiles of single cells, two cells may be identified as one cell and the expression profiles of the two cells may be identified as the expression profile for one cell (referred to as a doublet expression profile). For example,



when determining expression profiles of two cells using barcoding (e.g., stochastic barcoding), the mRNA molecules of the two cells may be associated with barcodes having the same cell label. As another example, two cells may be associated with one particle (e.g., a bead). The particle can include barcodes with the same cell label. After lysing the cells, the mRNA molecules in the two cells can be associated with the barcodes of the particle, thus the same cell label. Doublet expression profiles can skew the interpretation of the expression profiles.

[0667] A doublet can refer to a combined “cell” associated with two sample indexing oligonucleotides with different sample indexing sequences. A doublet can also refer to a combined “cell” associated with sample indexing oligonucleotides with two sample indexing sequences. A doublet can occur when two cells associated with two sample indexing oligonucleotides of different sequences (or two or more cells associated with sample indexing oligonucleotides with two different sample indexing sequences) are captured in the same microwell or droplet, the combined “cell” can be associated with two sample indexing oligonucleotides with different sample indexing sequences. A triplet can refer to a combined “cell” associated with three sample indexing oligonucleotides all with different sample indexing sequences, or a combined “cell” associated with sample indexing oligonucleotides with three different sample indexing sequences. A quartet can refer to a combined “cell” associated with four sample indexing oligonucleotides all with different sample indexing sequences, or a combined “cell” associated with sample indexing oligonucleotides with four different sample indexing sequences. A quintet can refer to a combined “cell” associated with five sample indexing oligonucleotides all with different sample indexing sequences, or a combined “cell” associated with sample indexing oligonucleotides with five different sample indexing sequences. A sextet can refer to a combined “cell” associated with six sample indexing oligonucleotides all with different sample indexing sequences, or a combined “cell” associated with sample indexing oligonucleotides with six different sample indexing sequences. A septet can refer to a combined “cell” associated with seven sample indexing oligonucleotides all with different sample indexing sequences, or a combined “cell” associated with sample indexing oligonucleotides with seven different sample indexing sequences. An octet can refer to a combined “cell” associated with eight sample indexing oligonucleotides all with different

sample indexing sequences, or a combined “cell” associated with sample indexing oligonucleotides with eight different sample indexing sequences. A nonet can refer to a combined “cell” associated with nine sample indexing oligonucleotides all with different sample indexing sequences, or a combined “cell” associated with sample indexing oligonucleotides with nine different sample indexing sequences. A multiplet can occur when two or more cells associated with two or more sample indexing oligonucleotides of different sequences (or two or more cells associated with sample indexing oligonucleotides with two or more different sample indexing sequences) are captured in the same microwell or droplet, the combined “cell” can be associated with sample indexing oligonucleotides with two or more different sample indexing sequences.

[0668] As another example, the method can be used for multiplet identification, whether in the context of sample overloading or in the context of loading cells onto microwells of a microwell array or generating droplets containing cells. When two or more cells are loaded into one microwell, the resulting data from the combined “cell” (or contents of the two or more cells) is a multiplet with aberrant gene expression profile. By using sample indexing, one can recognize some of these multiplets by looking for cell labels that are each associated with or assigned to two or more sample indexing oligonucleotides with different sample indexing sequences (or sample indexing oligonucleotides with two or more sample indexing sequences). With sample indexing sequence, the methods disclosed herein can be used for multiplet identification (whether in the context of sample overloading or not, or in the context of loading cells onto microwells of a microwell array or generating droplets containing cells). In some embodiments, the method comprises: contacting a first plurality of cells and a second plurality of cells with two sample indexing compositions respectively, wherein each of the first plurality of cells and each of the second plurality of cells comprise one or more cellular components, wherein each of the two sample indexing compositions comprises a cellular component binding reagent associated with a sample indexing oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular components, wherein the sample indexing oligonucleotide comprises a sample indexing sequence, and wherein sample indexing sequences of the two sample indexing compositions comprise different sequences; barcoding the sample indexing oligonucleotides using a plurality of barcodes to create a plurality of

barcoded sample indexing oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a barcode sequence (e.g., a molecular label sequence), and a target-binding region, wherein barcode sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded sample indexing oligonucleotides; and identifying one or more multiplet cell label sequences that is each associated with two or more sample indexing sequences in the sequencing data obtained.

[0669] The number of cells that can be loaded onto microwells of a microwell cartridge or into droplets generated using a microfluidics device can be limited by the multiplet rate. Loading more cells can result in more multiplets, which can be hard to identify and create noise in the single cell data. With sample indexing, the method can be used to more accurately label or identify multiplets and remove the multiplets from the sequencing data or subsequent analysis. Being able to identify multiplets with higher confidence can increase user tolerance for the multiplet rate and load more cells onto each microwell cartridge or generating droplets with at least one cell each.

[0670] In some embodiments, contacting the first plurality of cells and the second plurality of cells with the two sample indexing compositions respectively comprises: contacting the first plurality of cells with a first sample indexing compositions of the two sample indexing compositions; and contacting the first plurality of cells with a second sample indexing compositions of the two sample indexing compositions. The number of pluralities of cells and the number of pluralities of sample indexing compositions can be different in different implementations. In some embodiments, the number of pluralities of cells and/or sample indexing compositions can be, or be about, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, 1000000, or a number or a range between any two of these values. In some embodiments, the number of pluralities of cells and/or sample indexing compositions can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, or 1000000. The number of cells can be different in different implementations. In some embodiments, the number, or the average number, of cells can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200,

300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, 1000000, or a number or a range between any two of these values. In some embodiments, the number, or the average number, or cells can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, or 1000000.

[0671] In some embodiments, the method comprises: removing unbound sample indexing compositions of the two sample indexing compositions. Removing the unbound sample indexing compositions can comprise washing cells of the first plurality of cells and the second plurality of cells with a washing buffer. Removing the unbound sample indexing compositions can comprise selecting cells bound to at least one cellular component binding reagent of the two sample indexing compositions using flow cytometry. In some embodiments, the method comprises: lysing the one or more cells from each of the plurality of samples.

[0672] In some embodiments, the sample indexing oligonucleotide is configured to be detachable or non-detachable from the cellular component binding reagent. The method can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent. Detaching the sample indexing oligonucleotide can comprise detaching the sample indexing oligonucleotide from the cellular component binding reagent by UV photocleaving, chemical treatment (e.g., using reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0673] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the sample indexing oligonucleotides to generate barcodes hybridized to the sample indexing oligonucleotides; and extending the barcodes hybridized to the sample indexing oligonucleotides to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded sample indexing oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded sample indexing oligonucleotides.

[0674] In some embodiments, the method comprises: amplifying the plurality of barcoded sample indexing oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded sample indexing oligonucleotides can comprise amplifying, using

polymerase chain reaction (PCR), at least a portion of barcode sequence (e.g., the molecular label sequence) and at least a portion of the sample indexing oligonucleotide. In some embodiments, obtaining the sequencing data of the plurality of barcoded sample indexing oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data comprises sequencing at least a portion of the barcode sequence and at least a portion of the sample indexing oligonucleotide. In some embodiments, identifying the sample origin of the at least one cell comprises identifying sample origin of the plurality of barcoded targets based on the sample indexing sequence of the at least one barcoded sample indexing oligonucleotide.

[0675] In some embodiments, barcoding the sample indexing oligonucleotides using the plurality of barcodes to create the plurality of barcoded sample indexing oligonucleotides comprises stochastically barcoding the sample indexing oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded sample indexing oligonucleotides.

[0676] In some embodiments, the method includes: barcoding a plurality of targets of the cell using the plurality of barcodes to create a plurality of barcoded targets, wherein each of the plurality of barcodes comprises a cell label sequence, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes to create the plurality of barcoded targets can include: contacting copies of the targets with target-binding regions of the barcodes; and reverse transcribing the plurality targets using the plurality of barcodes to create a plurality of reverse transcribed targets.

[0677] In some embodiments, the method comprises: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets. Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using a plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0678] In some embodiments, the method for cell identification comprise: contacting a first plurality of one or more cells and a second plurality of one or more cells with two cell identification compositions respectively, wherein each of the first plurality of one or more cells and each of the second plurality of one or more cells comprise one or more cellular components, wherein each of the two cell identification compositions comprises a cellular component binding reagent associated with a cell identification oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular components, wherein the cell identification oligonucleotide comprises a cell identification sequence, and wherein cell identification sequences of the two cell identification compositions comprise different sequences; barcoding the cell identification oligonucleotides using a plurality of barcodes to create a plurality of barcoded cell identification oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a barcode sequence (e.g., a molecular label sequence), and a target-binding region, wherein the barcode sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded cell identification oligonucleotides; and identifying one or more cell label sequences that is each associated with two or more cell identification sequences in the sequencing data obtained; and removing the sequencing data associated with the one or more cell label sequences that is each associated with two or more cell identification sequences from the sequencing data obtained and/or excluding the sequencing data associated with the one or more cell label sequences that is each associated with two or more cell identification sequences from subsequent analysis (e.g., single cell mRNA profiling, or whole transcriptome analysis). In some embodiments, the cell identification oligonucleotide comprises a barcode sequence (e.g., a molecular label sequence), a binding site for a universal primer, or a combination thereof.

[0679] A multiplet (e.g., a doublet, triplet, etc) can occur when two or more cells associated with two or more cell identification oligonucleotides of different sequences (or two or more cells associated with cell identification oligonucleotides with two or more different cell identification sequences) are captured in the same microwell or droplet, the combined "cell" can be associated with cell identification oligonucleotides with two or more

different cell identification sequences.

[0680] Cell identification compositions can be used for multiplet identification, whether in the context of cell overloading or in the context of loading cells onto microwells of a microwell array or generating droplets containing cells. When two or more cells are loaded into one microwell, the resulting data from the combined “cell” (or contents of the two or more cells) is a multiplet with aberrant gene expression profile. By using cell identification, one can recognize some of these multiplets by looking for cell labels (e.g., cell labels of barcodes, such as stochastic barcodes) that are each associated with or assigned to two or more cell identification oligonucleotides with different cell identification sequences (or cell identification oligonucleotides with two or more cell identification sequences). With cell identification sequence, the methods disclosed herein can be used for multiplet identification (whether in the context of sample overloading or not, or in the context of loading cells onto microwells of a microwell array or generating droplets containing cells). In some embodiments, the method comprises: contacting a first plurality of one or more cells and a second plurality of one or more cells with two cell identification compositions respectively, wherein each of the first plurality of one or more cells and each of the second plurality of one or more cells comprise one or more cellular components, wherein each of the two cell identification compositions comprises a cellular component binding reagent associated with a cell identification oligonucleotide, wherein the cellular component binding reagent is capable of specifically binding to at least one of the one or more cellular components, wherein the cell identification oligonucleotide comprises a cell identification sequence, and wherein cell identification sequences of the two cell identification compositions comprise different sequences; barcoding the cell identification oligonucleotides using a plurality of barcodes to create a plurality of barcoded cell identification oligonucleotides, wherein each of the plurality of barcodes comprises a cell label sequence, a barcode sequence (e.g., a molecular label sequence), and a target-binding region, wherein barcode sequences of at least two barcodes of the plurality of barcodes comprise different sequences, and wherein at least two barcodes of the plurality of barcodes comprise an identical cell label sequence; obtaining sequencing data of the plurality of barcoded cell identification oligonucleotides; and identifying one or more multiplet cell label sequences that is each associated with two or more cell identification sequences in the sequencing data

obtained.

[0681] The number of cells that can be loaded onto microwells of a microwell cartridge or into droplets generated using a microfluidics device can be limited by the multiplet rate. Loading more cells can result in more multiplets, which can be hard to identify and create noise in the single cell data. With cell identification, the method can be used to more accurately label or identify multiplets and remove the multiplets from the sequencing data or subsequent analysis. Being able to identify multiplets with higher confidence can increase user tolerance for the multiplet rate and load more cells onto each microwell cartridge or generating droplets with at least one cell each.

[0682] In some embodiments, contacting the first plurality of one or more cells and the second plurality of one or more cells with the two cell identification compositions respectively comprises: contacting the first plurality of one or more cells with a first cell identification compositions of the two cell identification compositions; and contacting the first plurality of one or more cells with a second cell identification compositions of the two cell identification compositions. The number of pluralities of cell identification compositions can be different in different implementations. In some embodiments, the number of cell identification compositions can be, or be about, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, 1000000, or a number or a range between any two of these values. In some embodiments, the number of cell identification compositions can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, or 1000000. The number, or average number, of cells in each plurality of one or more cells can be different in different implementations. In some embodiments, the number, or average number, of cells in each plurality of one or more cells can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, 1000000, or a number or a range between any two of these values. In some embodiments, the number, or average number, of cells in each plurality of one or more cells can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 10000, 100000, or 1000000.

[0683] In some embodiments, the method comprises: removing unbound cell identification compositions of the two cell identification compositions. Removing the



unbound cell identification compositions can comprise washing cells of the first plurality of one or more cells and the second plurality of one or more cells with a washing buffer. Removing the unbound cell identification compositions can comprise selecting cells bound to at least one cellular component binding reagent of the two cell identification compositions using flow cytometry. In some embodiments, the method comprises: lysing the one or more cells from each of the plurality of samples.

[0684] In some embodiments, the cell identification oligonucleotide is configured to be detachable or non-detachable from the cellular component binding reagent. The method can comprise detaching the cell identification oligonucleotide from the cellular component binding reagent. Detaching the cell identification oligonucleotide can comprise detaching the cell identification oligonucleotide from the cellular component binding reagent by UV photocleaving, chemical treatment (e.g., using reducing reagent, such as dithiothreitol), heating, enzyme treatment, or any combination thereof.

[0685] In some embodiments, barcoding the cell identification oligonucleotides using the plurality of barcodes comprises: contacting the plurality of barcodes with the cell identification oligonucleotides to generate barcodes hybridized to the cell identification oligonucleotides; and extending the barcodes hybridized to the cell identification oligonucleotides to generate the plurality of barcoded cell identification oligonucleotides. Extending the barcodes can comprise extending the barcodes using a DNA polymerase to generate the plurality of barcoded cell identification oligonucleotides. Extending the barcodes can comprise extending the barcodes using a reverse transcriptase to generate the plurality of barcoded cell identification oligonucleotides.

[0686] In some embodiments, the method comprises: amplifying the plurality of barcoded cell identification oligonucleotides to produce a plurality of amplicons. Amplifying the plurality of barcoded cell identification oligonucleotides can comprise amplifying, using polymerase chain reaction (PCR), at least a portion of barcode sequence (e.g., the molecular label sequence) and at least a portion of the cell identification oligonucleotide. In some embodiments, obtaining the sequencing data of the plurality of barcoded cell identification oligonucleotides can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data comprises sequencing at least a portion of the barcode sequence and at least a portion of the cell identification oligonucleotide. In some

embodiments, identifying the sample origin of the at least one cell comprises identifying sample origin of the plurality of barcoded targets based on the cell identification sequence of the at least one barcoded cell identification oligonucleotide.

[0687] In some embodiments, barcoding the cell identification oligonucleotides using the plurality of barcodes to create the plurality of barcoded cell identification oligonucleotides comprises stochastically barcoding the cell identification oligonucleotides using a plurality of stochastic barcodes to create a plurality of stochastically barcoded cell identification oligonucleotides.

#### Determining Cellular Component-Cellular Component Interactions

[0688] Disclosed herein include methods, for example multiplexed proximity ligation methods, for determining interactions between cellular components. Such methods, kits and systems can be used in combination with any suitable methods, kits and systems disclosed herein, for example the methods, kits and systems for measuring cellular component expression level (such as protein expression level) using cellular component binding reagents associated with oligonucleotides.

[0689] The methods disclosed herein can be used to detect, identify, determine, and/or measure interactions between one or multiple cellular components (for example, proteins). For example, the methods can be used to detect, identify, determine, and/or measure interactions between two, three, four, five, six, seven, eight, nine, ten, twenty, thirty, forty, fifty, one hundred, two hundred, three hundred, or more pairs of cellular components. The cellular component targets in one pair of cellular components targets can overlap with the cellular component targets in another pair of cellular component targets. For example, the first pair of cellular component targets can include protein A and protein B, and the second pair of cellular component targets can include protein A and protein C.

[0690] In some embodiments, the method uses cellular component binding reagents associated with interaction determination oligonucleotides to determine interactions between targets of the cellular component binding reagents (referred to herein as cellular component targets). For example, the method can use antibody-oligonucleotide conjugates to detect protein-protein interactions in single cells. The antibodies can be homodimers or heterodimers. Each cellular component binding reagent can be associated with (e.g., attached

to or bind to) an interaction determination oligonucleotide containing a unique interaction determination sequence. For example, each antibody can be conjugated to an oligonucleotide containing a unique barcode sequence. Determining the pairs of interaction determination oligonucleotides that are ligated together can reveal the cellular component targets that interact with each other. For example, determining the barcode pairs that are ligated together can reveal the proteins interact with each other. This method can also be used in conjunction with single cell mRNA sequencing methods, such as well-based and droplet-based single cell mRNA sequencing methods.

[0691] In some embodiments, the method can utilize linear amplification to generate amplicons for sequencing. The method may not generate a circular template and use rolling circle amplification to generate sufficient template for downstream detection by fluorescent probes, or other methods. The method can be used to determine interactions between cellular component targets in single cells. For example, the method can be used to detect hundreds of cellular component-cellular component interactions (e.g., protein-protein interactions) across thousands to tens of thousands of single cells and uses sequencing as readouts. Interactions between the same proteins or cellular component targets can also be determined. The method does not rely on fluorescence and quantitative PCR (qPCR) as the readouts, which require bulk RNA as input, can only process few cells (e.g., less than 100 cells) at a time, and can detect a limited amount of cellular component-cellular component interactions at a time. The multiplexed method can utilize barcoded oligonucleotides associated with or conjugated to antibodies or cellular component binding reagents can be used to determine cellular component-cellular component interactions in a large number of single cells in a high throughput manner.

[0692] Furthermore, the method can also be used with other single cell RNA sequencing methods. For example, one single workflow can be used to determine mRNA expression levels, protein expression levels (or levels of cellular component targets), and/or protein-protein interactions (or interactions between cellular components). As a result, in a single experiment, data about mRNA expression, protein expression, and protein-protein interactions from in a single cell (or multiple single cells) can be obtained.

[0693] In some embodiments, the method can be used to detect cellular component-cellular component interactions between two cells and intracellular cellular

component- cellular component interactions. For example, the method can be used to detect protein-protein interactions between two cells, as well as intracellular protein-protein interactions. The method may require additional cell preparation before antibody staining or incubation (e.g., cell fixation to allow binding of antibodies to intracellular proteins or targets).

[0694] FIGS. 14A-14F show a schematic illustration of an exemplary workflow of determining interactions between cellular components (e.g., proteins) using a pair of interaction determination compositions. The method can utilize two types of antibody-oligonucleotide compositions 1404a, 1404b (e.g., antibody-oligonucleotide conjugates) illustrated in FIG. 14A. The two types of antibody-oligonucleotide compositions 1404a, 1404b can include two identical or different antibodies 1408a, 1408b associated with two types of oligonucleotides 1412a, 1412b. The association (e.g., conjugation) can be reversible or non-reversible (e.g., cleavable or non-cleavable). One type of oligonucleotides 1412a (referred to herein as the type 1 oligonucleotide 1412a) can have a universal sequencing region 1416. In some embodiments, the universal sequencing region 1416 can be common to all antibody-oligonucleotide compositions for amplification with a universal primer. The other type of oligonucleotides 1412b (referred to herein as the type 2 oligonucleotide 1412b) can have a poly(dA) sequence 1420 for hybridization to barcodes, such as stochastic barcodes. For example, the poly(dA) sequence 1420 of the type two oligonucleotide 1412b can hybridize to stochastic barcodes associated with a bead (e.g., immobilized on the bead, partially immobilized on the bead, releasably enclosed in the bead, partially enclosed in the bead, or any combination thereof). Both oligonucleotide types 1412a, 1412b also include a barcode sequence 1424a, 1424b unique to each antibody 1404a, 1404b. The two oligonucleotide types 1412a, 1412b can include different sequences 1428a, 1428b for hybridization to a bridge oligo 1432 illustrated in FIG. 14B. Although the type one oligonucleotide 1412a and the type two oligonucleotide 1412b are shown to be associated with the antibodies 1408a, 1408b in the 5' to 3' direction and the 3' to 5' direction, they are illustrative only and are not intended to be limiting. In some embodiments, the type one oligonucleotide 1412a and the type two oligonucleotide 1412b can be associated with the antibodies 1408a, 1408b in the 3' to 5' direction and the 5' to 3' direction.

[0695] Referring to FIG. 14B, the bridge oligonucleotide 1432 can bind to the

bridge oligonucleotide hybridization regions 1428a, 1428b to bring the antibody-oligonucleotide compositions 1404a, 1404b together for ligation. The sequences of the bridge oligonucleotide hybridization regions 1428a, 1428b can be different for each composition 1404a, 1404b. In some embodiments, a different bridge oligonucleotide 1432 can be used for each antibody pair.

[0696] The methods can be based on bringing together pairs of antibody-oligonucleotide compositions 1404a, 1404b so that the associated oligonucleotides 1412a, 1412b can be joined together or ligated to each other, and later captured by barcodes (e.g., stochastic barcodes) associated with beads for sequencing. The oligonucleotides 1412a, 1412b can be ligated to each other if the targets 1436a, 1436b of the antibodies 1408a, 1408b are within a certain threshold distance or range of each other, such as 30-40 nm.

[0697] The multiplexed method can utilize barcoded oligonucleotides associated with or conjugated to antibodies or cellular component binding reagents to determine protein-protein interactions, or interactions between cellular components, in a large number of single cells in a high throughput manner. In some embodiments, the interactions between cellular components in, or in about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ , or a number or a range between any two of these values, cells can be determined in a cell workflow. In some embodiments, the interactions between cellular components in at least, or in at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ ,  $10^9$ , or a number or a range between any two of these values, cells can be determined in a cell workflow.

#### Staining

[0698] To perform the assay, cells can be first stained with antibody-oligonucleotide compositions 1404a, 1404b or pairs of antibody-oligonucleotide compositions 1404a, 1404b (or cellular component binding reagent-oligonucleotide compositions). The number of antibody-oligonucleotide compositions 1404a, 1404b (or cellular component binding reagents) can be different in different implementations. In some embodiments, the number of antibody-oligonucleotide compositions 1404a, 1404b can be, or

be about, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of antibody-oligonucleotide compositions 1404a, 1404b can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000.

**[0699]** The antibodies 1408a, 1408b (or cellular component binding reagents) of different compositions 1404a, 1404b can be associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different or identical barcode sequences 1424a, 1424b. In some embodiments, the number of antibodies 1408a, 1408b of different compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different barcode sequences or identical barcode sequences 1424a, 1424b can be, or be about, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of antibodies 1408a, 1408b of different compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different barcode sequences or identical barcode sequences 1424a, 1424b can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000. In some embodiments, the percentage of antibodies 1408a, 1408b of different compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different barcode sequences or identical barcode sequences 1424a, 1424b can be, or be about, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of antibodies 1408a, 1408b of different compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different barcode sequences or identical barcode sequences 1424a, 1424b can be at least, or be at most, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, or 100%.

**[0700]** Some, all, or none of the antibodies 1408a, 1408b (or cellular component binding reagents) can be associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a,

1428b with identical sequence. In some embodiments, the number of antibodies 1408a, 1408b associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a, 1428b with identical sequence or different sequences can be, or be about, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of antibodies 1408a, 1408b associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a, 1428b with identical sequence or different sequences can be at least, or be at most, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000. In some embodiments, the percentage of antibodies 1408a, 1408b associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a, 1428b with identical sequence or different sequences can be, or be about, 0%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of antibodies 1408a, 1408b associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a, 1428b with identical sequence or different sequences can be at least, or be at most, 0%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, or 100%.

**[0701]** Some, all, or none of the antibodies 1408a (or cellular component binding reagents) can be associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions with identical sequence or different sequences. In some embodiments, the number of antibodies 1408a associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions 1416a with identical sequence or different sequences can be, or be about, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of antibodies 1408a associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions 1416a with identical sequence or different sequences can be at least, or be at most, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300,

400, 500, 600, 700, 800, 900, or 1000. In some embodiments, the percentage of antibodies 1408a associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions 1416a with identical sequence or different sequences can be, or be about, 0%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, 100%, or a number or a range between any two of these values. In some embodiments, the number of antibodies 1408a associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions 1416a with identical sequence or different sequences can be at least, or be at most, 0%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, or 100%.

**[0702]** The number of pairs of antibody-oligonucleotide compositions 1404a, 1404b (cellular component binding reagent-oligonucleotide compositions) can be different in different implementations. In some embodiments, the number of pairs of antibody-oligonucleotide compositions 1404a, 1404b can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of pairs of antibody-oligonucleotide compositions 1404a, 1404b can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000.

**[0703]** The antibodies 1408a, 1408b (or cellular component binding reagents) of different pairs of compositions 1404a, 1404b can be associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different or the identical barcode sequences 1424a, 1424b. In some embodiments, the number of antibodies 1408a, 1408b of different pairs of compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different barcode sequences or identical barcode sequences 1424a, 1424b can be, or be about, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of antibodies 1408a, 1408b of different pairs of compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different barcode sequences or identical barcode sequences 1424a, 1424b can be at least, or be at most, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30,



40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000. In some embodiments, the percentage of antibodies 1408a, 1408b of different pairs of compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different barcode sequences or identical barcode sequences 1424a, 1424b can be, or be about, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of antibodies 1408a, 1408b of different pairs of compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b with different barcode sequences or identical barcode sequences 1424a, 1424b can be at least, or be at most, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, or 100%.

**[0704]** Some, all, or none of the antibodies 1408a, 1408b (or cellular component binding reagents) of different pairs of compositions 1404a, 1404b can be associated with or bind to interaction determination oligonucleotides 1412a, 1412b with identical bridge oligonucleotide hybridization regions. In some embodiments, the number of antibodies 1408a, 1408b of different pairs of compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a, 1428b with identical sequence or different sequences can be, or be about, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of antibodies 1408a, 1408b of different pairs of compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a, 1428b with identical sequence or different sequences can be at least, or be at most, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000. In some embodiments, the % of antibodies 1408a, 1408b of different pairs of compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a, 1428b with identical sequence or different sequences can be, or be about, 0%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of antibodies

1408a, 1408b of different pairs of compositions 1404a, 1404b associated with or bind to interaction determination oligonucleotides 1412a, 1412b having bridge oligonucleotide hybridization regions 1428a, 1428b with identical sequence or different sequences can be at least, or be at most, 0%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, or 100%.

[0705] Some, all, or none of the antibodies 1408a (or cellular component binding reagents) of different compositions 1404a can be associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions with identical sequence or different sequences. In some embodiments, the number of antibodies 1408a of different compositions 1404a associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions 1416a with identical sequence or different sequences can be, or be about, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of antibodies 1408a of different compositions 1404a associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions 1416a with identical sequence or different sequences can be at least, or be at most, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000. In some embodiments, the percentage of antibodies 1408a of different compositions 1404a associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions 1416a with identical sequence or different sequences can be, or be about, 0%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, 100%, or a number or a range between any two of these values. In some embodiments, the percentage of antibodies 1408a of different compositions 1404a associated with or bind to interaction determination oligonucleotides 1412a having universal sequencing regions 1416a with identical sequence or different sequences can be at least, or be at most, 0%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 33%, 40%, 50%, 60%, 70%, 80%, 90%, 99.9%, 99.99%, or 100%.

[0706] The number of targets 1436a, 1436b can be different in different implementations, such as 2 to at least 50-100 protein targets. In some embodiments, the number of targets 1436a, 1436b can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50,

60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, or a number or a range between any two of these values. In some embodiments, the number of targets 1436a, 1436b can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000.

#### Ligation

[0707] After washing away unbound compositions 1404a, 1404b, a bridge oligonucleotide 1432 can be added. Pairs made of type 1 and type 2 compositions 1404a, 1404b can be linked together via the bridge oligonucleotide 1432 which can hybridize to bridge oligonucleotide hybridization regions. For example, pairs made of type 1 and type 2 compositions 1404a, 1404b can be linked together via the bridge oligonucleotide 1432 which can hybridize to bridge oligonucleotide hybridization regions if the corresponding protein targets 1436a, 1436b are within a threshold distance, such as 30-40nm. DNA ligase can then be added to ligate the two oligonucleotides 1412a, 1412b to form a ligated oligonucleotide 1412, resulting in the compositions 1436a, 1436b being ligated together. For example, the bridge oligonucleotide 1432 can make the ligation region double-stranded to facilitate ligation. The bridge oligonucleotide 1432 can bind to the hybridization regions 1428a, 1428b on each oligonucleotide 1412a, 1412b, which allows a DNA ligase to ligate the two oligonucleotides 1412a, 1412b together.

[0708] The threshold distance can be different in different implementations. In some embodiments, the threshold distance can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 nm, or a number or a range between any two of these values. For example, the distance can be in the range 30-40 nm. In some embodiments, the threshold distance can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nm. The threshold distance can be affected by the lengths of the oligonucleotides 1412a, 1412b, locations of the associations or attachments of the oligonucleotides 1412a, 1412b to the antibodies 1408a, 1408b, the size of the targets 1436a, 1436b, or any combination thereof.

#### Cell Lysis and Ligated Oligonucleotide Hybridization to Barcode

[0709] Then, cells can be lysed to allow capture of the ligated oligonucleotide 1412 as illustrated in FIG. 14C. For example, a bead 1438 with a barcode 1440 (e.g., a

stochastic barcode) having a poly(dT) region 1444 can capture the ligated oligonucleotide 1412 via the poly(dA) region. The bead 1438 can additionally include labels, such as cell and/or molecular labels 1448 and a universal sequence 1452 for subsequence amplification.

#### Reverse Transcription

[0710] Reverse transcription can then be performed to generate the complementary sequence 1456 of the ligated oligonucleotide 1412 shown in FIG. 14D. For example, a Moloney Murine Leukemia Virus (MMLV)-based or Taq-based reverse transcriptase can be used to generate the complementary sequence of the ligated oligonucleotide 1412. The MMLV-based reverse transcriptase has a strand-displacing mechanism to displace the bridge oligonucleotide 1432 during generation of the complementary strand. The Taq-based reverse transcriptase has 5' to 3' exonuclease activity to remove the bridge oligonucleotide 1432.

#### Library Preparation

[0711] After reverse transcription, a sequencing library can be prepared, for example, following standard protocols for single cell mRNA sequencing. Referring to FIG. 14E, the complementary sequence 1456 (or a reverse complement thereof 1456r) can be amplified via the universal sequence 1452 (e.g., the universal P5 sequence or a partial universal P5 sequence) on the barcode 1440 and the universal sequencing region 1416 (e.g., the universal N1 sequence) of the type one antibody-oligonucleotide composition 1404a. The superscript r denotes a reverse complement. The first round of PCR (referred to herein as PCR1) can use universal primers (e.g., primers with P5 and N1 sequences, or reverse complements thereof) to generate full length (or close to full length) amplicons that correspond to ligated oligonucleotide 1456r with one or more labels and the universal sequence.

[0712] The same primers can be used for the second round of PCR amplification (referred to herein as PCR2), or the N1 primer can be replaced by an N2 primer that binds to both bridge oligonucleotide hybridization regions 1428a, 1428b for further specificity as illustrated in FIG. 14F. When this N2 primer is used, a majority of ligated oligonucleotide 1456r with both the universal sequencing region 1416 and the universal sequence 1452 can be amplified. In some embodiments, when this N2 primer is used for the second round of PCR, a majority of ligated oligonucleotide 1456r with both the universal sequencing region 1416 and the universal sequence 1452 can be amplified.

Using Pairs of Interaction Determination Compositions to Determine Interactions between Cellular Components

[0713] Disclosed herein include systems, methods, and kits for determining protein-protein interactions. In some embodiments, the method comprises: contacting a cell with a first pair of interaction determination compositions 1404a, 1404b illustrated in FIG. 14A. The cell can comprise a first protein target 1436a and a second protein target 1436b. Each of the first pair of interaction determination compositions 1404a, 1404b comprises a protein binding reagent 1408a, 1408b associated with an interaction determination oligonucleotide 1412a, 1412b. The protein binding reagent 1408a, 1408b of one of the first pair of interaction determination compositions is capable of specifically binding to the first protein target 1436a, and the protein binding reagent 1408b of the other of the first pair of interaction determination compositions is capable of specifically binding to the second protein target 1436b. The interaction determination oligonucleotide 1412a, 1412b comprises an interaction determination sequence 1424a, 1434b and a bridge oligonucleotide hybridization region 1428a, 1428b. The interaction determination sequences 1424a, 1424b of the first pair of interaction determination compositions 1404a, 1404b can comprise different sequences.

[0714] Referring to FIG. 14B, the method can include ligating the interaction determination oligonucleotides 1412a, 1412b of the first pair of interaction determination compositions 1404a, 1404b using a bridge oligonucleotide 1432 to generate a ligated interaction determination oligonucleotide 1412. The bridge oligonucleotide 1432 can comprise two hybridization regions capable of specifically binding to the bridge oligonucleotide hybridization regions 1428a, 1428b of the first pair of interaction determination compositions 1404a, 1404b

[0715] Referring to FIG. 14C, the method can include barcoding the ligated interaction determination oligonucleotide 1412 using a plurality of barcodes 1440 to create a plurality of barcoded interaction determination oligonucleotides 1456. Each of the plurality of barcodes 1440 can comprise a barcode sequence 1448 and a capture sequence 1444. The method can include obtaining sequencing data of the plurality of barcoded interaction determination oligonucleotides 1456. The method can include determining an interaction between the first and second protein targets 1436a, 1436b based on the association of the

interaction determination sequences 1424a, 1424b of the first pair of interaction determination compositions 1404a, 1404b in the obtained sequencing data.

[0716] Disclosed herein include systems, methods, and kits for determining interactions between cellular component targets. In some embodiments, the method comprises: contacting a cell with a first pair of interaction determination compositions 1404a, 1404b illustrated in FIG. 14A. The cell can comprise a first cellular component target and a second cellular component target (e.g., a first protein target 1436a and a second protein target 1436b). Each of the first pair of interaction determination compositions can comprise a cellular component binding reagent (e.g., the antibody 1408a, 1408b) associated with an interaction determination oligonucleotide 1412a, 1412b. The cellular component binding reagent of one of the first pair of interaction determination compositions 1404a, 1404b is capable of specifically binding to the first cellular component target (e.g., the protein target 1436a), and the cellular component binding reagent of the other 1404b of the first pair of interaction determination compositions is capable of specifically binding to the second cellular component target (e.g., the protein target 1436b). The interaction determination oligonucleotide 1412a, 1412b can comprise an interaction determination sequence 1424a, 1424b and a bridge oligonucleotide hybridization region 1428a, 1428b. The interaction determination sequences 1424a, 1424b of the first pair of interaction determination compositions 1404a, 1404b can comprise different sequences.

[0717] Referring to FIG. 14B, the method can include ligating the interaction determination oligonucleotides 1412a, 1412b of the first pair of interaction determination compositions 1404a, 1404b using a bridge oligonucleotide 1432 to generate a ligated interaction determination oligonucleotide 1412. The bridge oligonucleotide 1432 can comprise two hybridization regions capable of specifically binding to the bridge oligonucleotide hybridization regions 1428a, 1428b of the first pair of interaction determination compositions 1404a, 1404b.

[0718] Referring to FIG. 14C, the method can include barcoding the ligated interaction determination oligonucleotide 1412 using a plurality of barcodes 1440 to create a plurality of barcoded interaction determination oligonucleotides 1456. Each of the plurality of barcodes 1440 can comprise a barcode sequence 1448 and a capture sequence 1444. The method can include obtaining sequencing data of the plurality of barcoded interaction

determination oligonucleotides 1456. The method can include determining an interaction between the first and second cellular component targets 1436a, 1436b based on the association of the interaction determination sequences 1424a, 1424b of the first pair of interaction determination compositions 1404a, 1404b in the obtained sequencing data. In some embodiments, at least one of the two cellular component binding reagent can comprise a protein binding reagent. The protein binding reagent can be associated with one of the two interaction determination oligonucleotides. The one or more cellular component targets can comprise at least one protein target.

**[0719]** In some embodiments, contacting the cell with the first pair of interaction determination compositions 1404a, 1404b comprises: contacting the cell with each of the first pair of interaction determination compositions 1404a, 1404b sequentially or simultaneously. The first protein target 1436a can be the same as the second protein target 1436b. The first protein target 1436a can be different from the second protein target 1436b. The first and the second cellular component targets can be the same or different.

**[0720]** The interaction determination oligonucleotides 1412a, 1412b can have different lengths in different implementations. In some embodiments, an interaction determination oligonucleotide 1412a, 1412b is, or is about, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 128, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000, or a number or a range between any two of these values, nucleotides in length. In some embodiments, an interaction determination oligonucleotide 1412a, 1412b is at least, or is at most, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 128, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, or 1000, nucleotides in

length.

[0721] In some embodiments, the method comprises: contacting the cell with a second pair of interaction determination compositions. The cell can comprise a third cellular component target and a fourth cellular component target (e.g., a third protein target and a fourth protein target). Each of the second pair of interaction determination compositions can comprise a cellular component binding reagent associated with an interaction determination oligonucleotide. The cellular component binding reagent of one of the second pair of interaction determination compositions can be capable of specifically binding to the third cellular component target and the cellular component binding reagent of the other of the second pair of interaction determination compositions can be capable of specifically binding to the fourth cellular component target. In some embodiments, at least one of the third and fourth cellular component targets can be different from one of the first and second cellular component targets. In some embodiments, at least one of the third and fourth cellular component targets and at least one of the first and second cellular component targets can be identical.

[0722] In some embodiments, the method comprises: contacting the cell with three or more pairs of interaction determination compositions. The interaction determination sequences of a number of interaction determination compositions of the plurality of pairs of interaction determination compositions can comprise the same or different sequences. In some embodiments, the number of interaction determination compositions having interaction determination sequences with the same or different sequences can be, or be about, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or a number or a range between any two of these values. In some embodiments, the number of interaction determination compositions having interaction determination sequences with the same or different sequences can be at least, or be at most, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000.

[0723] In some embodiments, the bridge oligonucleotide hybridization regions 1428a, 1428b of the first pair of interaction determination compositions 1404a, 1404b comprise different sequences. At least one of the bridge oligonucleotide hybridization



regions 1428a, 1428b can be complementary to at least one of the two hybridization regions of the bridge oligonucleotide 1432.

[0724] In some embodiments, ligating the interaction determination oligonucleotides 1412a, 1412b of the first pair of interaction determination compositions 1404a, 1404b using the bridge oligonucleotide 1432 comprises: hybridizing a first hybridization regions of the bridge oligonucleotide 1432 with a first bridge oligonucleotide hybridization region 1428a of the bridge oligonucleotide hybridization regions of the interaction determination oligonucleotides; hybridizing a second hybridization region of the bridge oligonucleotide with a second bridge oligonucleotide hybridization region 1428b of the bridge oligonucleotide hybridization regions of the interaction determination oligonucleotides; and ligating the interaction determination oligonucleotides 1428a, 1428b that are hybridized to the bridge oligonucleotide 1432 to generate a ligated interaction determination oligonucleotide.

[0725] In some embodiments, the at least one of the one or more cellular component targets (e.g., protein targets 1436a, 1436b) is on a cell surface. In some embodiments, the method can comprise: fixating the cell prior to contacting the cell with the first pair of interaction determination compositions 1404a, 1404b. In some embodiments, the method can comprise: removing unbound interaction determination compositions of the first pair of interaction determination compositions 1404a, 1404b. Removing the unbound interaction determination compositions can comprise washing the cell with a washing buffer. Removing the unbound interaction determination compositions can comprise selecting the cell using flow cytometry. In some embodiments, the method can comprise: lysing the cell.

[0726] In some embodiments, the interaction determination oligonucleotide is configured to be detachable or non-detachable from the protein binding reagent. The method can comprise: detaching the interaction determination oligonucleotide 1412a, 1412b or the ligated interaction determination oligonucleotide 1412 from the cellular component binding reagent. Detaching the interaction determination oligonucleotide 1412a, 1412b can comprise detaching the interaction determination oligonucleotide 1412a, 1412b from the cellular component binding reagent by UV photocleaving, chemical treatment, heating, enzyme treatment, or any combination thereof. The interaction determination oligonucleotide 1412a, 1412b may be not homologous to genomic sequences of the cell. The interaction

determination oligonucleotide 1412a, 1412b can be homologous to genomic sequences of a species.

[0727] In some embodiments, the interaction determination oligonucleotide 1412b of the one of the first pair of interaction determination compositions 1404a, 1404b comprises a sequence 1420 complementary to the capture sequence 1444. The capture sequence 1444 can comprise a poly(dT) region. The sequence 1420 of the interaction determination oligonucleotide complementary to the capture sequence 1444 can comprise a poly(dA) region. In some embodiments, the interaction determination oligonucleotide 1412a, 1412b comprises a second barcode sequence. The interaction determination oligonucleotide 1412a of the other of the first pair of interaction identification compositions 1404a, 1404b can comprise a binding site for a universal primer 1416. The interaction determination oligonucleotide 1404a, 1404b can be associated with a detectable moiety.

[0728] In some embodiments, the cellular component binding reagent can be associated with two or more interaction determination oligonucleotides 1412a, 1412b with the same or different interaction determination sequences 1428a, 1428b. In some embodiments, one of the plurality of interaction determination compositions 1404a, 1404b comprises a second protein binding reagent not associated with the interaction determination oligonucleotide. The cellular component binding reagent and the second cellular component binding reagent can be identical. The cellular component binding reagent can be associated with a detectable moiety. The cellular component binding reagent 1404a, 1404b can be associated with two or more interaction determination oligonucleotides 1412a, 1412b with an identical sequence.

[0729] In some embodiments, the method comprises: contacting two or more cells with the first pair of interaction determination compositions 1404a, 1404b. Each of the two or more cells can comprise the first and the second cellular component targets (e.g., the first and the second protein targets 1436a, 1436b). At least one of the two or more cells can comprise a single cell.

[0730] Referring to FIG. 14B, the barcode 1440 comprises a cell label sequence 1448, a binding site for a universal primer, or any combination thereof. At least two barcodes of the plurality of barcodes can comprise an identical cell label sequence 1448.

[0731] In some embodiments, the cellular component target is, or comprises, an

extracellular protein, an intracellular protein, or any combination thereof. The cellular component target can be, or comprise, a lipid, a carbohydrate, or any combination thereof. The cellular component target can be selected from a group comprising 10-100 different cellular component targets.

**[0732]** In some embodiments, the plurality of barcodes is associated with a particle (e.g., a bead 1438). At least one barcode the plurality of barcodes can be immobilized on the particle. At least one barcode of the plurality of barcodes can be partially immobilized on the particle. At least one barcode of the plurality of barcodes can be enclosed in the particle. At least one barcode of the plurality of barcodes can be partially enclosed in the particle. The particle can be disruptable. In some embodiments, the barcodes of the particle comprise barcode sequences selected from at least 1000, 10000, or more different barcode sequences

**[0733]** Referring to FIG. 14C, barcoding the interaction determination oligonucleotides 1412 using the plurality of barcodes 1440 comprises: contacting the plurality of barcodes 1440 with the interaction determination oligonucleotides 1412 to generate barcodes hybridized to the interaction determination oligonucleotides; and extending the barcodes hybridized to the interaction determination oligonucleotides to generate the plurality of barcoded interaction determination oligonucleotides 1456. Extending the barcodes 1440 can comprise extending the barcodes 1440 using a DNA polymerase to generate the plurality of barcoded interaction determination oligonucleotides 1456. Extending the barcodes 1440 can comprise extending the barcodes 1440 using a reverse transcriptase to generate the plurality of barcoded interaction determination oligonucleotides 1456. Extending the barcodes 14 can comprise extending the barcodes 14 using a Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase or a Taq DNA polymerase to generate the plurality of barcoded interaction determination oligonucleotides 1456. Extending the barcodes 14 can comprise displacing the bridge oligonucleotide 1432 from the ligated interaction determination oligonucleotide 1412.

**[0734]** Referring to FIGS. 14D and 14E, the method can comprise: amplifying the plurality of barcoded interaction determination oligonucleotides 1456 to produce a plurality of amplicons. Amplifying the plurality of barcoded interaction determination oligonucleotides 1456 can comprise amplifying, using polymerase chain reaction (PCR), at

least a portion of the barcode sequence 1448 and at least a portion of the interaction determination oligonucleotide 1428a, 1428b. Obtaining the sequencing data of the plurality of barcoded interaction determination oligonucleotides 1456 can comprise obtaining sequencing data of the plurality of amplicons. Obtaining the sequencing data can comprise sequencing at least a portion of the barcode sequence 1448 and at least a portion of the interaction determination oligonucleotide 1428a, 1428b. In some embodiments, obtaining sequencing data of the plurality of barcoded interaction determination oligonucleotides 1456 comprises obtaining partial and/or complete sequences of the plurality of barcoded interaction determination oligonucleotides 1456 (or a reverse, a complement, a reverse complement 1456r, or any combination thereof).

[0735] Referring to FIG. 14C, the plurality of barcodes 1440 can comprise a plurality of stochastic barcodes. The barcode sequence of each of the plurality of stochastic barcodes can comprise a molecular label sequence 1448. The molecular label sequences 1448 of at least two stochastic barcodes of the plurality of stochastic barcodes can comprise different sequences. Barcoding the interaction determination oligonucleotides 1412 using the plurality of barcodes 1440 to create the plurality of barcoded interaction determination oligonucleotides 1456 can comprise stochastically barcoding the interaction determination oligonucleotides 1412 using the plurality of stochastic barcodes to create a plurality of stochastically barcoded interaction determination oligonucleotides.

[0736] The method can also be used with other single cell RNA sequencing methods. For example, one single workflow can be used to determine mRNA expression levels, protein expression levels (or expression levels of cellular component targets), and/or protein-protein interactions (or interactions between cellular components). As a result, in a single experiment, data about mRNA expression, protein expression, and protein-protein interactions from in a single cell (or multiple single cells) can be obtained. In some embodiments, the method comprises: barcoding a plurality of targets (e.g., mRNA species of interest) of the cell using the plurality of barcodes 1440 to create a plurality of barcoded targets; and obtaining sequencing data of the barcoded targets. Barcoding the plurality of targets using the plurality of barcodes 1440 to create the plurality of barcoded targets can comprise: contacting copies of the targets with target-binding regions (e.g., the poly(dT) region 1444) of the barcodes 1440; and reverse transcribing the plurality targets using the

plurality of barcodes 1440 to create a plurality of reverse transcribed targets. The method can comprise: prior to obtaining the sequencing data of the plurality of barcoded targets, amplifying the barcoded targets to create a plurality of amplified barcoded targets (similar to amplifying the barcoded interaction determination oligonucleotides 1456 illustrated in FIGS. 14E and 14F). Amplifying the barcoded targets to generate the plurality of amplified barcoded targets can comprise: amplifying the barcoded targets by polymerase chain reaction (PCR). Barcoding the plurality of targets of the cell using the plurality of barcodes 1440 to create the plurality of barcoded targets can comprise stochastically barcoding the plurality of targets of the cell using the plurality of stochastic barcodes to create a plurality of stochastically barcoded targets.

[0737] Embodiments disclosed herein also include kits for identifying cellular component-cellular component interactions (e.g., protein-protein interactions). In some embodiments, the kit comprises: a first pair of interaction determination compositions, wherein each of the first pair of interaction determination compositions comprises a cellular component binding reagent associated with an interaction determination oligonucleotide, wherein the cellular component binding reagent of one of the first pair of interaction determination compositions is capable of specifically binding to a first cellular component target and a cellular component binding reagent of the other of the first pair of interaction determination compositions is capable of specifically binding to the second cellular component target, wherein the interaction determination oligonucleotide comprises an interaction determination sequence and a bridge oligonucleotide hybridization region, and wherein the interaction determination sequences of the first pair of interaction determination compositions comprise different sequences; and a plurality of bridge oligonucleotides each comprising two hybridization regions capable of specifically binding to the bridge oligonucleotide hybridization regions of the first pair of interaction determination compositions.

[0738] In some embodiments, the kit comprises: a second pair of interaction determination compositions, wherein each of the second pair of interaction determination compositions comprises a cellular component binding reagent associated with an interaction determination oligonucleotide, wherein the cellular component binding reagent of one of the second pair of interaction determination compositions is capable of specifically binding to a

third cellular component target and the cellular component binding reagent of the other of the second pair of interaction determination compositions is capable of specifically binding to a fourth cellular component target. In some embodiments, the kit comprises: three or more pairs of interaction determination compositions.

[0739] In some embodiments, the kit comprises: a plurality of barcodes, wherein each of the plurality of barcodes comprises a barcode sequence and a capture sequence. The plurality of barcodes can comprise a plurality of stochastic barcodes, wherein the barcode sequence of each of the plurality of stochastic barcodes comprises a barcode sequence (e.g., a molecular label sequence), wherein the barcode sequences of at least two stochastic barcodes of the plurality of stochastic barcodes comprise different sequences. In some embodiments, the plurality of barcodes is associated with a particle. At least one barcode of the plurality of barcodes can be immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or any combination thereof. The particle can be disruptable. The particle can comprise a bead.

[0740] In some embodiments, the kit comprises: a DNA polymerase. The kit can comprise a reverse transcriptase. The kit can comprise: a Moloney Murine Leukemia Virus (M-MLV) reverse transcriptase or a Taq DNA polymerase. In some embodiments, the method comprises a fixation agent (e.g., formalin, paraformaldehyde, glutaraldehyde/osmium tetroxide, Alcoholic fixatives, Hepes-glutamic acid buffer-mediated organic solvent protection effect (HOPE), Bouin solution, or any combination thereof).

### EXAMPLES

[0741] Some aspects of the embodiments discussed above are disclosed in further detail in the following examples, which are not in any way intended to limit the scope of the present disclosure.

#### Example 1

##### Oligonucleotides for Conjugation with Protein Binding Reagents

[0742] This example demonstrates designing of oligonucleotides that can be conjugated with protein binding reagents. The oligonucleotides can be used to determine protein expression and gene expression simultaneously. The oligonucleotides can also be used for sample indexing to determine cells of the same or different samples.

### 95mer Oligonucleotide Design

[0743] The following method was used to generate candidate oligonucleotide sequences and corresponding primer sequences for simultaneous determination of protein expression and gene expression or sample indexing.

[0744] 1. Sequence generation and elimination

[0745] The following process was used to generate candidate oligonucleotide sequences for simultaneous determination of protein expression and gene expression or sample indexing.

[0746] Step 1a. Randomly generate a number of candidate sequences (50000 sequences) with the desired length (45 bps).

[0747] Step 1b. Append the transcriptional regulator LSRR sequence to the 5' end of the sequences generated and a poly(A) sequence (25 bps) to the 3' end of the sequences generated.

[0748] Step 1c. Remove sequences generated and appended that do not have GC contents in the range of 40% to 50%.

[0749] Step 1d. Remove remaining sequences with one or more hairpin structures each.

[0750] The number of remaining candidate oligonucleotide sequences was 423.

[0751] 2. Primer design

[0752] The following method was used to design primers for the remaining 423 candidate oligonucleotide sequences.

[0753] 2.1 N1 Primer: Use the universal N1 sequence: 5'-GTTGTCAAGATGCTACCGTTCAGAG-3' (LSRR sequence; SEQ ID NO. 5) as the N1 primer.

[0754] 2.2 N2 Primer (for amplifying specific sample index oligonucleotides; e.g., N2 primer in FIGS. 15B-15D):

[0755] 2.2a. Remove candidate N2 primers that do not start downstream of the N1 sequence.

[0756] 2.2b. Remove candidate N2 primers that overlap in the last 35 bps of the candidate oligonucleotide sequence.

[0757] 2.2c. Remove the primer candidates that are aligned to the transcriptome

of the species of cells being studied using the oligonucleotides (e.g., the human transcriptome or the mouse transcriptome).

[0758] 2.2d. Use the ILR2 sequence as the default control (ACACGACGCTCTTCCGATCT; SEQ ID NO. 8) to minimize or avoid primer-primer interactions.

[0759] Of the 423 candidate oligonucleotide sequences, N2 primers for 390 candidates were designed.

[0760] 3. Filtering

[0761] The following process was used to filter the remaining 390 candidate primer sequences.

[0762] 3a. Eliminate any candidate oligonucleotide sequence with a random sequence ending in As (i.e., the effective length of the poly(A) sequence is greater than 25 bps) to keep poly(A) tail the same length for all barcodes.

[0763] 3b. Eliminate any candidate oligonucleotide sequences with 4 or more consecutive Gs (> 3Gs) because of extra cost and potentially lower yield in oligo synthesis of runs of Gs.

[0764] FIG. 15A shows a non-limiting exemplary candidate oligonucleotide sequence generated using the method above.

200mer Oligonucleotide Design

[0765] The following method was used to generate candidate oligonucleotide sequences and corresponding primer sequences for simultaneous determination of protein expression and gene expression and sample indexing.

[0766] 1. Sequence generation and elimination

[0767] The following was used to generate candidate oligonucleotide sequences for simultaneous determination of protein expression and gene expression and sample indexing.

[0768] 1a. Randomly generate a number of candidate sequences (100000 sequences) with the desired length (128 bps).

[0769] 1b. Append the transcriptional regulator LSRR sequence and an additional anchor sequence that is non-human, non-mouse to the 5' end of the sequences generated and a poly(A) sequence (25 bps) to the 3' end of the sequences generated.



[0770] 1c. Remove sequences generated and appended that do not have GC contents in the range of 40% to 50%.

[0771] 1d. Sort remaining candidate oligonucleotide sequences based on hairpin structure scores.

[0772] 1e. Select 1000 remaining candidate oligonucleotide sequences with the lowest hairpin scores.

[0773] 2. Primer design

[0774] The following method was used to design primers for 400 candidate oligonucleotide sequences with the lowest hairpin scores.

[0775] 2.1 N1 Primer: Use the universal N1 sequence: 5'-GTTGTCAAGATGCTACCGTTCAGAG-3' (LSRR sequence; SEQ ID NO. 5) as the N1 primer.

[0776] 2.2 N2 Primer (for amplifying specific sample index oligonucleotides; e.g., N2 primer in FIGS. 15B and 15C):

[0777] 2.2a. Remove candidate N2 primers that do not start 23 nts downstream of the N1 sequence (The anchor sequence was universal across all candidate oligonucleotide sequences).

[0778] 2.2b. Remove candidate N2 primers that overlap in the last 100 bps of the target sequence. The resulting primer candidates can be between the 48th nucleotide and 100th nucleotide of the target sequence.

[0779] 2.2c. Remove the primer candidates that are aligned to the transcriptome of the species of cells being studied using the oligonucleotides (e.g., the human transcriptome or the mouse transcriptome).

[0780] 2.2d. Use the ILR2 sequence, 5'-ACACGACGCTCTTCCGATCT-3' (SEQ ID NO. 8) as the default control to minimize or avoid primer-primer interactions.

[0781] 2.2e. Remove N2 primer candidates that overlap in the last 100 bps of the target sequence.

[0782] Of the 400 candidate oligonucleotide sequences, N2 primers for 392 candidates were designed.

[0783] 3. Filtering

[0784] The following was used to filter the remaining 392 candidate primer

sequences.

[0785] 3a. Eliminate any candidate oligonucleotide sequence with a random sequence ending in As (i.e., the effective length of the poly(A) sequence is greater than 25 bps) to keep poly(A) tail the same length for all barcodes.

[0786] 3b. Eliminate any candidate oligonucleotide sequences with 4 or more consecutive Gs (> 3Gs) because of extra cost and potentially lower yield in oligo synthesis of runs of Gs.

[0787] FIG. 15B shows a non-limiting exemplary candidate oligonucleotide sequence generated using the method above. The nested N2 primer shown in FIG. 15B can bind to the antibody or sample specific sequence for targeted amplification. FIG. 15C shows the same non-limiting exemplary candidate oligonucleotide sequence with a nested universal N2 primer that corresponds to the anchor sequence for targeted amplification. FIG. 15D shows the same non-limiting exemplary candidate oligonucleotide sequence with a N2 primer for one step targeted amplification.

[0788] Altogether, these data indicate that oligonucleotide sequences of different lengths can be designed for simultaneous determination of protein expression and gene expression or sample indexing. The oligonucleotide sequences can include a universal primer sequence, an antibody specific oligonucleotide sequence or a sample indexing sequence, and a poly(A) sequence.

#### Example 2

##### Comparison of Detection Sensitivity with Different Antibody:Oligonucleotide Ratios

[0789] This example demonstrates detection sensitivity of CD4 protein using an anti-CD4 antibody conjugated with 1, 2, or 3 oligonucleotides.

[0790] Frozen peripheral blood mononuclear cells (PBMCs) of a subject were thawed. The thawed PBMCs were stained with three types of anti-CD4 antibody at 0.06 µg/100 µl (1:333 dilution of oligonucleotide-conjugated antibody stocks) at room temperature for 20 minutes. Each type of the types of anti-CD4 antibody was conjugated with 1, 2, or 3 oligonucleotides ("antibody oligonucleotides"). The sequence of the antibody oligonucleotide is shown in FIG. 16. The cells were washed to remove unbound anti-CD4 antibodies. The cells were stained with Calcein AM (BD (Franklin Lake, New Jersey)) and Draq7<sup>TM</sup> (Abcam (Cambridge, United Kingdom)) for sorting with flow cytometry to obtain

live cells. The cells were washed to remove excess Calcein AM and Draq7<sup>TM</sup>. Single cells stained with Calcein AM (live cells) and not Draq7<sup>TM</sup> (cells that were not dead or permeabilized) were sorted, using flow cytometry, into a BD Rhapsody<sup>TM</sup> cartridge.

[0791] Of the wells containing a single cell and a bead, 3500 of the single cells in the wells were lysed in a lysis buffer with 5 mM DTT. The CD4 mRNA expression profile was determined using BD Rhapsody<sup>TM</sup> beads. The CD4 protein expression profile was determined using BD Rhapsody<sup>TM</sup> beads and the antibody oligonucleotides. Briefly, the mRNA molecules were released after cell lysis. The Rhapsody<sup>TM</sup> beads were associated with stochastic barcodes each containing a molecular label, a cell label, and a polyT region. The poly(A) regions of the mRNA molecules released from the lysed cells hybridized to the polyT regions of the stochastic barcodes. The poly(A) regions of the oligonucleotides hybridized to the polyT regions of the stochastic barcodes. The mRNA molecules were reverse transcribed using the stochastic barcodes. The antibody oligonucleotides were replicated using the stochastic barcodes. The reverse transcription and replication occurred in one sample aliquot at the same time.

[0792] The reverse transcribed products and replicated products were PCR amplified for 15 cycles at 60 degrees annealing temperature using primers for determining the mRNA expression profiles of 488 blood panel genes, using blood panel N1 primers, and the expression profile of CD4 protein, using the antibody oligonucleotide N1 primers ("PCR 1"). Excess stochastic barcodes were removed with Ampure cleanup. The products from PCR1 were divided into two aliquots, one aliquot for determining the mRNA expression profiles of the 488 blood panel genes, using the blood panel N2 primers, and one aliquot for determining the expression profile of CD4 protein, using the antibody oligonucleotide N2 primers ("PCR 2"). Both aliquots were PCR amplified for 15 cycles at 60 degrees annealing temperature. The expression of CD4 protein in the lysed cells was determined based on the antibody oligonucleotides as illustrated in FIG. 16 ("PCR 2"). Sequencing data was obtained and analyzed after sequencing adaptor ligation ("PCR 3"). Cell types were determined based on the expression profiles of the 488 blood panel genes.

[0793] FIGS. 17A-17F are non-limiting exemplary t-Distributed Stochastic Neighbor Embedding (tSNE) projection plots showing results of using oligonucleotide-conjugated antibodies to measure CD4 protein expression and gene expression

simultaneously in a high throughput manner. CD4 protein expression was distinctly and robustly detected in CD4 expressing cell types (e.g., CD4 T cells) with anti-CD4 antibodies conjugated to 1, 2, or 3 antibody oligonucleotides (FIGS. 17B, 17D, and 17F respectively). FIGS. 18A-18F are non-limiting exemplary bar charts showing the expressions of CD4 mRNA and protein in CD4 T cells (high CD4 expression), CD8 T cells (minimal CD4 expression), and Myeloid cells (some CD4 expression). With similar sequencing depth, detection sensitivity for CD4 protein level increased with higher ratios of antibody:oligonucleotide, with the 1:3 ratio performing better than the 1:1 and 1:2 ratios (FIG. 19). The expression of CD4 protein on cell surface of cells sorted using flow cytometry was confirmed using FlowJo (FlowJo (Ashland, OR)) as shown in FIGS. 20A-20D. FIGS. 21A-21F are non-limiting exemplary bar charts showing the expressions of CD4 mRNA and CD4 protein in CD4 T cells, CD8 T cells, and Myeloid cells of two samples. The second sample was prepared using two different sample preparation protocols. FIG. 22 is a non-limiting exemplary bar chart showing detection sensitivity for CD4 protein level determined using different sample preparation protocols with an antibody:oligonucleotide ratio of 1:3.

[0794] Altogether, these data indicate that CD4 protein expression can be distinctly and robustly detected based on oligonucleotide-conjugated with anti-CD4 antibodies. Detection sensitivity for CD4 protein level can increase with higher antibody:oligonucleotide ratios.

### Example 3

#### Sample Indexing

[0795] This example demonstrates identifying cells of different samples using sample indexing with high labeling efficiency and low non-specific labeling or spill over. This example also demonstrates the effects of the lengths of the sample indexing oligonucleotides and the cleavability of sample indexing oligonucleotides on sample indexing.

[0796] FIG. 23 shows a non-limiting exemplary experimental design for performing sample indexing and determining the effects of the lengths of the sample indexing oligonucleotides and the cleavability of sample indexing oligonucleotides on sample indexing. The left column in FIG. 23 shows that Jurkat cells were labeled with anti-

CD147 antibodies with different sample indexing oligonucleotides to determine the effects of the lengths of the sample indexing oligonucleotides and the cleavability of sample indexing oligonucleotides on sample indexing. The oligonucleotide-conjugated with the anti-CD147 antibodies were 200 nucleotides in length that were cleavable from the antibody (T-linker), 95 nucleotides in length that were cleavable (T-linker), and 95 nucleotides in length that were not cleavable. After further sample preparation, the effects of the lengths of the sample indexing oligonucleotides and the cleavability of sample indexing oligonucleotides on sample indexing were determined using a Rhapsody™ flowcell (labeled “Flowcell 1” in FIG. 23, abbreviated as “FC1”). Sample preparation included Calcein AM and Draq7™ staining, cell sorting using flow cytometry, and stochastic barcoding.

[0797] The middle column in FIG. 23 shows that cells were labeled with anti-CD147 antibodies conjugated with three different sample indexing oligonucleotides that were cleavable (the sample indexing oligonucleotides, labeled “Short 1,” “Short 2,” and “Short 3” in FIG. 23, can be abbreviated as S1, S2, and S3). Ramos cells were labeled with two types of anti-CD147 antibodies conjugated with different sample indexing oligonucleotides that are cleavable (the sample indexing oligonucleotides are labeled “Short 1” and “Short 2” in FIG. 23). Jurkat cells were labeled with two types of anti-CD147 antibodies conjugated with different sample indexing oligonucleotides that are cleavable (the sample indexing oligonucleotides are labeled “Short 1” and “Short 3” in FIG. 23). The labeled cells were mixed in a 1:1 ratio, followed by further sample preparation and determination of the efficiency and specificity of sample indexing using a Rhapsody™ flowcell (labeled “Flowcell 2” in FIG. 23, abbreviated as “FC2”). The right column in FIG. 23 shows the control experiment where Jurkat cells and Ramos cells that were not labeled were mixed in a 1:1 ratio prior to further sample preparation and analysis using a Rhapsody™ flowcell (labeled as “Flowcell 3” in FIG. 23, abbreviated as “FC3”).

[0798] The cells were stained with antibodies conjugated with sample indexing oligonucleotides a 1:3 antibody:oligonucleotide ratio at room temperature for 20 minutes. The antibodies were diluted from oligonucleotide-conjugated antibody stocks at 1:20 dilution in 100 µl using a pH 7.5 diluent. The cells were washed to remove unbound anti-CD147 antibodies. The cells were stained with Calcein AM (staining for live cells) and Draq7™ (staining for cells that were dead or permeabilized)) for sorting with flow cytometry to obtain

live cells. The cells were washed to remove excess Calcein AM and Draq7<sup>TM</sup>. Single cells stained with Calcein AM and not Draq7<sup>TM</sup> were sorted, using flow cytometry, into a BD Rhapsody<sup>TM</sup> cartridge containing a flowcell.

[0799] Of the wells containing a single cell and a bead, 1000 of the single cells in the wells were lysed in a lysis buffer. The CD147 mRNA expression profile was determined using BD Rhapsody<sup>TM</sup> beads. Briefly, the mRNA molecules were released after cell lysis. The Rhapsody<sup>TM</sup> beads were associated with stochastic barcodes each containing a molecular label, a cell label, and a polyT region. The poly(A) regions of the mRNA molecules released from the lysed cells hybridized to the polyT regions of the stochastic barcodes. The poly(A) regions of the sample indexing oligonucleotides, cleaved from the antibodies if cleavable, hybridized to the polyT regions of the stochastic barcodes. The mRNA molecules were reverse transcribed using the stochastic barcodes. The antibody oligonucleotides were replicated using the stochastic barcodes. The reverse transcription and replication occurred in one sample aliquot at the same time for 15 cycles at 60 degrees annealing temperature.

[0800] The reverse transcribed products and replicated products were PCR amplified for 15 cycles at 60 degrees annealing temperature using primers for determining the mRNA expression profiles of 488 blood panel genes, using blood panel N1 primers, and the expression of CD147 protein, using the sample indexing oligonucleotide N1 primers ("PCR 1"). Excess primers were removed with Ampure cleanup. The products from PCR1 were further PCR amplified ("PCR 2") for 15 cycles at 60 degrees annealing temperature using blood panel N2 primers and sample indexing oligonucleotide N1 primers with a flanking sequence for adaptor ligation. Sequencing data was obtained and analyzed after sequencing adaptor ligation ("PCR 3"). Cell types were determined based on the expression profiles of the 488 blood panel genes.

[0801] Table 1 shows metrics of the sequencing data obtained using the experimental design illustrated in FIG. 23. Three types of anti-CD147 antibody, conjugated with any one of the three types of sample indexing oligonucleotides (cleavable 95mer, non-cleavable 95mer, and cleavable 200mer shown at the left column in FIG. 23), were successfully used for sample indexing. More than 11% of total numbers of reads in the sequencing data were attributed to the sample indexing oligonucleotides.

**Table 1.** Sequencing metrics of sample indexing oligonucleotides.

|                            | FC1 (Oligos)     |                      |                   | FC2 (sample indexing) |           |      |
|----------------------------|------------------|----------------------|-------------------|-----------------------|-----------|------|
| Sample                     | Jurkat (J)       |                      |                   | J & R                 | Ramos (R) | J    |
| Oligo type                 | 95mer, cleavable | 95mer, non-cleavable | 200mer, cleavable | 95mer, cleavable      |           |      |
| % of total reads           | 11.4             | 22.4                 | 11.8              | 16.7                  | 12.9      | 11   |
| % of total molecules       | 29               | 41.4                 | 7.5               | 28.1                  | 20.3      | 16.3 |
| Average # of mols per cell | 1043             | 1679                 | 303               | NA                    | NA        | NA   |
| Median # of mols per cell  | 895              | 1536                 | 256               | NA                    | NA        | NA   |
| Oligo RSEC seq depth       | 3.9              | 4.6                  | 11.54             | 1.74                  | 1.87      | 2    |
| Blood panel RSEC seq depth | 10.72            | 10.72                | 10.72             | 4.1                   | 4.1       | 4.1  |

[0802] Three types of anti-CD147 antibody, conjugated with any one of the three cleavable 95mers (labeled “Short 1,” “Short 2,” and “Short 3” at the middle column in FIG. 23), were successfully used for sample indexing to distinguish Ramos and Jurkat cells of different samples. More than 11% of the total number of reads in the sequencing data were attributed to the sample indexing oligonucleotides. The percentage was the highest (16.7%) for a sample containing both Jurkat and Ramos cells.

[0803] FIGS. 24A-24C are non-limiting exemplary tSNE plots of expression profiles of cells showing that the three types of anti-CD147 antibody conjugated with different sample indexing oligonucleotides (cleavable 95mer, non-cleavable 95mer, and cleavable 200mer) can be used for determining the protein expression level of CD147. The overlays of the CD147 expression, determined using sample indexing oligonucleotides, on tSNE projection plots of the expression profiles of the cells show that the CD147 expression patterns determined using different sample indexing oligonucleotides were similar. Detection of the three sample indexing oligonucleotides was 100%. FIG. 25 is a non-limiting exemplary tSNE plot with an overlay of GAPDH expression per cell. “Cells” at the top right corner of the plot were low in GAPDH, suggesting that they were not real cells. These “cells” were high in sample indexing oligonucleotides.

[0804] FIGS. 26A-26C are non-limiting exemplary histograms showing the number of molecules of sample indexing oligonucleotides detected using the three types of sample indexing oligonucleotides. The non-cleavable sample indexing oligonucleotide was the most sensitive in terms of the number of molecules detected.

[0805] FIGS. 27A-27C are plots and bar charts showing that CD147 expression was higher in dividing cells (determined using “Flowcell 1” in FIG. 23). CD147 expression was not uniform across cells in the sample of Jurkat cells. Cells can be classified based on the cell cycle gene AURKB. The mRNA expression of the AURKB gene correlated with the CD147 protein expression determined using sample indexing. FIG. 27C shows a comparison of CD147 expression, determined using the three sample indexing oligonucleotides, in AURKB+ and AURKB- cells.

[0806] FIGS. 28A-28C are non-limiting exemplary tSNE projection plots of expression profiles of cells, showing that sample indexing can be used to identify cells of different samples. Reads of sample indexing oligonucleotides in the sequencing data, containing expression profiles of the 488 blood panel genes, were filtered. The tSNE projection plots in FIGS. 28A-28C were generated using the filtered sequencing data. The cluster corresponding to Jurkat cells and the cluster corresponding to Ramos cells (determined based on the abundance of the sample indexing oligonucleotides labeled “Short 1” in FIG. 23) are clearly separated in the tSNE projection plot of FIG. 28A (similarly for FIGS. 28B-28C).

[0807] Sample indexing correctly matched the cell type. For example, FIG. 28B shows that Ramos cells, which were labeled with the sample indexing oligonucleotides labeled “Short 2” in FIG. 23, had high expression of CD147, while Jurkat cells which were not labeled had minimal expression of CD147. FIG. 28C shows that Jurkat cells, which were labeled with the sample indexing oligonucleotides labeled “Short 3” in FIG. 23, had high expression of CD147, while Ramos cells which were not labeled had low expression of CD147. The small doublet cluster at 3.6% frequency closely matched the expected doublet rate of 3.92% from image analysis of occupancy of wells.

[0808] FIGS. 29A-29C are non-limiting exemplary histograms of the sample indexing sequences per cell based on the numbers of molecules of the sample indexing oligonucleotides determined. The sequencing data was corrected with distribution based error correction (DBEC) with a cutoff of 250. DBEC has been described in U.S. Patent Application Number 15/605874, filed on May 25, 2017, the content of which is incorporated herein by reference in its entirety. After DBEC, there seemed to be a clear “noise” peak in the 100s that was distinguishable from the “signal” peak in the 1000s (FIGS. 29B-29C).



FIGS. 29B-29C also show that the detection of sample indexing oligonucleotides to be in the high 100s after a noise subtraction of 250.

[0809] FIGS. 30A-30D are non-limiting exemplary plots comparing annotations of cell types determined based on mRNA expression of CD3D (for Jurkat cells) and JCHAIN (for Ramos cells) and sample indexing of Jurkat and Ramos cells. FIG. 30A is a non-limiting exemplary histogram showing clear separation of a signal peak of JCHAIN expression and noise signals with low numbers of molecules of sample indexing oligonucleotides per cell. FIG. 30B is a non-limiting exemplary histogram showing clear separation of a signal peak of Ramos expression and noise signals with low numbers of molecules of sample indexing oligonucleotides per cell. Doublets were identified as cells with more than one sample indexing sequence with more than 250 counts. In FIG. 30C, Ramos cells were identified using the mRNA expression of JCHAIN, and Jurkat cells were identified using the mRNA expression of CD3D.

[0810] Ramos cells were labeled with two types of anti-CD147 antibodies conjugated with different sample indexing oligonucleotides that are cleavable (the sample indexing oligonucleotides are labeled “Short 1” and “Short 2” in FIG. 23). Jurkat cells were labeled with two types of anti-CD147 antibodies conjugated with different sample indexing oligonucleotides that are cleavable (the sample indexing oligonucleotides are labeled “Short 1” and “Short 3” in FIG. 23). In FIG. 30D, the identity of the Ramos cells were identified based on the numbers of counts of the “Short 2” sample indexing oligonucleotide (greater than 250) in the sequencing data, and the identity of the Jurkat cells were identified based on the numbers of counts of the “Short 3” sample indexing oligonucleotide (greater than 250) in the sequencing data.

[0811] FIGS. 31A-31C are non-limiting tSNE projection plots of the mRNA expression profiles of Jurkat and Ramos cells with overlays of the mRNA expressions of CD3D and JCHAIN (FIG. 31A), JCHAIN (FIG. 31B), and CD3D (FIG. 31C). The mRNA expression of JCHAIN was higher in Ramos cells (FIG. 31B) and minimal in Jurkat cells (FIG. 31C) as expected. The mRNA expression of CD3D was higher in Jurkat cells (FIG. 31C) and minimal in Ramos cells (FIG. 31B).

[0812] The comparisons of FIGS. 28B-28C with FIGS. 31B-31C revealed the high performance of sample indexing in distinguishing cells of different samples. The

performance of sample indexing in determining sample origin is shown in FIG. 32. FIG. 32 is a non-limiting exemplary tSNE projection plot of expression profiles of Jurkat and Ramos cells with an overlay of the cell types determined using sample indexing with a DBEC cutoff of 250. Table 2 is a summary of the cell types determined using sample indexing shown in FIG. 32. Table 3 shows the sensitivity and specificity of sample indexing, with doublets and undefined cells in Table 2 excluded, in determining sample origin.

**Table 2.** Summary of sample origin determination using sample indexing.

|                 | S2  | S2+S3 | S3  | not labeled |
|-----------------|-----|-------|-----|-------------|
| <b>Doublets</b> | 0   | 36    | 1   | 0           |
| <b>Jurkat</b>   | 0   | 5     | 658 | 12          |
| <b>Ramos</b>    | 833 | 4     | 0   | 0           |
| <b>Unknown</b>  | 30  | 0     | 17  | 0           |

**Table 3.** Specificity and sensitivity of sample indexing in determining sample origin.

|                    | Specificity          | Sensitivity          |
|--------------------|----------------------|----------------------|
| <b>S2 (Ramos)</b>  | 99.5% (833/(833+4))  | 100% (833/(833+0))   |
| <b>S3 (Jurkat)</b> | 99.25% (658/(658+4)) | 98.2% (658/(658+12)) |

[0813] FIGS. 33A-33C are non-limiting exemplary bar charts of the numbers of molecules of sample indexing oligonucleotides per cell for Ramos & Jurkat cells (FIG. 33), Ramos cells (FIG. 33B), and Jurkat cells (FIG. 33C) that were not labeled or labeled with “Short 3” sample indexing oligonucleotides, “Short 2” & “Short 3” sample indexing oligonucleotides, and “Short 2” sample indexing oligonucleotides. Less than 1% of single cells were labeled with both the “Short 2” and “Short 3” sample indexing oligonucleotides (FIGS. 34A-34C).

[0814] FIGS. 35A-35C are non-limiting exemplary tSNE plots showing batch effects on expression profiles of Jurkat and Ramos cells between samples prepared using different flowcells as outlined in FIG. 23. FIG. 35A shows a non-limiting exemplary tSNE projection plot of expression profiles of Jurkat and Ramos cells with an overlay of different sample preparations of the cells. The variations in mRNA expression profiles in the 488 blood panel genes (FIG. 35B) and the CD3D gene (FIG. 35C) between different sample preparations may be a result of batch effect due to different sequencing depths of the samples prepared using the different flowcells.

[0815] Altogether, these data indicate that all three versions of sample indexing oligonucleotides (95mer cleavable, 95mer non-cleavable, and 200mer cleavable) can be used for sample indexing, with the 95mer non-cleavable having the highest efficiency. Sample indexing for determining sample origin can have high specificity and sensitivity.

#### Example 4

##### Hot:Cold Antibody Titration

[0816] This example demonstrates determining a ratio of oligonucleotide-conjugated antibodies (“hot antibodies”) and antibodies not conjugated with oligonucleotides (“cold antibody”) such that the antibody oligonucleotides account for a desired percentage (e.g., 2%) of total reads in sequencing data.

[0817] An anti-CD147 antibody stock was diluted at 1:20, 1:100, 1:200, 1:300, 1:400, 1:600, 1:800, and 1:1000 dilutions with PE buffer. Around 150000 Jurkat cells in 100 µl of staining buffer (FBS) (BD (Franklin Lake, New Jersey)) were stained at various antibody dilutions for 20 minutes at room temperature. After staining, the cells were washed once with 500 µl of staining buffer and resuspended in 200 µl for measurement of fluorescence intensity. Muse™ Autophagy LC3-antibody (EMD Millipore (Billerica, MA)) was used to detect the anti-CD147 antibody bound to the Jurkat cells. The fluorescence intensities from cells stained at various anti-CD147 antibody dilutions or cells not stained were determined and compared to determine an optimal dilution for the antibody (FIGS. 36A1-36A9, 36B and 36C). Fluorescence intensity decreased with higher dilution. More than 99% of the cells were stained with a dilution ratio of 1:800. Fluorescence signals began to drop out at 1:800. Cells were stained to saturation up to a dilution ratio of 1:200. Cells were stained close to saturation up to a dilution ratio of 1:400.

[0818] FIG. 37 shows a non-limiting exemplary experimental design for determining a staining concentration of oligonucleotide-conjugated antibodies such that the antibody oligonucleotides account for a desired percentage of total reads in sequencing data. An anti-CD147 antibody was conjugated with a cleavable 95mer antibody oligonucleotide at an antibody:oligonucleotide ratio of 1:3 (“hot antibody”). The hot antibody was diluted using a pH 7.5 diluent at a 1:100 ratio or a 1:800 ratio. A mixture of 10% hot antibody:90% cold antibody was prepared using 9 µl of cold anti-CD147 antibody and 1 µl of the hot antibody. A mixture of 1% hot antibody:90% cold antibody was prepared using 9 µl of the

cold anti-CD147 antibody and 1  $\mu$ l of the mixture of 10% hot antibody:90% cold antibody.

[0819] Thawed peripheral blood mononuclear cells (PBMCs) with around 0.5 million cells were stained in 100  $\mu$ l of staining buffer (FBS) with the 1:100 diluted stock with 100% hot antibody (1% of the stock hot antibody), the mixture of 10% hot antibody:90% cold antibody (0.1% of the stock hot antibody), the mixture of 1% hot antibody:99% cold antibody (0.01% of the stock hot antibody), and the 1:800 diluted stock with 100% hot antibody (0.0125% of the stock hot antibody). After staining, the cells were washed to remove unbound antibody molecules. The cells were stained with Calcein AM and Draq7<sup>TM</sup> for sorting with flow cytometry to obtain live cells. The cells were washed to remove excess Calcein AM and Draq7<sup>TM</sup>. Single cells stained with Calcein AM (live cells) and not Draq7<sup>TM</sup> (cells that were not dead or permeabilized) were sorted, using flow cytometry, into a BD Rhapsody<sup>TM</sup> cartridge.

[0820] Of the wells containing a single cell and a bead, 1000 of the single cells in the wells were lysed in a lysis buffer. For each single cell, the mRNA molecules were reverse transcribed and the antibody oligonucleotides were replicated using stochastic barcodes conjugated with a bead for the cell. The samples after reverse transcription and replication were PCR amplified for 15 cycles at 60 degrees annealing temperature using primers for determining the mRNA expression profiles of 488 blood panel genes, using blood panel N1 primers, and the expression of CD147 protein, using the sample indexing oligonucleotide N1 primers ("PCR 1"). Excess primers were removed with Ampure cleanup. The products from PCR1 were further PCR amplified ("PCR 2") for 15 cycles at 60 degrees annealing temperature using blood panel N2 primers and sample indexing oligonucleotide N1 primers with a flanking sequence for adaptor ligation. Sequencing data was obtained and analyzed after sequencing adaptor ligation ("PCR 3").

[0821] FIGS. 38A-38D are non-limiting exemplary bioanalyzer traces showing peaks (indicated by arrows) consistent with the expected size of the antibody oligonucleotide. The antibody oligonucleotide peaks decreased as the hot antibody was titrated with the cold antibody.

[0822] Table 4 is a summary of sequencing data metrics. By staining the cells with the mixture of 1% hot antibody:99% cold antibody prepared using the 1:100 diluted stock, the antibody oligonucleotides accounted for 2.4% of the total raw reads in the

sequencing data. However, as shown in FIGS. 39A1-39A3 and 39B1-39B3 and FIGS. 40B, 41B, and 42B, a distribution histogram of the numbers of molecules of antibody oligonucleotides detected after recursive substitution error correction (RSEC) or distribution-based error correction (DBEC) did not include a clear signal peak if the cells were stained with the mixture of 1% hot antibody:99% cold antibody prepared using the 1:100 diluted stock. RSEC has been described in U.S. Patent Application Number 15/605874, filed on May 25, 2017, the content of which is incorporated herein by reference in its entirety.

[0823] FIGS. 40A-40C are non-limiting exemplary plots showing that oligonucleotide-conjugated anti-CD147 antibody molecules can be used to label various cell types. The cell types were determined using the expression profiles of 488 genes in a blood panel (FIG. 40A). The cells were stained with a mixture of 10% hot antibody:90% cold antibody prepared using a 1:100 diluted stock, resulting in a clear signal in a histogram showing the numbers of molecules of antibody oligonucleotides detected (FIG. 40B). The labeling of the various cell types by the antibody oligonucleotide is shown in FIG. 40C.

[0824] FIGS. 41A-41C are non-limiting exemplary plots showing that oligonucleotide-conjugated anti-CD147 antibodies can be used to label various cell types. The cell types were determined using the expression profiles of 488 genes in a blood panel (FIG. 41A). The cells were stained with a mixture of 1% hot antibody:99% cold antibody prepared using a 1:100 diluted stock, resulting in no clear signal in a histogram showing the numbers of molecules of antibody oligonucleotides detected (FIG. 41B). The labeling of the various cell types by the antibody oligonucleotide is shown in FIG. 41C.

[0825] FIGS. 42A-42C are non-limiting exemplary plots showing that oligonucleotide-conjugated anti-CD147 antibody molecules can be used to label various cell types. The cell types were determined using the expression profiles of 488 genes in a blood panel (FIG. 42A). The cells were stained with a 1:800 diluted stock, resulting in a clear signal in a histogram showing the numbers of molecules of antibody oligonucleotides detected (FIG. 42B). The labeling of the various cell types by the antibody oligonucleotide is shown in FIG. 42C.

**Table 4.** Summary of sequencing data metrics.

| Sample | FC4 - 1:800<br>Dilution, No Cold<br>Antibody | FC3 - 1:100<br>Dilution, 1:100<br>Cold Antibody | FC2 - 1:100<br>Dilution, 1:10 Cold<br>Antibody |
|--------|--|---|--|
|--------|--|---|--|

|   |                 |               |                 |
|---|-----------------|---------------|-----------------|
| <b>Total Raw Reads</b>                    | 31.3M           | 27.3M         | 29.2M           |
| <b>Total Raw Reads Assigned to Oligos</b> | 9161642 (29.2%) | 660044 (2.4%) | 4013438 (13.7%) |
| <b>Cell Detected</b>                      | 1010            | 983           | 907             |
| <b>RSEC Oligo MI</b>                      | 577110          | 20054         | 170742          |
| <b>DBEC Oligo MI</b>                      | 477216          | 9319          | 110629          |
| <b>% Q30</b>                              | 76.45           | 71.89         | 73.45           |
| <b>% assigned to cell labels</b>          | 84.63           | 79.47         | 82.58           |
| <b>% aligned uniquely to amplicons</b>    | 73.19           | 65.4          | 69.58           |
| <b>Mean raw seq depth</b>                 | 5.58            | 6.52          | 6.4             |
| <b>Mean RSEC seq depth</b>                | 8.69            | 10.41         | 10.25           |
| <b>Mean DBEC seq depth</b>                | 15.96           | 23.29         | 22.48           |
| <b>AbOligo RSEC seq depth</b>             | 8.08            | 12.9          | 11.2            |
| <b>AbOligo DBEC Seq depth</b>             | 12.4            | 30.9          | 21.2            |
| <b>Mean reads per cell</b>                | 11257           | 14414         | 15150           |
| <b>mean molecules per cell</b>            | 553.8           | 608.3         | 647.1           |
| <b>Median mols per cell</b>               | 246.5           | 279           | 278             |
| <b>No. of genes in panel</b>              | 489             | 489           | 489             |
| <b>Total genes detected</b>               | 438             | 439           | 441             |
| <b>Mean genes per cell</b>                | 68.39           | 73.6          | 73.12           |

[0826] Altogether, these data indicate that the ratio of oligonucleotide-conjugated antibodies (“hot antibodies”) and antibodies not conjugated with oligonucleotides (“cold antibody”) can be adjusted such that the antibody oligonucleotides account for a desired percentage of total reads in sequencing data and data representing signal antibody oligonucleotides is clearly separated from data representing noise antibody oligonucleotides.

#### Example 5

##### Normalization

[0827] This example demonstrates how normalization, using a mixture of oligonucleotide-conjugated antibodies (“hot antibodies”) and antibodies not conjugated with oligonucleotides (“cold antibody”), can result in the antibody oligonucleotides accounting for a desired percentage of total reads in sequencing data with a desired coverage, irrespective of

the abundance of the protein targets of the antibodies.

[0828] Table 5 shows a comparison of quantification of three cell surface markers of varying abundance in 10000 B cells using hot antibodies. Total number of reads required to resolve relative expression levels of the three cell surface markers was 47.52 million reads.

**Table 5.** Example protein quantification using hot antibodies.

| Antigen | Molecules per Cell | Relative abundance | Ratio of Hot:Cold Antibodies | Number of Molecules per Cell Detected by Sequencing | Number of reads given sequencing depth of 4 |
|---------|--------------------|--------------------|------------------------------|---|---|
| CD21    | 210000             | 105                | 1:0                          | 840   | 33.6M                                       |
| HLA-DR  | 85000              | 42.5               | 1:0                          | 340   | 13.6M                                       |
| CD40    | 2000               | 1                  | 1:0                          | 8   | 0.32M                                       |

**Table 6.** Example protein quantification using hot and cold antibodies.

| Antigen | Molecules per Cell | Relative abundance | Ratio of Hot:Cold Antibodies | Number of Molecules per Cell detected by sequencing | Number of reads given sequencing depth of 4 | Expected number of molecules based on Antibody ratio | Relative abundance by sequencing |
|---------|--------------------|--------------------|------------------------------|---|---|--|----------------------------------|
| CD21    | 210000             | 105                | 1:100                        | 8.3   | 0.33M                                       | $8.3 \times 100 = 830$                               | 103.75                           |
| HLA-DR  | 85000              | 42.5               | 1:40                         | 8.5   | 0.34M                                       | $8.5 \times 40 = 340$                                | 42.5                             |
| CD40    | 2000               | 1                  | 1:0                          | 8   | 0.32M                                       | $8 \times 1 = 8$                                     | 1                                |

[0829] Table 6 shows that the total number of reads required to resolve relative expression levels of the three cell surface markers was 1 million reads using mixtures of hot antibodies:cold antibodies. Also, only 2% of the number of reads, compared to the quantification result shown in Table 5 (1 million reads vs. 47.52 million reads), is needed to achieve optimal coverage (e.g., sequencing depth of 4) of all three protein markers when mixtures of hot antibodies:cold antibodies were used to quantify expression levels of the three cell surface markers. Normalizing high expressing protein molecules, using a mixture with higher percentage of cold antibodies, decreased tradeoffs between detection of low abundance proteins, number of parameters, and sequencing cost, making the assay more attractive as a tool.

[0830] Altogether, these data indicate that a desired number of total reads in sequencing data with a desired coverage can be achieved for protein targets (e.g., antigens) of different abundance using mixtures of hot antibodies:cold antibodies.

#### Example 6

##### Control Particles

[0831] This example demonstrates generating control particles comprising control particle oligonucleotides with different sequences and use of the functionalized control particles to determine capture efficiency.

##### Materials

[0832] BD CompBead Plus Anti-Mouse Ig (7.5um) Particles Set (51-9006274)

[0833] BD staining buffer (FBS)

##### Procedure

[0834] 1. Vortex BD CompBead Plus thoroughly before use (1 minute at least).

[0835] 2. Add 800uL of staining buffer to tube (Table 7 shows the composition of the staining buffer with CD147 conjugated to oligonucleotides with five sequences at different abundance. FIG. 43A is a plot showing the composition of the staining buffer.).

[0836] 3. Add 5 full drops (approximately 300uL) of CompBead Plus Anti-Mouse.

[0837] 4. Add 20uL of the staining cocktail below to the tube. Vortex.

[0838] 5. Incubate 30 minutes at room temperature away from light.

[0839] 6. Spin beads at 200g for 10 minutes.

[0840] 7. Remove supernatant carefully and resuspend with 1mL staining buffer.

[0841] 8. Spin beads at 200g for 10 minutes.

[0842] 9. Remove supernatant carefully and resuspend with 1mL staining buffer to generate the functionalized CompBead stock solution.

[0843] 10. Count beads.

**Table 7.** Staining Cocktail Composition

| Antibodies | Final % in Staining Buffer | Prior Dilution | Staining Solution (μl) |
|------------|----------------------------|----------------|------------------------|
| CD147-LZ15 | 1                          | 1:1            | 1                      |
| CD147-LZ16 | 0.2                        | 1:5            | 1                      |
| CD147-LZ17 | 0.1                        | 1:10           | 1                      |



|                        |      |       |    |
|------------------------|------|-------|----|
| <b>CD147-LZ18</b>      | 0.02 | 1:50  | 1  |
| <b>CD147-LZ19</b>      | 0.01 | 1:100 | 1  |
| <b>Staining Buffer</b> |      |       | 95 |

### Results

[0844] FIGS. 44A-44B are brightfield images of cells (FIG. 44A, white circles) and control particles (FIG. 44B, black circles) in a hemocytometer. FIGS. 45A-45B are phase contrast (FIG. 45A, 10X) and fluorescent (FIG. 45B, 10X) images of control particles bound to oligonucleotide-conjugated antibodies associated with fluorophores. Fluorescent microscope was used to determine that 5ul of the functionalized CompBead stock solution contained ~2000 cells (4% of total input) with ~ 400000 functionalized CompBeads made.

[0845] FIG. 46 is an image of a control particle showing cells and a particle being loaded into microwells of a cartridge. CompBeads can be used with regular Rhapsody™ experiments. 522 functionalized CompBeads were added into a plurality of cells. Of the 20000 cells (including control particles) sequences, 156 had a sum of all control particle oligonucleotides greater than 20. Thus, 156 control particles were sequences. FIG. 43B is a plot showing the number of control particle oligonucleotides with the five different control barcode sequences (LZ15-LZ19) correlated with their abundance in the staining buffer.

[0846] Altogether, these data show that particles (e.g., CompBead Plus) can be functionalized with oligonucleotides (e.g., control particle oligonucleotides). Functionalized particles can be used with single cell sequencing workflow to determine the number of particles captured and sequenced.

### Example 7

#### Antibody Cocktail for Sample Indexing

[0847] This example demonstrates using an antibody cocktail for sample indexing can increase labeling sensitivity.

[0848] Sample indexing can utilize oligonucleotide-conjugated antibodies against multiple protein targets to label samples. In this example, instead of using a single antibody, a cocktail of two antibodies were used instead. The cocktail of antibodies were labeled with the same sample indexing oligonucleotides. FIGS. 47A-47C are plots showing using an antibody cocktail for sample indexing can increase labeling sensitivity. FIGS. 47A-47C

show the labeling sensitivity of PBMCs with a CD147 antibody, a CD47 antibody, and both CD147 and CD47 antibodies. FIG. 47C shows that more cells were labeled when both CD147 and CD47 antibodies were used and better signal to noise separation was achieved. Table 8 shows increased labeling sensitivity when both CD147 and CD47 antibodies were used. The use of both CD147 and CD47 antibodies resulted in higher sensitivity for labeling heterogeneous sample types.

**Table 8.** Labeling sensitivity with CD147, CD47, or both CD147 and CD47 antibodies.

| Antibody   | Sensitivity |
|------------|-------------|
| CD147      | 86.2%       |
| CD47       | 82.5%       |
| CD147+CD47 | 92.6%       |

[0849] Altogether, the data show that using an antibody cocktail can increase labeling sensitivity. This may be because protein expression can vary between cell types and cell states, making finding a universal antibody that labels all cell types challenging. Using an antibody cocktail can allow for more sensitive and efficient labeling of more sample types. An antibody cocktail can also include a wider range of antibodies. Antibodies that label different sample types well can be pooled together to create a cocktail that sufficiently labels all cell types or cell types of interest.

#### Example 8

##### Presence of Multiplets in Datasets

[0850] This example describes that multiplets in sequencing data can be identified and removed by tagging cells with different sample indexing oligonucleotides.

[0851] FIG. 48 is a non-limiting exemplary plot showing that multiplets can be identified and removed from sequencing data using sample indexing. As shown in FIG. 48, the rate of multiplets in droplets and single cell expression data can be close to eight percent when a sample with single 20000 cells are partitioned into droplets using a droplet-based platform. The rate of multiplets in a microwell array and single cell expression data can be much lower (e.g., close to only four percent) when a sample with single 20000 cells are partitioned into wells of a Rhapsody<sup>TM</sup> (Becton, Dickinson and Company (Franklin Lakes, NJ)). Multiplets in sequencing data can be identified and removed using sample indexing

oligonucleotides with different sample indexing sequences. For example, when a sample with 20000 single cells is divided into two subpopulations, cells in each subpopulation are labeled with sample indexing oligonucleotides with an identical sample indexing sequence, and cells in different subpopulations are labeled with sample indexing oligonucleotides with two different sample indexing sequences, sequence data associated with multiplets can be identified and removed, with the resulting sequencing data containing around two % multiplets. With sample indexing oligonucleotides with eight different sample indexing sequences, sequencing data of 20000 single cells can include only around 0.5% multiplets. With a higher number of sample indexing sequences, the rate of multiplet in sequencing data can be even lower.

[0852] Thus, the rate of multiplets remaining in single cell sequencing data can be substantially lowered, or eliminated, by labeling cells with sample indexing oligonucleotides with different sample indexing sequences.

#### Example 9

##### Sample Indexing for Identifying Multiplets

[0853] This example demonstrates using sample tags to identify and remove multiplets in sequencing data.

[0854] Six tissues (bone marrow, fat (gonadal white adipose tissue (gWAT)), colon, liver, lung, and spleen) from two mice were obtained. CD45+ single cells from the isolated tissues were isolated and sorted using fluorescence-activated cell sorting (FACS) to create 12 samples. CD45+ single cells of the 12 samples were tagged with 12 different sample tags (referred to herein as sample indexing compositions) using BD™ Single-Cell Multiplexing Kit for RNA-Sequencing and loaded onto a Rhapsody™ cartridge. A sample tag was an antibody conjugated to an oligonucleotide (referred to herein as sample indexing oligonucleotides) that can be captured and amplified in a 3' RNA-seq assay (such as the BD Rhapsody™ assay). Such a sample tag had high specificity and sensitivity (> 99%). The mRNA molecules from the cells and the sample indexing oligonucleotides were captured using Rhapsody™ magnetic beads, and barcoded and amplified using Rhapsody™ Immune Response Panel-Mouse (Mouse). Expression profiles of the cells were determined and identified as singlet expression profiles and multiplet expression profiles using synthetic multiplet expression profiles.

[0855] FIG. 49 is a non-limiting exemplary tSNE projection plot of expression profiles of CD45+ single cells from 12 samples of six tissues from two mice identified as singlets or multiplets using sample tags. FIG. 49 shows the multiplet expression profiles as multiple clusters. Sample tag sequences (also referred to herein as sample indexing indices of sample indexing oligonucleotides) for tagging the different tissues were used to identify multiplet expression profiles each including the expression profiles of cells of two or more cell types or subtypes. Multiplets of CD45+ cells from two different tissues were identified based on the presence of the sample tag sequences in the sequence data obtained from the cells. The performance of using nucleic acid sample tags to index samples and identify multiplets was comparable to the performance of using synthetic multiplet expression profiles to identify multiplets. FIG. 49B is a non-limiting exemplary tSNE projection plot of expression profiles of CD45+ single cells of 12 samples of six tissues from two mice with multiplets identified using sample indexing oligonucleotides removed.

[0856] FIG. 50A is a non-limiting exemplary tSNE projection plot of expression profiles of CD45+ single cells from two mice with multiplets identified using sample indexing oligonucleotides removed. With multiplet expression profiles removed, the two mice, which were biological replicates, exhibited similar expression profiles as shown in FIGS. 50A-50C. The results of immune profiling of the six different tissues after removing multiplet expression profiles are shown in FIGS. 51A-51F. The different cell types (e.g., B cells, macrophages, etc.) in the six sample shown in FIGS. 51A-51F were identified based on the expression profiles of the cells.

[0857] Exemplary expression profiles of macrophages, T cells, and B cells from the six different tissues are shown in FIGS. 52A-52C. FIGS. 52A-52C show that the same cell types in different tissues exhibited unique expression profiles and colon immune populations were the most distinct out of the six tissues. FIGS. 53A-53D are non-limiting exemplary plots comparing macrophages in the colon and spleen. FIG. 53A is a non-limiting exemplary tSNE projection plot comparing the expression profiles of macrophages in the colon and spleen, showing that macrophages in the two tissues had distinct expression profiles. FIG. 53B is a non-limiting exemplary plot showing the expression profiles of macrophages in the colon and spleen had low correlation. The low correlation further supported that macrophages in the two tissues had distinct expression profiles shown in FIG.

53A. FIGS. 53C and 53D are non-limiting exemplary tSNE projection plots of expression profiles of the Areg gene and Rora gene in macrophages in the colon and spleen. These two genes were expressed in 5.2% and 9.5% of the macrophages in the colon and spleen. FIGS. 53A, 53C, and 53D show that these two genes were expressed only in macrophages in the colon. Single-cell profiling of more than 28000 mouse immune cells from six different tissues as shown in FIGS. 52A-52C and 53A-53D revealed tissue specific gene expression signatures for major immune cell types.

[0858] Altogether, these data show that sample tagging can allow pooling of multiple samples onto the same single-cell experiment. Sample Tagging can increase sample throughput, eliminate batch effect between samples due to technical errors and enable detection of multiplets.

#### Example 10

##### Sample Indexing and Expression Profiles

[0859] This example demonstrates using sample tagging of cells does not alter mRNA expression profiles of cells.

[0860] FIG. 54 is a non-limiting exemplary plot showing that tagging or staining cells with sample indexing compositions did not alter mRNA expression profiles in peripheral blood mononuclear cell (PBMCs). FIG. 54 shows an overlay of tSNE projection plots of expression profiles of PBMCs with or without sample tagging. The overlap shows that sample tagging did not alter mRNA expression profiles in PBMCs.

[0861] Altogether, the data show that sample tagging or staining with antibodies conjugated with oligonucleotides do not alter expression profiles of cells. Thus, sample tagging can allow pooling of multiple samples, increase sample throughput, eliminate batch effect between samples, and/or enable detection of multiplets without altering the expression profiles of the tagged cells.

#### Terminology

[0862] In at least some of the previously described embodiments, one or more elements used in an embodiment can interchangeably be used in another embodiment unless such a replacement is not technically feasible. It will be appreciated by those skilled in the art that various other omissions, additions and modifications may be made to the methods

and structures described above without departing from the scope of the claimed subject matter. All such modifications and changes are intended to fall within the scope of the subject matter, as defined by the appended claims.

[0863] With respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity. As used in this specification and the appended claims, the singular forms “a,” “an,” and “the” include plural references unless the context clearly dictates otherwise. Any reference to “or” herein is intended to encompass “and/or” unless otherwise stated.

[0864] It will be understood by those within the art that, in general, terms used herein, and especially in the appended claims (*e.g.*, bodies of the appended claims) are generally intended as “open” terms (*e.g.*, the term “including” should be interpreted as “including but not limited to,” the term “having” should be interpreted as “having at least,” the term “includes” should be interpreted as “includes but is not limited to,” etc.). It will be further understood by those within the art that if a specific number of an introduced claim recitation is intended, such an intent will be explicitly recited in the claim, and in the absence of such recitation no such intent is present. For example, as an aid to understanding, the following appended claims may contain usage of the introductory phrases “at least one” and “one or more” to introduce claim recitations. However, the use of such phrases should not be construed to imply that the introduction of a claim recitation by the indefinite articles “a” or “an” limits any particular claim containing such introduced claim recitation to embodiments containing only one such recitation, even when the same claim includes the introductory phrases “one or more” or “at least one” and indefinite articles such as “a” or “an” (*e.g.*, “a” and/or “an” should be interpreted to mean “at least one” or “one or more”); the same holds true for the use of definite articles used to introduce claim recitations. In addition, even if a specific number of an introduced claim recitation is explicitly recited, those skilled in the art will recognize that such recitation should be interpreted to mean at least the recited number (*e.g.*, the bare recitation of “two recitations,” without other modifiers, means at least two recitations, or two or more recitations). Furthermore, in those instances where a convention analogous to “at least one of A, B, and C, etc.” is used, in general such a construction is

intended in the sense one having skill in the art would understand the convention (*e.g.*, “a system having at least one of A, B, and C” would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). In those instances where a convention analogous to “at least one of A, B, or C, etc.” is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (*e.g.*, “a system having at least one of A, B, or C” would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms. For example, the phrase “A or B” will be understood to include the possibilities of “A” or “B” or “A and B.”

[0865] In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0866] As will be understood by one skilled in the art, for any and all purposes, such as in terms of providing a written description, all ranges disclosed herein also encompass any and all possible sub-ranges and combinations of sub-ranges thereof. Any listed range can be easily recognized as sufficiently describing and enabling the same range being broken down into at least equal halves, thirds, quarters, fifths, tenths, etc. As a non-limiting example, each range discussed herein can be readily broken down into a lower third, middle third and upper third, etc. As will also be understood by one skilled in the art all language such as “up to,” “at least,” “greater than,” “less than,” and the like include the number recited and refer to ranges which can be subsequently broken down into sub-ranges as discussed above. Finally, as will be understood by one skilled in the art, a range includes each individual member. Thus, for example, a group having 1-3 articles refers to groups having 1, 2, or 3 articles. Similarly, a group having 1-5 articles refers to groups having 1, 2, 3, 4, or 5 articles, and so forth.

[0867] While various aspects and embodiments have been disclosed herein, other

aspects and embodiments will be apparent to those skilled in the art. The various aspects and embodiments disclosed herein are for purposes of illustration and are not intended to be limiting, with the true scope and spirit being indicated by the following claims.



In the specification the term “comprising” shall be understood to have a broad meaning similar to the term “including” and will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps. This definition also applies to variations on the term “comprising” such as “comprise” and “comprises”.

The reference to any prior art in this specification is not, and should not be taken as an acknowledgement or any form of suggestion that the referenced prior art forms part of the common general knowledge in Australia.

WHAT IS CLAIMED IS:

1. A method, comprising:

(a) contacting a first sample comprising a first plurality of cells and a second sample comprising a second plurality of cells with a first oligonucleotide-conjugated antibody and a second oligonucleotide-conjugated antibody, respectively, to form a first barcoded sample and a second barcoded sample, respectively, each comprising cells associated with the first oligonucleotide-conjugated antibody and the second oligonucleotide-conjugated antibody, respectively,

wherein the first antibody is capable of specifically binding to a protein target of each of the first plurality of cells, and is conjugated with a first sample indexing oligonucleotide comprising a poly(A) tail and a first sample indexing sequence,

wherein the second antibody is capable of specifically binding to a protein target of each of the second plurality of cells, and is conjugated with a second sample indexing oligonucleotide comprising a poly(A) tail and a second sample indexing sequence,

wherein the first sample indexing sequence and the second sample indexing sequence comprise different sequences, and

wherein the protein target of each of the first plurality of cells and the protein target of each of the second plurality of cells are the same;

(b) pooling the first barcoded sample and the second barcoded sample to form a combined barcoded sample comprising the first plurality of cells associated with the first oligonucleotide-conjugated antibody and the second plurality of cells associated with the second oligonucleotide-conjugated antibody;

(c) partitioning the first plurality of cells associated with the first oligonucleotide-conjugated antibody and the second plurality of cells associated with the second oligonucleotide-conjugated antibody of the combined barcoded sample to a plurality of partitions, wherein partitions of the plurality of partitions each comprises a single cell of the combined barcoded sample;

(d) in each of the partitions of the plurality of partitions, contacting a barcoding particle with the first sample indexing oligonucleotide or the second

sample indexing oligonucleotide, wherein the barcoding particle comprises a plurality of oligonucleotide probes each comprising a poly(T) region and a barcode sequence common to the plurality of oligonucleotide probes of the barcoding particle;

(e) extending each oligonucleotide probe hybridized to the first sample indexing oligonucleotide or the second indexing oligonucleotide via the hybridization between the poly(A) tail of the first sample indexing oligonucleotide or the second sample indexing oligonucleotide and the poly(T) region of the oligonucleotide probe to produce a first labeled nucleic acid or a second labeled nucleic acid, wherein the first labeled nucleic acid and the second labeled nucleic acid comprises a first sample indexing sequence and a second sample indexing sequence, respectively, or a complementary sequence thereof, and the barcode sequence; and

(f) obtaining sequencing information of the first labeled nucleic acid comprising the first sample indexing sequence or the second labeled nucleic acid comprising the second sample indexing sequence, or a portion thereof; and

(g) determining each single cell in the partitions of the plurality of partitions of the combined barcoded sample is from the first sample or the second sample based on the first sample indexing sequence or the second sample indexing sequence, respectively, a complementary sequence thereof, or a portion thereof, in the sequence information.

2. The method of claim 1, comprising after step (a), removing the first oligonucleotide-conjugated antibody and the second oligonucleotide-conjugated antibody that are not associated with the first plurality of cells and the second plurality of cells, respectively.

3. The method of any one of claims 1-2, wherein the first antibody capable of specifically binding to the protein target of each of the first plurality of cells and the second antibody capable of specifically binding to the protein target of each of the second plurality of cells are the same.

4. The method of any one of claims 1-3, comprising detaching the first sample indexing oligonucleotide and the second sample indexing oligonucleotide from the first oligonucleotide-conjugated antibody and the second oligonucleotide-conjugated antibody, respectively.

5. The method of any one of claims 1-4, wherein the at least one of the protein target of each of the first plurality of cells and the protein target of each of the second plurality of cells is a cell surface protein, an intracellular protein, a cell marker, a B-cell receptor, a T-cell receptor, an antibody, a major histocompatibility complex, a tumor antigen, a receptor, or a combination thereof.

6. The method of any one of claims 1-5, wherein partitioning the first plurality of cells and the second plurality of cells of the combined barcoded sample in step (c) comprises partitioning the first plurality of cells associated with the first oligonucleotide-conjugated antibody and the second plurality of cells associated with the second oligonucleotide-conjugated antibody of the combined sample and a plurality of barcoding particles comprising the barcoding particle to the plurality of partitions.

7. The method of any one of claims 1-6, wherein step (f) obtaining sequencing information of the first labeled nucleic acid or the second labeled nucleic acid, or a portion thereof comprises subjecting the first labeled nucleic acid or the second labeled nucleic acid to one or more reactions to generate a first set of nucleic acids or a second set of nucleic acids for nucleic acid sequencing.

8. The method of any one of claims 1-7, wherein each of the oligonucleotide probes comprises a cell label, a binding site for a universal primer, an amplification adaptor, a sequencing adaptor, or a combination thereof.

9. The method of any one of claims 1-8, wherein at least one oligonucleotide probe of the plurality of oligonucleotide probes is immobilized on the barcoding particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or a combination thereof.

10. The method of any one of claims 1-9, wherein step (d) comprises releasing the plurality of oligonucleotide probes from the barcoding particle in the partition comprising the single cell.

11. The method of any one of claims 1-10, wherein the barcoding particle is a Sepharose bead, a streptavidin bead, an agarose bead, a magnetic bead, a silica bead, a silica-like bead, an anti-biotin microbead, an anti-fluorochrome microbead, a hydrogel bead, or a disruptable hydrogel bead.

12. The method of any one of claims 1-11, wherein a single cell of one of the partitions of the plurality of partitions comprises a plurality of proteins of interest, and wherein:

step (a) further comprises contacting a third plurality of oligonucleotide-conjugated antibodies with the single cell, wherein each of the third plurality of oligonucleotide-conjugated antibodies comprises an antibody conjugated with an antibody specific oligonucleotide comprising a unique identifier for the antibody conjugated therewith and a poly(A) tail, and wherein the antibody is capable of specifically binding to at least one of the plurality of proteins of interest;

step (e) further comprises extending the oligonucleotide probes hybridized to the antibody specific oligonucleotides via the hybridization between the poly(A) tails of the antibody specific oligonucleotides and the poly(T) regions of the oligonucleotide probes to produce a first plurality of labeled nucleic acids, wherein each of the first plurality of labeled nucleic acids comprises a unique identifier, or a complementary sequence thereof, and a barcode sequence; and

step (f) further comprises obtaining sequence information of the first plurality of labeled nucleic acids or a portion thereof to determine the quantity of one or more of the plurality of proteins of interest in the single cell.

13. The method of claim 12, wherein the antibody specific oligonucleotide comprises a molecular label, a cell label, a binding site for a universal primer, an amplification adaptor, a sequencing adaptor, or a combination thereof.

14. The method of any one of claims 12-13, wherein a protein of interest of the plurality of proteins of interest is a cell surface protein, an intracellular protein, a cell marker, a B-cell receptor, a T-cell receptor, an antibody, a major histocompatibility complex, a tumor antigen, a receptor, or a combination thereof.

15. The method of any one of claims 12-14, wherein the single cell comprises a plurality of mRNA target molecules, and wherein:

step (d) further comprises hybridizing mRNA target molecules from the single cell with oligonucleotide probes of the plurality of oligonucleotide probes between the poly(A) tails of the mRNA target molecules and the poly(T) regions of the oligonucleotide probes;

step (e) further comprises extending the oligonucleotide probes hybridized to the mRNA target molecules via the hybridization between the poly(A) tails of the mRNA target molecules and the poly(T) regions of the oligonucleotide probes to produce a second plurality of labeled nucleic acids, wherein each of the second plurality of labeled nucleic acids comprises a complementary sequence of one of the mRNA target molecules and a barcode sequence; and

step (f) further comprises obtaining sequence information of the second plurality of labeled nucleic acids or a portion thereof to determine the quantity of one or more of the plurality of mRNA target molecules in the single cell.

16. A method, comprising:

(a) contacting a first sample comprising a first plurality of cells and a second sample comprising a second plurality of cells with a first sample indexing composition and a second sample indexing composition, respectively, to form a first labeled sample and a second labeled sample, respectively, each comprising cells associated with the first sample indexing composition and the second sample indexing composition, respectively,

wherein the first sample indexing composition comprises a first cellular component binding reagent associated with a first sample indexing oligonucleotide, wherein the first cellular component binding reagent is capable of specifically binding to a cellular component target of each of the first plurality of cells,

wherein the second sample indexing composition comprises a second cellular component binding reagent associated with a second sample indexing

oligonucleotide, wherein the second cellular component binding reagent is capable of specifically binding to a cellular component target of each of the second plurality of cells,

wherein the first and second sample indexing oligonucleotides comprise a first sample indexing sequence and a second sample indexing sequence, respectively, and wherein the first sample indexing sequence and the second sample indexing sequence comprise different sequences, and

wherein the cellular component target of each of the first plurality of cells and the cellular component target of each of the second plurality of cells are the same;

(b) pooling the first labeled sample and the second labeled sample to form a combined labeled sample comprising cells associated with the first sample indexing composition and cells associated with the second sample indexing composition, respectively;

(c) partitioning the cells associated with the first sample indexing composition and the cells associated with the second sample indexing composition of the combined labeled sample to a plurality of partitions, wherein the partitions of the plurality of partitions each comprises a single cell of the combined labeled sample;

(d) subjecting the first sample indexing oligonucleotide or the second sample indexing oligonucleotide to one or more reactions to generate a first barcoded sample indexing oligonucleotide comprising the first sample indexing sequence, or a complement thereof, or a second barcoded sample indexing oligonucleotide comprising the second sample indexing sequence, or a complement thereof;

(e) obtaining sequencing information of the first barcoded sample indexing oligonucleotide or a second barcoded sample indexing oligonucleotide, or a portion thereof; and

(f) determining cells in the combined labeled sample is from the first sample or the second sample based on the first sample indexing sequence or the second sample indexing sequence, respectively, a complementary sequence thereof, or a portion thereof, in the sequence information obtained in (e).

17. The method of claim 16, wherein the cellular component binding reagent comprises a protein binding reagent, wherein the protein binding reagent is associated with the sample indexing oligonucleotide associated with the corresponding cellular component binding reagent, and wherein the cellular component target comprises a protein target.

18. The method of any one of claims 16-17, wherein subjecting the first sample indexing oligonucleotide or the second sample indexing oligonucleotide to one or more reactions to generate a first barcoded sample indexing oligonucleotide comprising the first sample indexing sequence, or complement thereof, or a second barcoded sample indexing oligonucleotide comprising the second sample indexing sequence, or a complementary sequence thereof, in (d) comprises:

barcoding the first sample indexing oligonucleotide or the second sample oligonucleotide using a plurality of barcodes to generate the first barcoded sample indexing oligonucleotide or the second barcoded sample indexing oligonucleotide.

19. The method of any one of claims 16-18, wherein determining the cells in the combined labeled sample is from the first sample or the second sample comprises identifying the presence or absence of the first sample indexing sequence, the second sample index sequence, a complementary sequence thereof, or a portion thereof, in the barcoded sample indexing compositions.

20. The method of any one of claims 16-19, wherein at least one of the cellular component targets of each of the first plurality of cells and the cellular component target of each of the second plurality of cells is on a cell surface.

21. The method of any one of claims 16-19, comprising removing the first and second sample indexing compositions that are not associated with the first plurality and the second plurality of cells.

22. The method of any one of claims 16-21, wherein the first sample indexing oligonucleotide is configured to be non-detachable from the first cellular component binding



reagent, or the second sample indexing oligonucleotide is configured to be non-detachable from the second component binding reagent, or both.

23. The method of any one of claims 16-21, comprising detaching the first sample indexing oligonucleotide from the first cellular component binding reagent, detaching the second sample indexing oligonucleotide from the second cellular component binding reagent, or both.

24. The method of any one of claims 16-23, wherein the first and/or second sample indexing oligonucleotide comprises a sequence complementary to a capture sequence configured to capture the sequence of the first and/or second sample indexing oligonucleotide, and wherein each of the barcodes comprises a target-binding region which comprises the capture sequence.

25. The method of claim 24, wherein the target-binding region comprises a poly(dT) region, and wherein the sequence of the sample indexing oligonucleotide complementary to the capture sequence comprises a poly(dA) region.

26. The method of any one of claims 16-25, wherein at least one of the first sample indexing oligonucleotide and the second sample indexing oligonucleotide comprises a molecular label sequence, a binding site for a universal primer, or both.

27. The method of any one of claim 16-26, wherein the cellular component target of each of the first plurality of cells, the cellular component target of each of the second plurality of cells, or both comprises a carbohydrate, a lipid, a protein, an extracellular protein, a cell-surface protein, a cell marker, a B-cell receptor, a T-cell receptor, a major histocompatibility complex, a tumor antigen, a receptor, an intracellular protein, or any combination thereof.

28. The method of any one of claims 16-27, wherein the plurality of barcodes is associated with a particle, and optionally the particle is a disruptable bead or a disruptable hydrogel bead.

29. The method of claim 28, wherein at least one barcode of the plurality of barcodes is immobilized on the particle, partially immobilized on the particle, enclosed in the particle, partially enclosed in the particle, or a combination thereof.

30. The method of any one of claims 28-29, wherein barcoding the first sample indexing oligonucleotide or the second sample indexing oligonucleotide using the plurality of barcodes comprises:

contacting the plurality of barcodes with the first sample indexing oligonucleotide or the second sample indexing oligonucleotide to allow barcodes to hybridize to the first sample indexing oligonucleotide or the second sample indexing oligonucleotide, wherein each of the barcodes comprises a target binding region capable of specifically binding to the first sample indexing oligonucleotide or the second sample indexing oligonucleotide, or both; and

extending the barcodes hybridized to the first sample indexing oligonucleotide or the second sample indexing oligonucleotide to generate the first barcoded sample indexing oligonucleotide or the second barcoded sample indexing oligonucleotide.

31. The method of any one of claims 28-30, wherein each of the plurality of barcodes comprises an identical barcode sequence.

32. The method of any one of claims 1-31, wherein (1) at least one of the first sample indexing sequence and the second sample indexing sequence is 6-60 nucleotides in length; (2) at least one of the first sample indexing oligonucleotide and the second sample indexing oligonucleotide is 50-500 nucleotides in length; or both.

33. The method of any one of claims 16-32, wherein at least one of the first cellular component binding reagent and the second cellular component binding reagent comprises an antibody, a tetramer, an aptamer, a protein scaffold, or a combination thereof.

34. The method of any one of claims 16-33, wherein the first sample indexing oligonucleotide is conjugated to the first cellular component binding reagent through a linker, or

the second sample indexing oligonucleotide is conjugated to the second cellular component binding reagent through a linker, or both.

35. The method of any one of claims 16-34, wherein the first cellular component binding reagent and the second cellular component binding reagent are the same.

36. The method of any one of claims 16-35, wherein at least one of the cellular component target of each of the first plurality of cells and the cellular component target of each of the second plurality of cells comprises a cell surface protein, an intracellular protein, a cell marker, a B-cell receptor, a T-cell receptor, an antibody, a major histocompatibility complex, a tumor antigen, a receptor, or a combination thereof.

37. The method of any one of claims 1-36, wherein the plurality of partitions comprises a well or a droplet.

38. The method of any one of claims 1-37, wherein the first plurality of cells, the second plurality of cells, or both comprise T cells, B cells, tumor cells, myeloid cells, blood cells, normal cells, fetal cells, maternal cells, or a mixture thereof.

39. The method of any one of claims 18-38, wherein each of the plurality of barcodes comprises a cell label, a binding site for a universal primer, an amplification adaptor, a sequencing adaptor, or a combination thereof.

40. The method of claim 28, wherein partitioning the first plurality of cells associated with the first sample indexing composition and the second plurality of cells associated with the second sample indexing composition of the combined barcoded sample to a plurality of partitions in (c) comprises partitioning the first plurality of cells associated with the first sample indexing composition, the second plurality of cells associated with the second sample indexing composition of the combined barcoded sample, and a plurality of particles comprising the particle that comprises the plurality of barcodes.

41. The method of claim 28 or 40, comprising releasing one or more of the plurality of barcodes from the particle in the partition comprising the single cell prior to (d).

42. The method of any one of claims 16-41, wherein subjecting the first sample indexing oligonucleotide or the second sample indexing oligonucleotide to one or more reactions comprises in the partitions of the plurality of partitions, subjecting the first sample indexing oligonucleotide or the second sample indexing oligonucleotide to one or more reactions to generate a first barcoded sample indexing oligonucleotide comprising the first sample indexing sequence, or a complement thereof, or a second barcoded sample indexing oligonucleotide comprising the second sample indexing sequence, or a complement thereof.

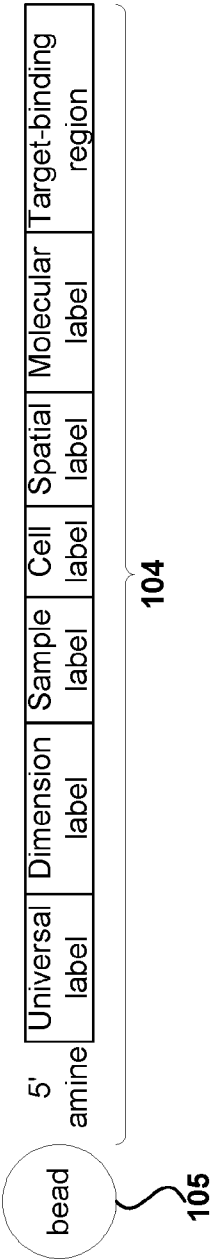


FIG. 1

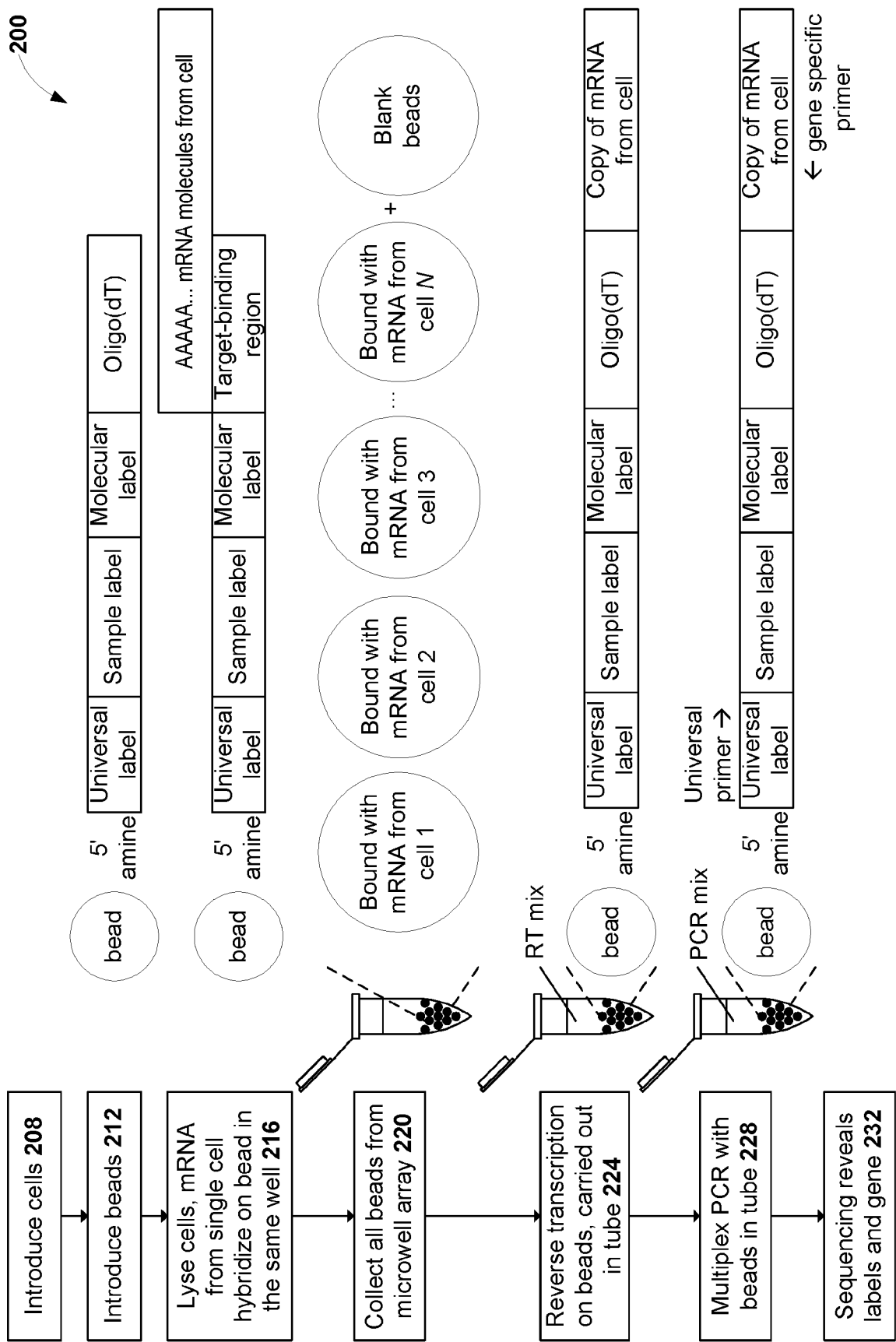


FIG. 2

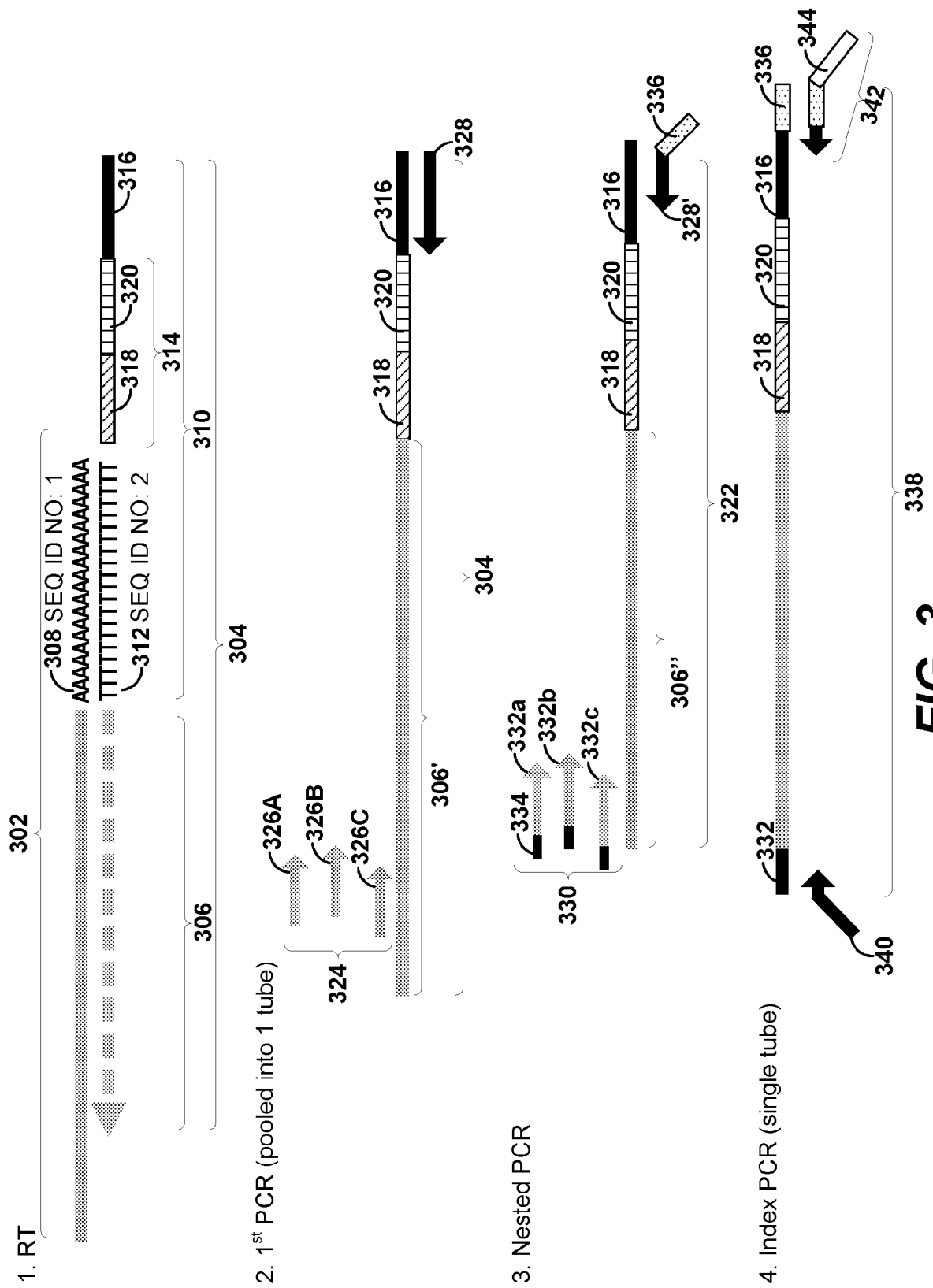


FIG. 3

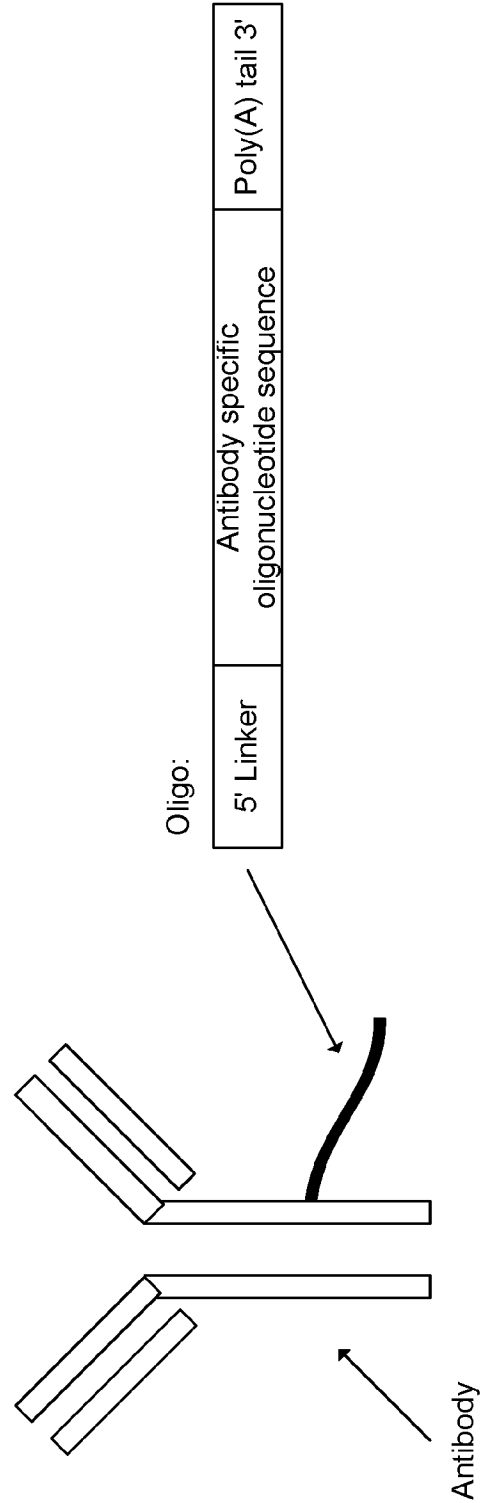


FIG. 4



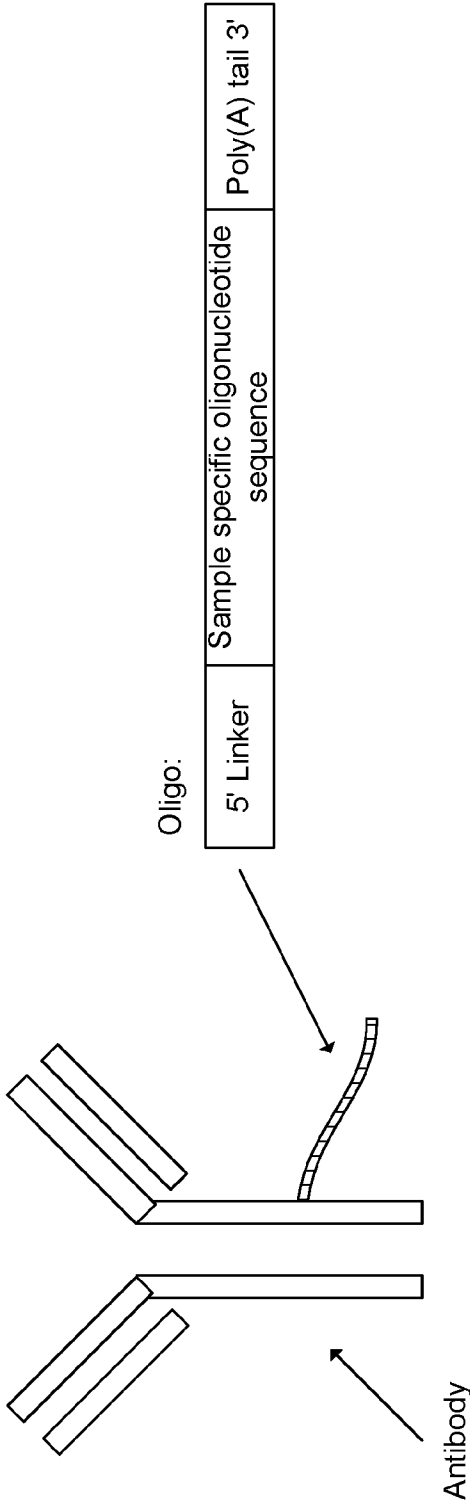


FIG. 5

6/112

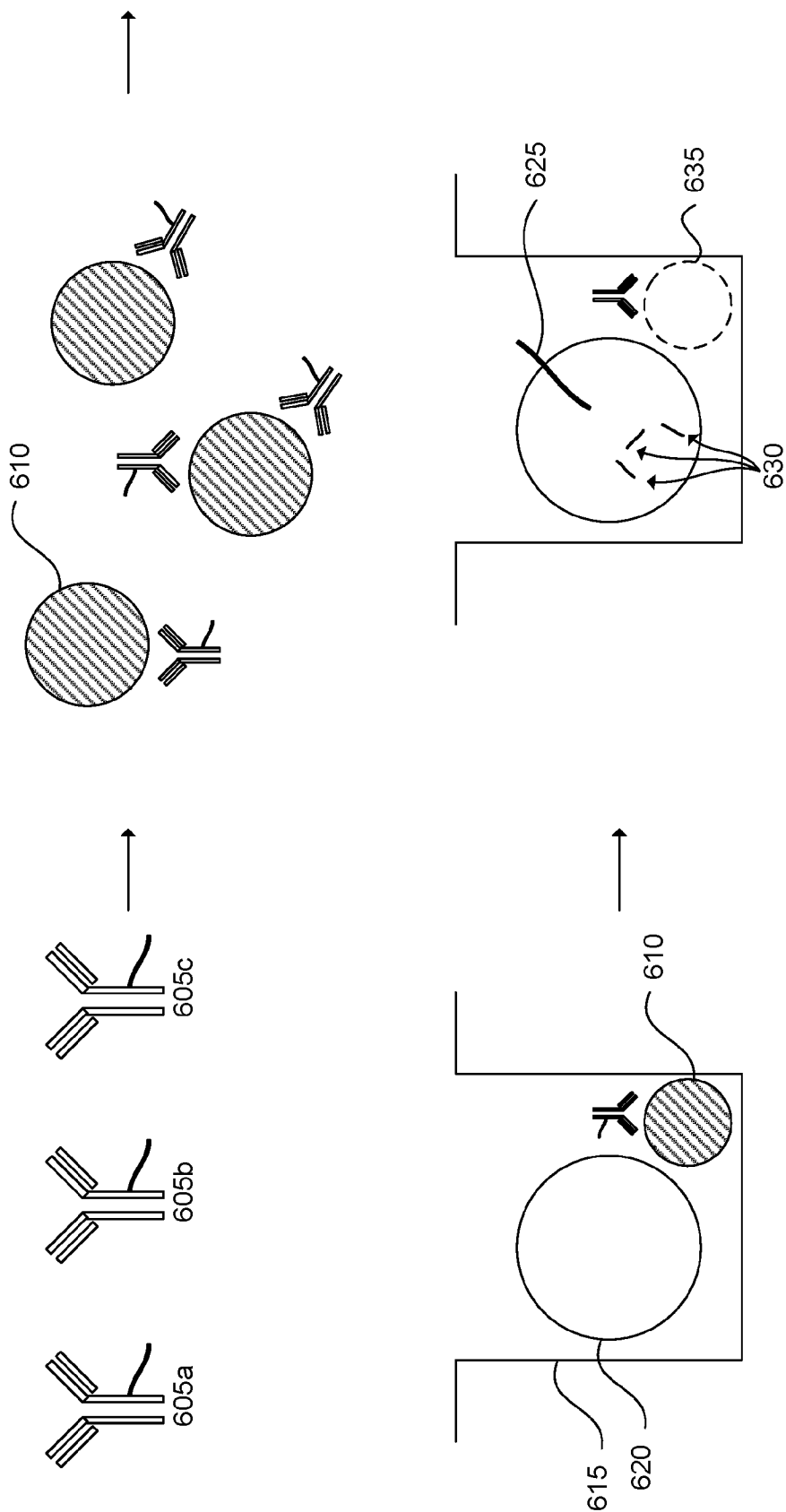


FIG. 6

7/112

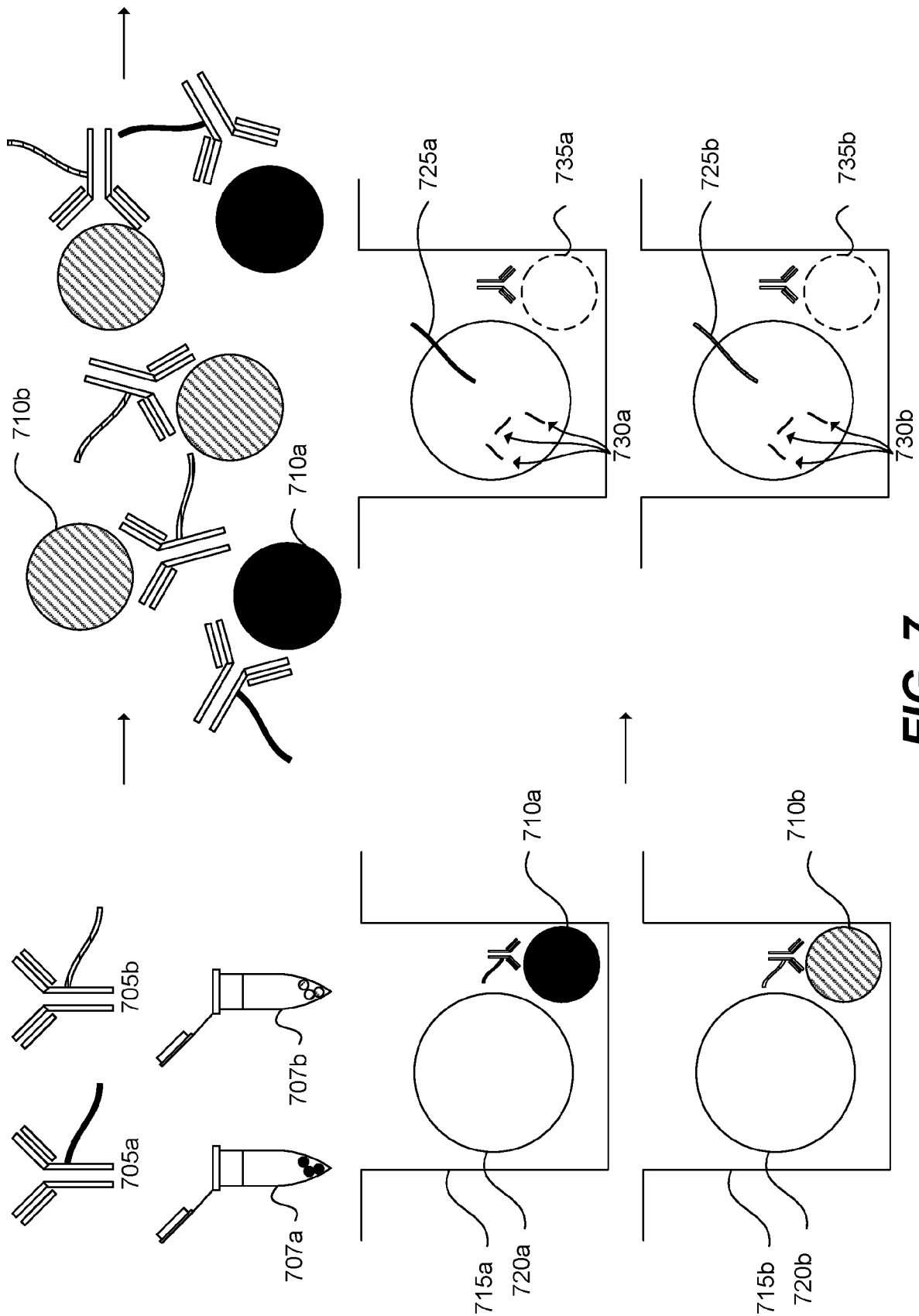


FIG. 7

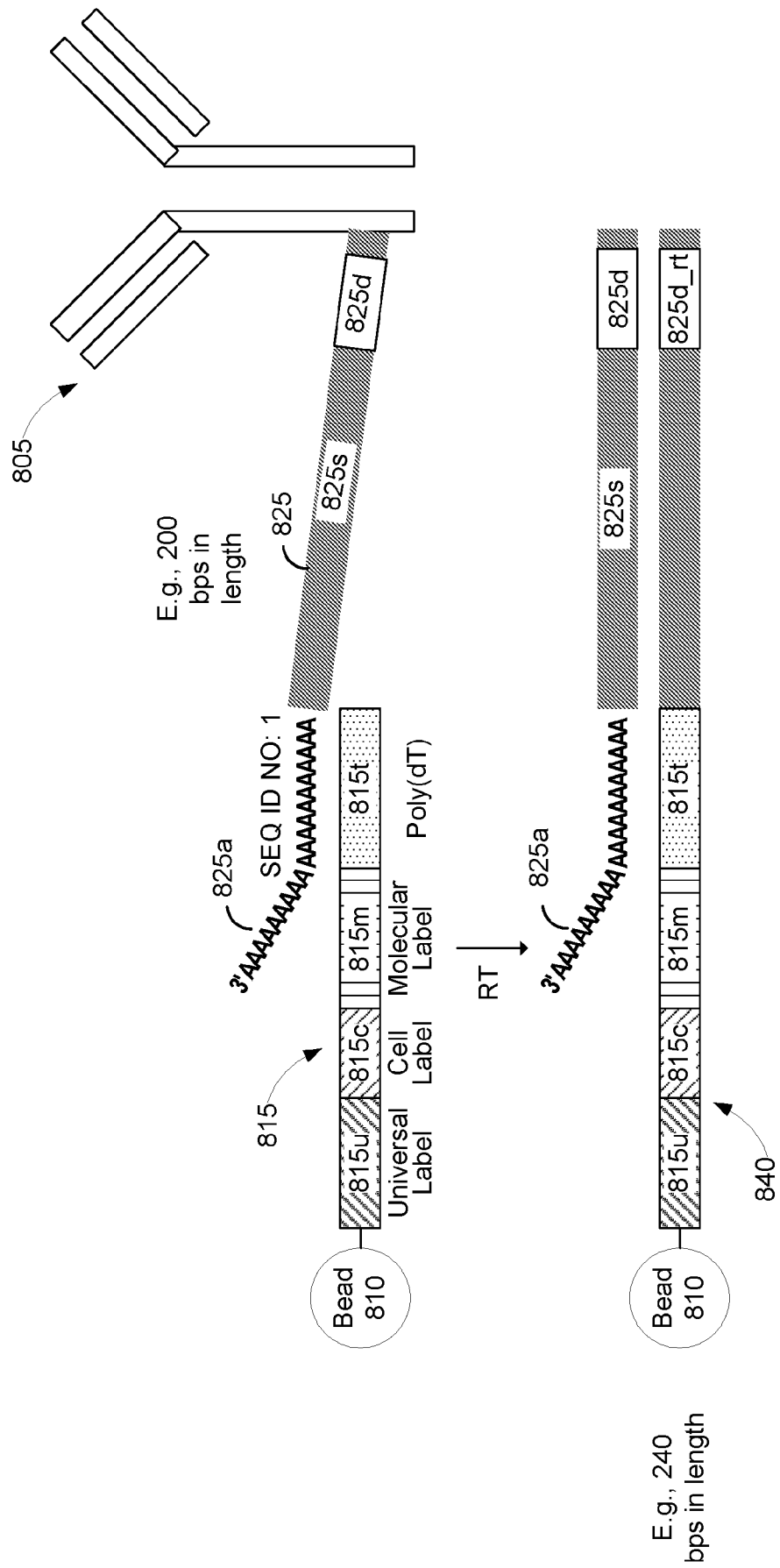
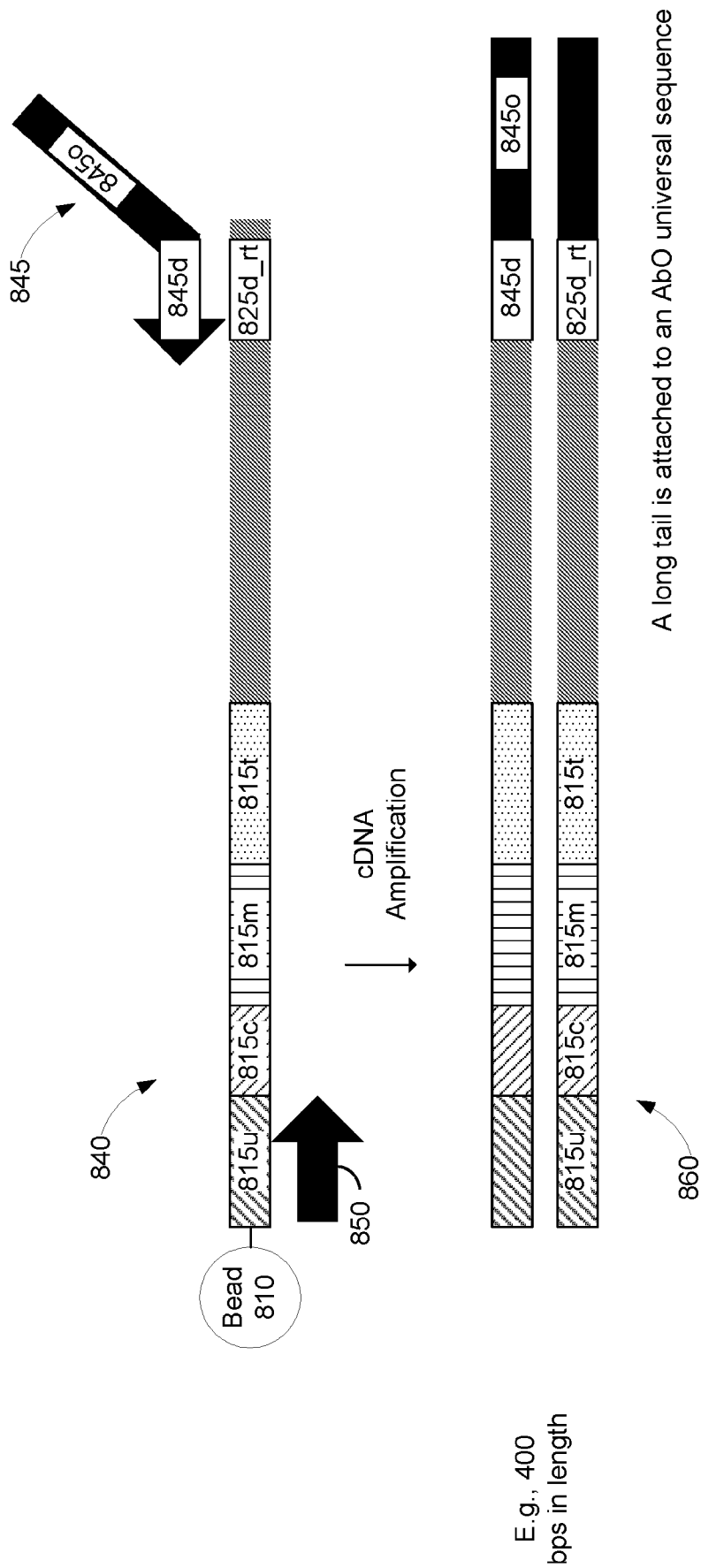


FIG. 8A1



**FIG. 8A2**

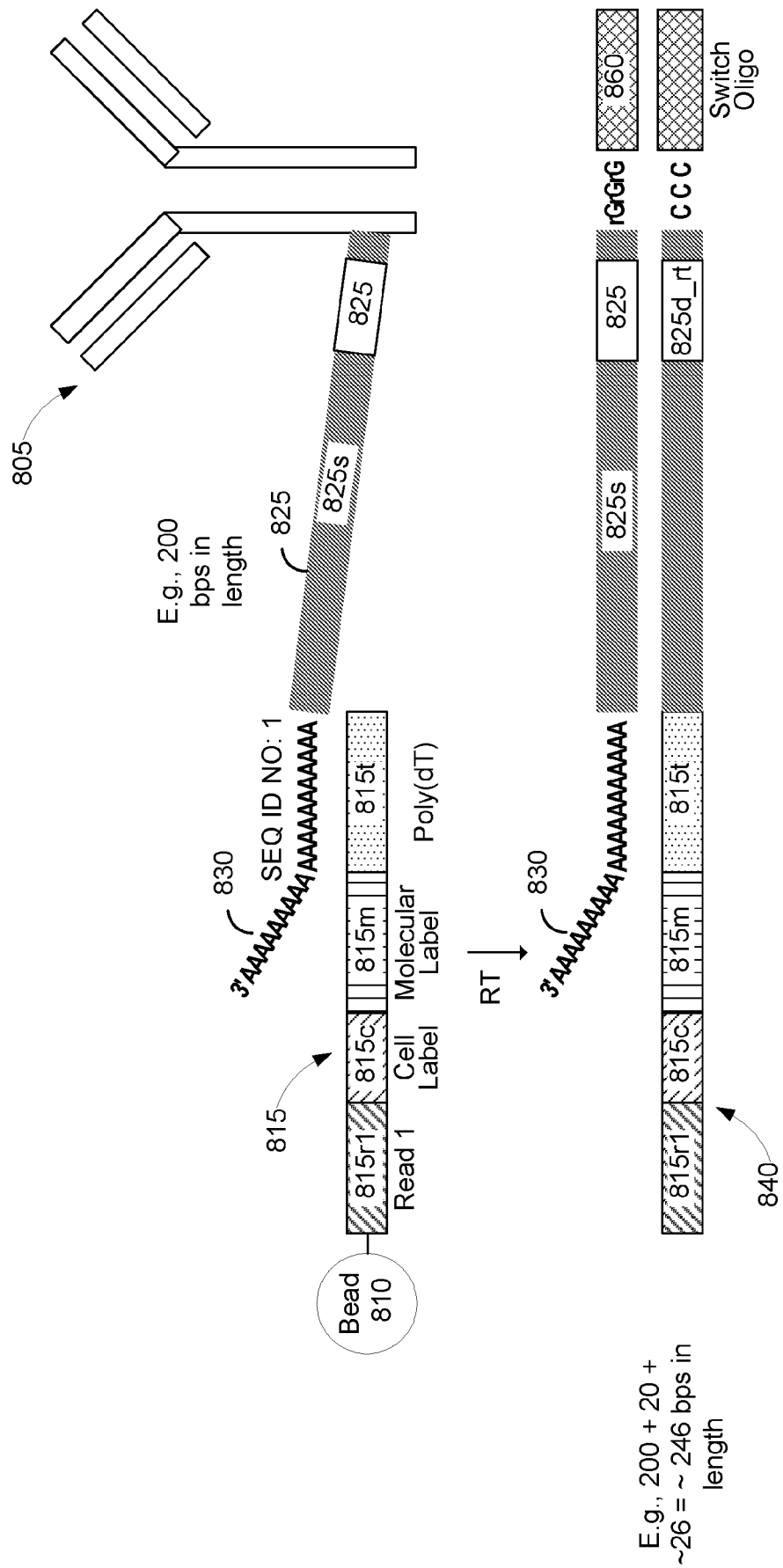


FIG. 8B1

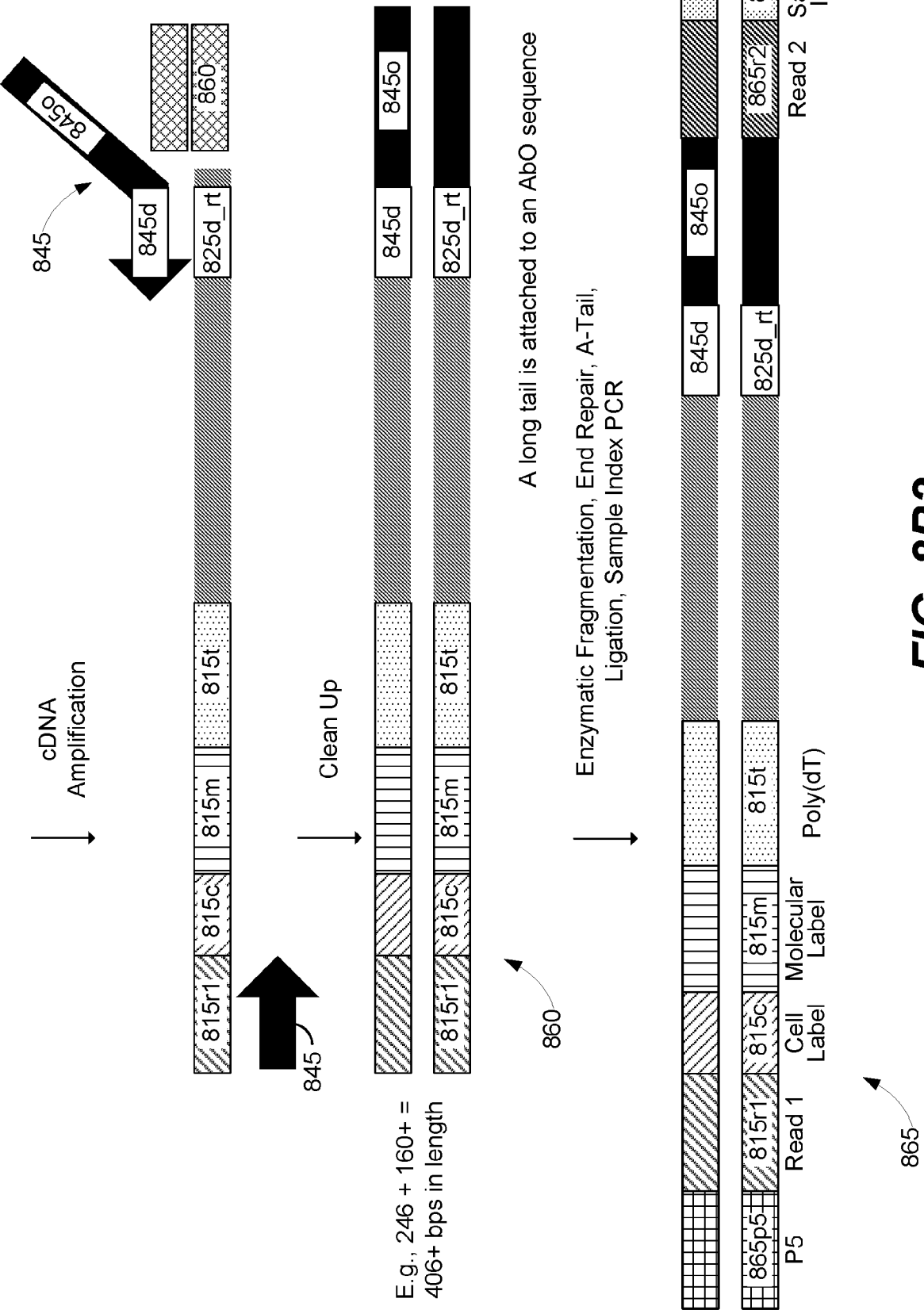


FIG. 8B2

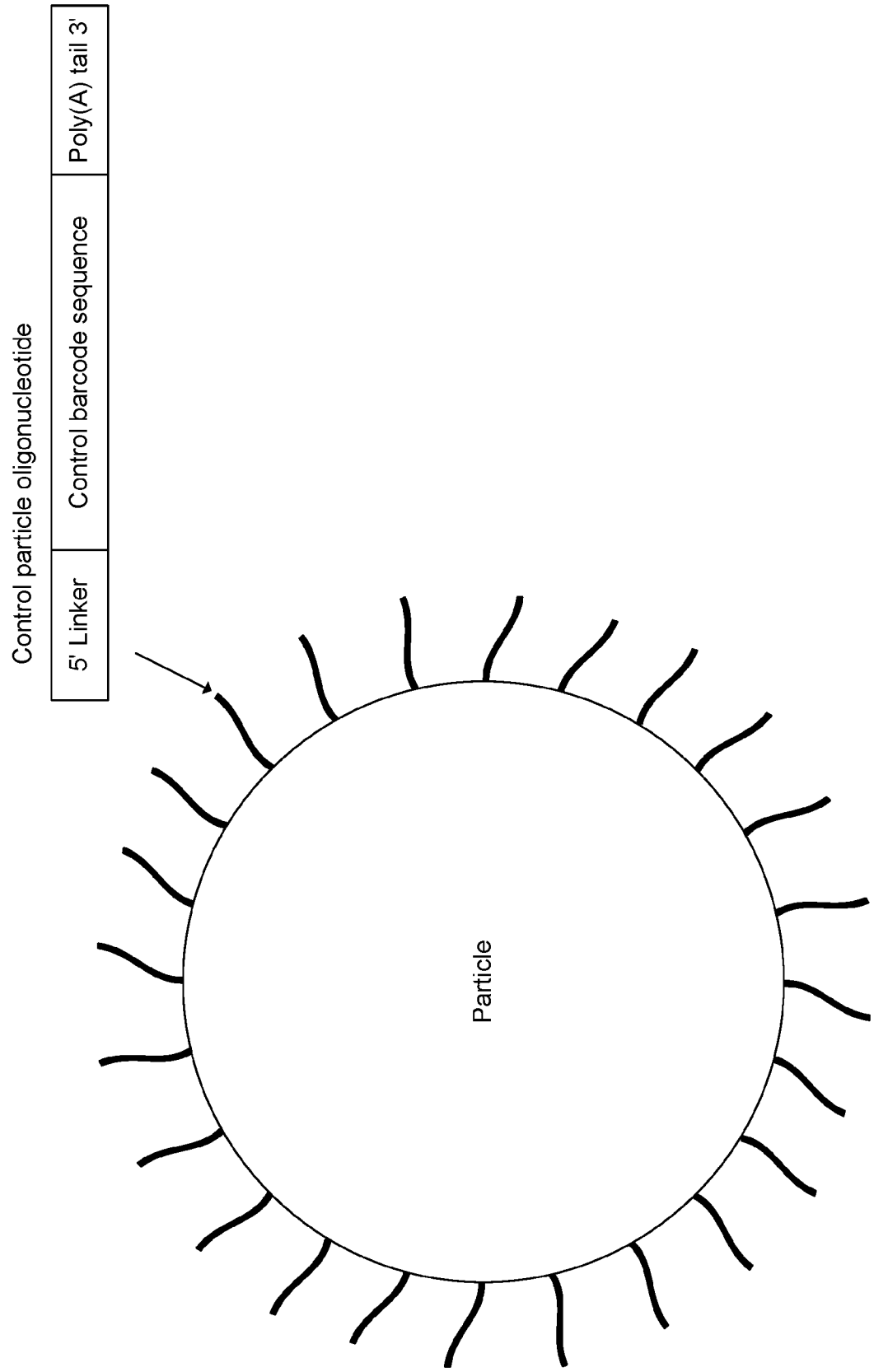


FIG. 9A



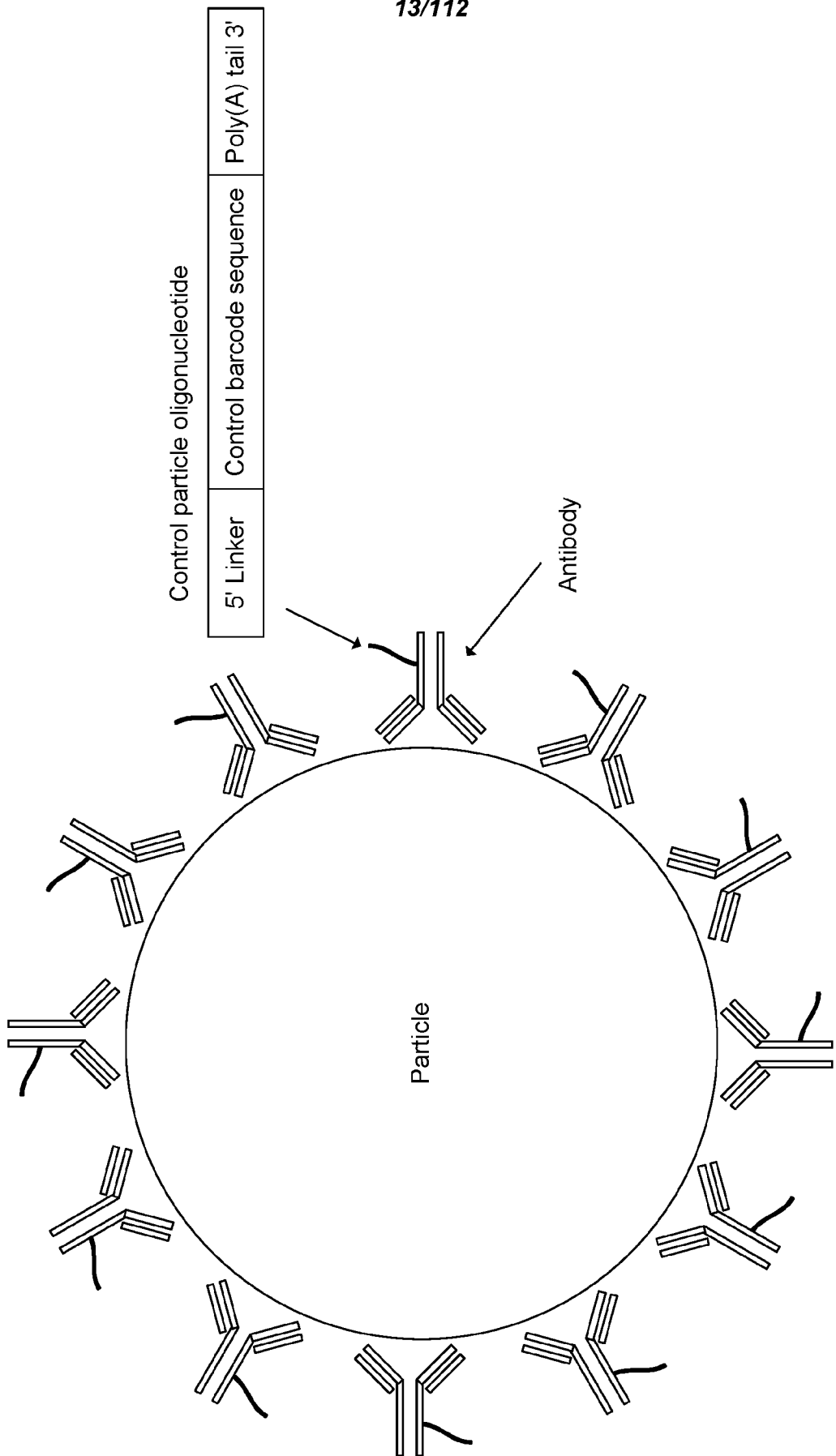


FIG. 9B

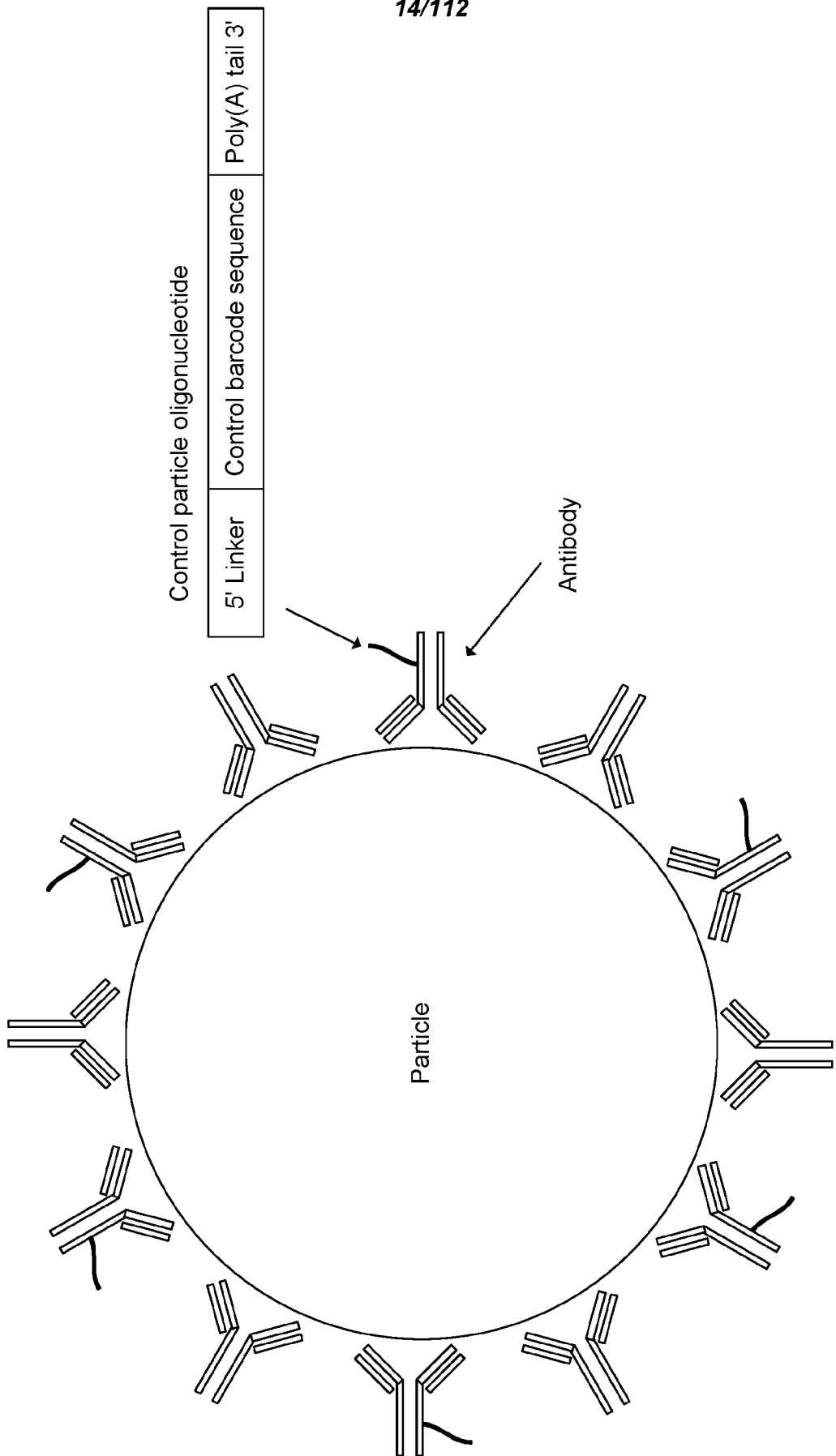


FIG. 9C

15/112

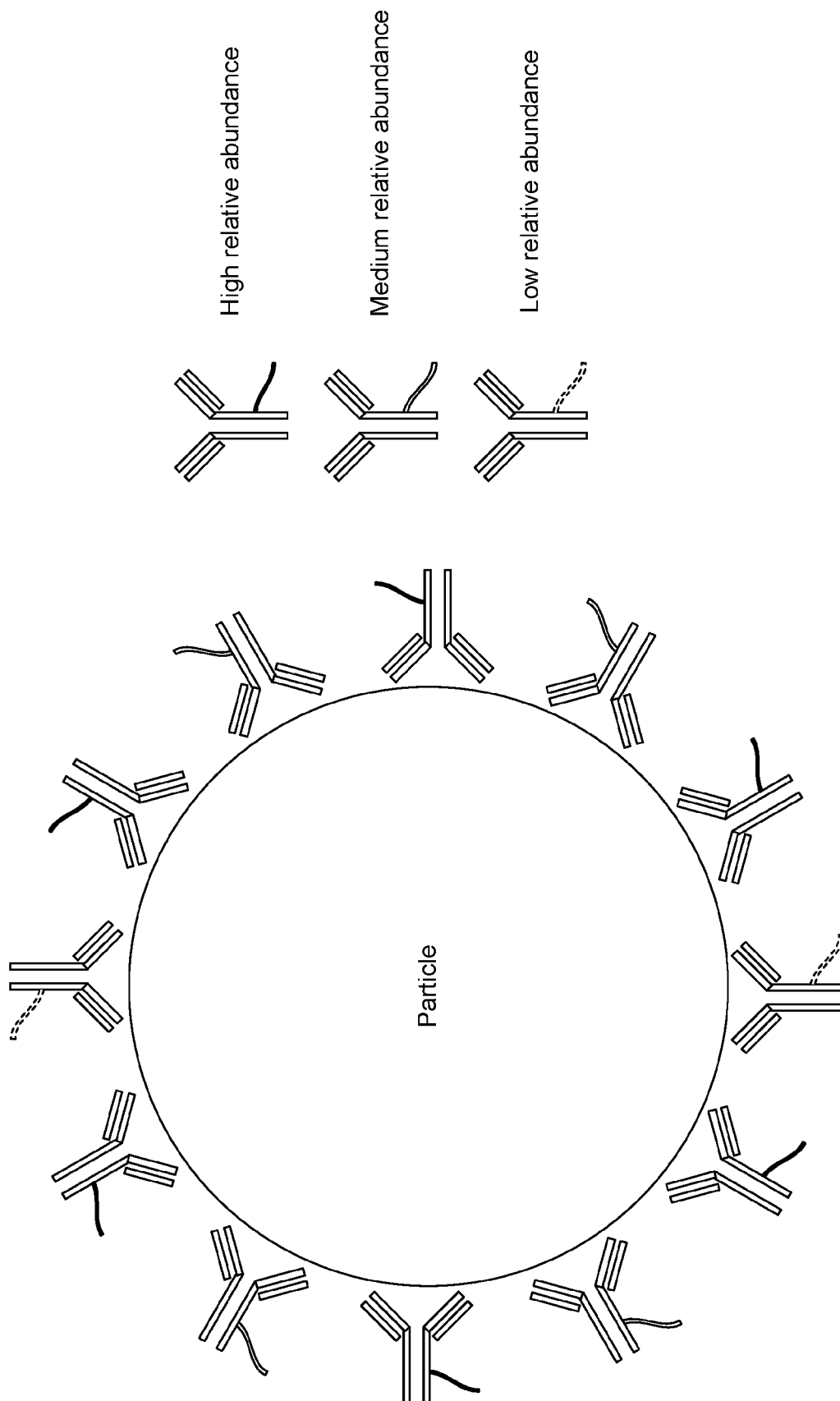
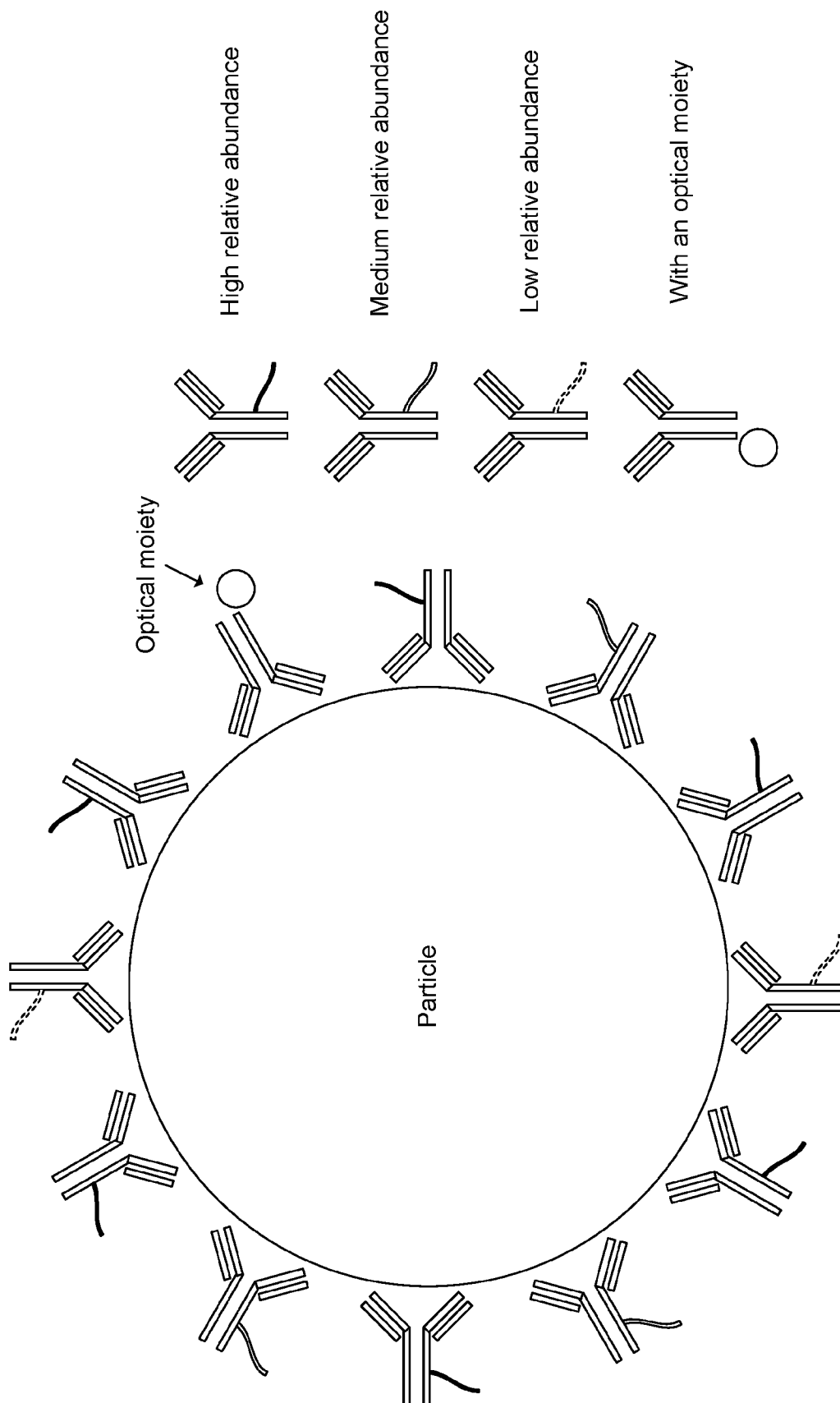
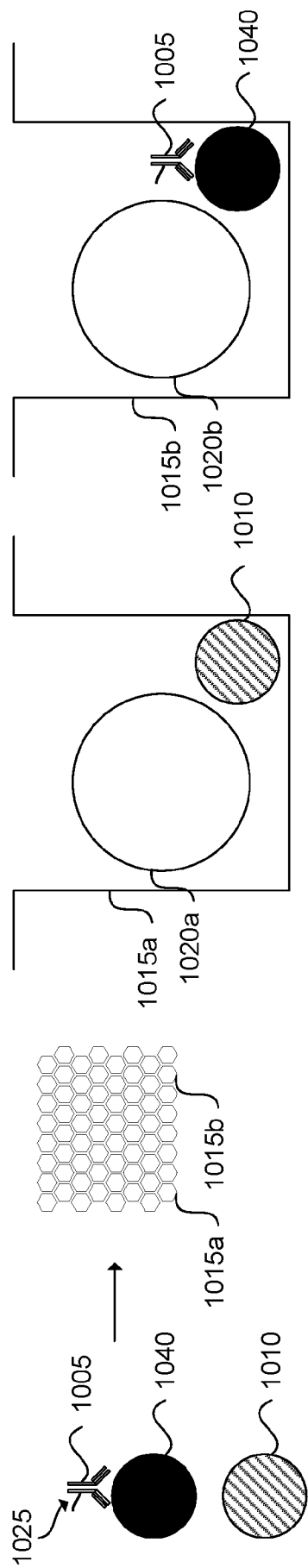


FIG. 9D

16/112



**FIG. 9E**



17/112

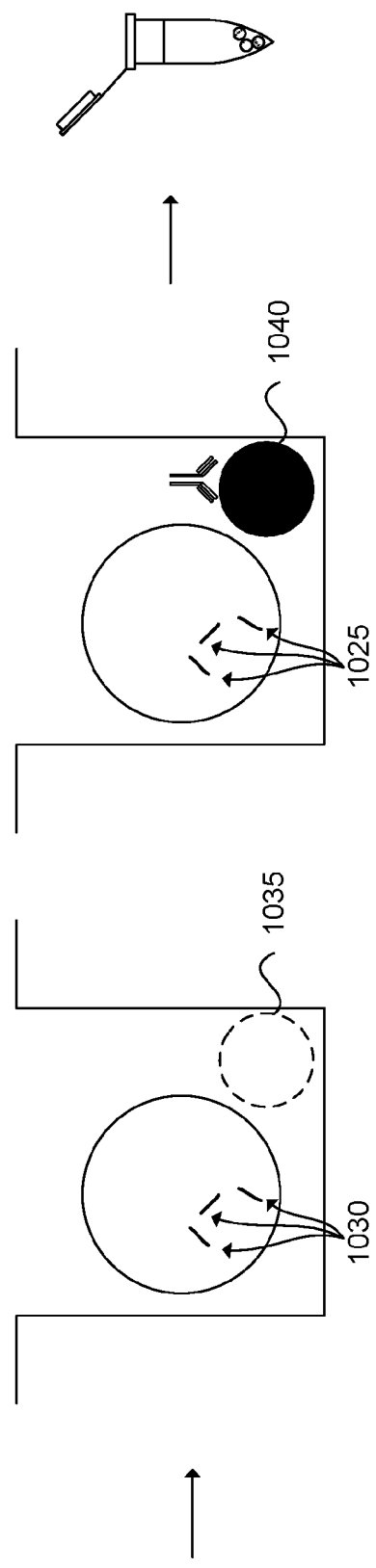


FIG. 10A

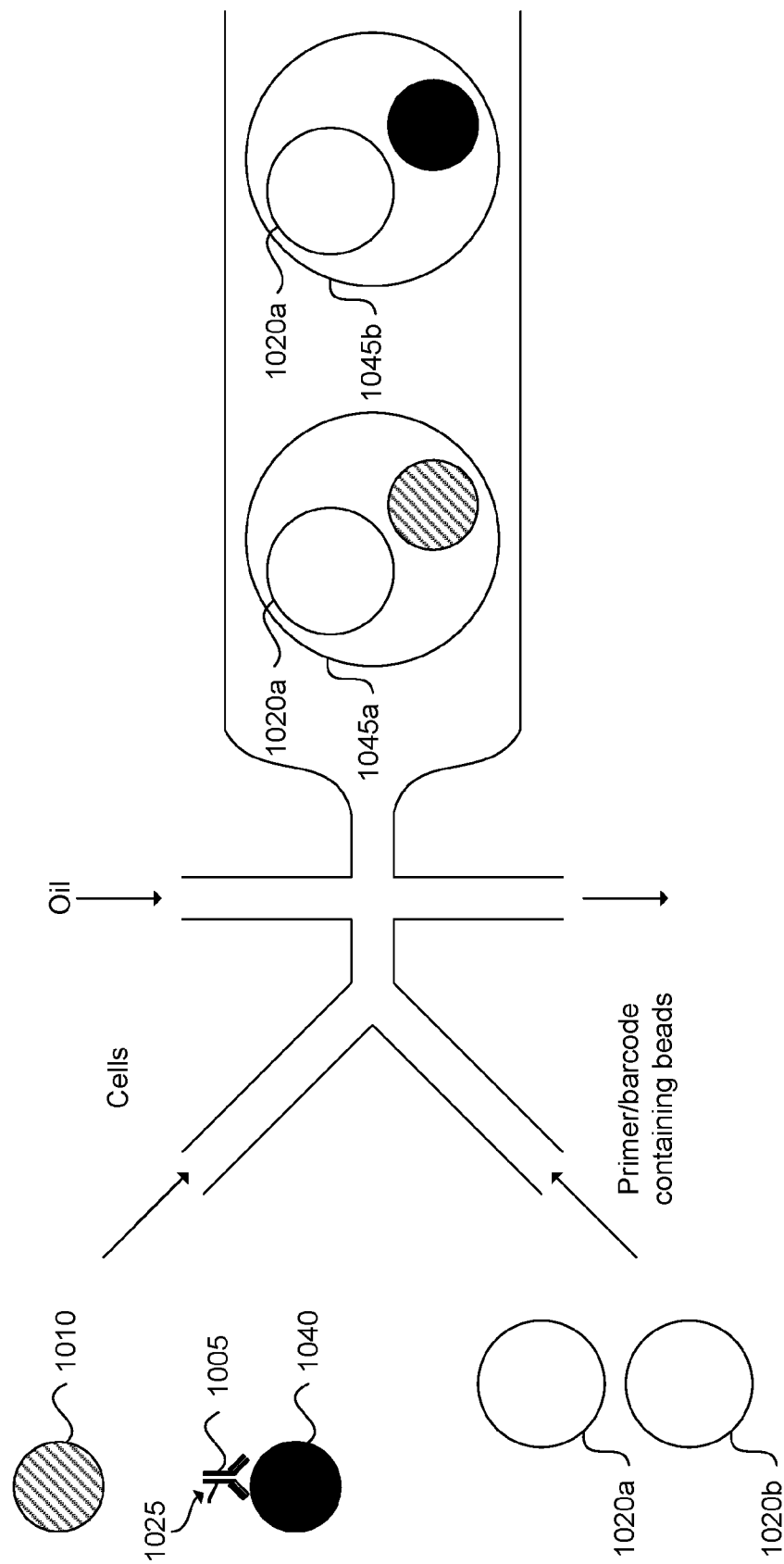


FIG. 10B

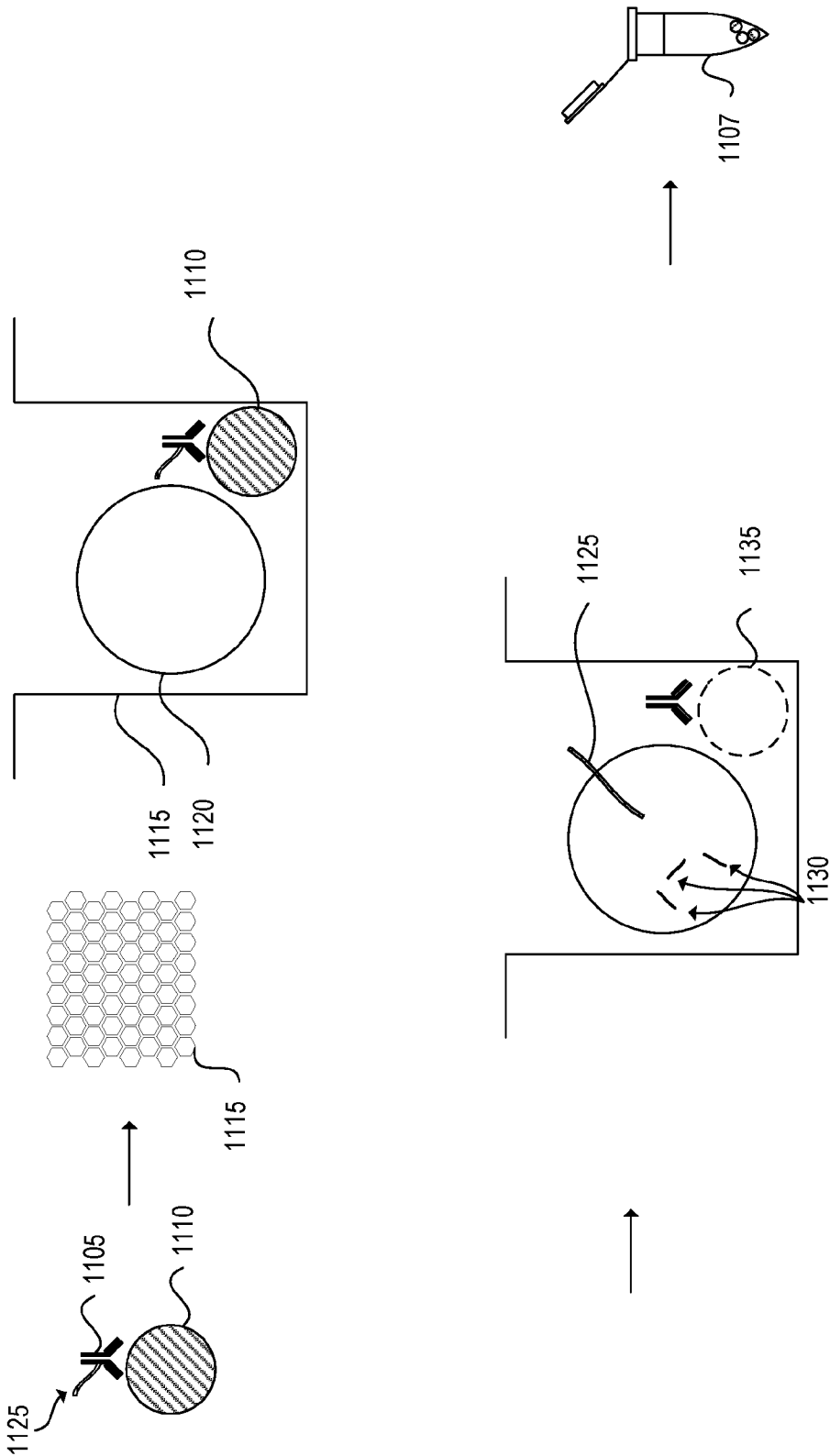


FIG. 11

20/112

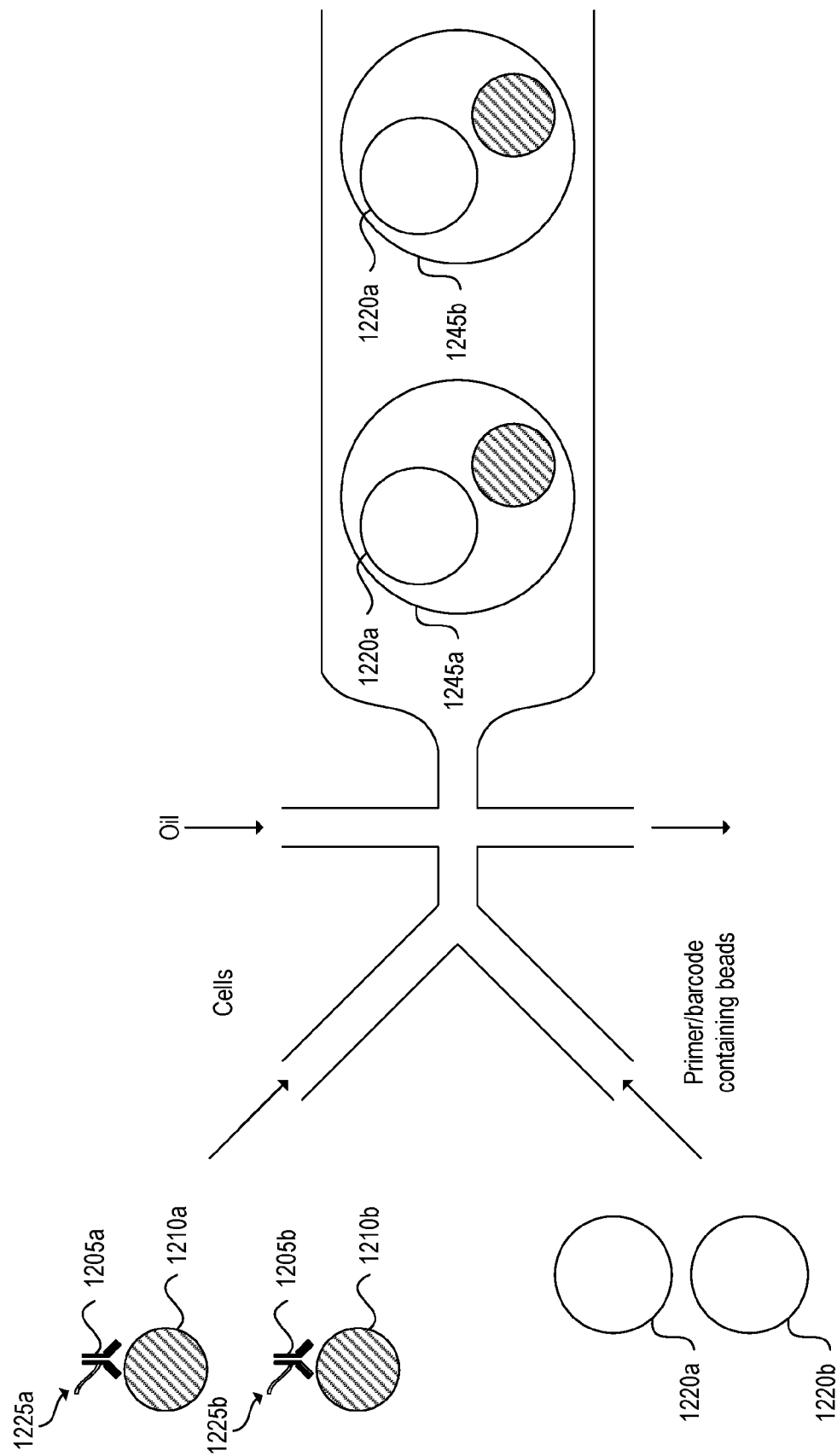


FIG. 12



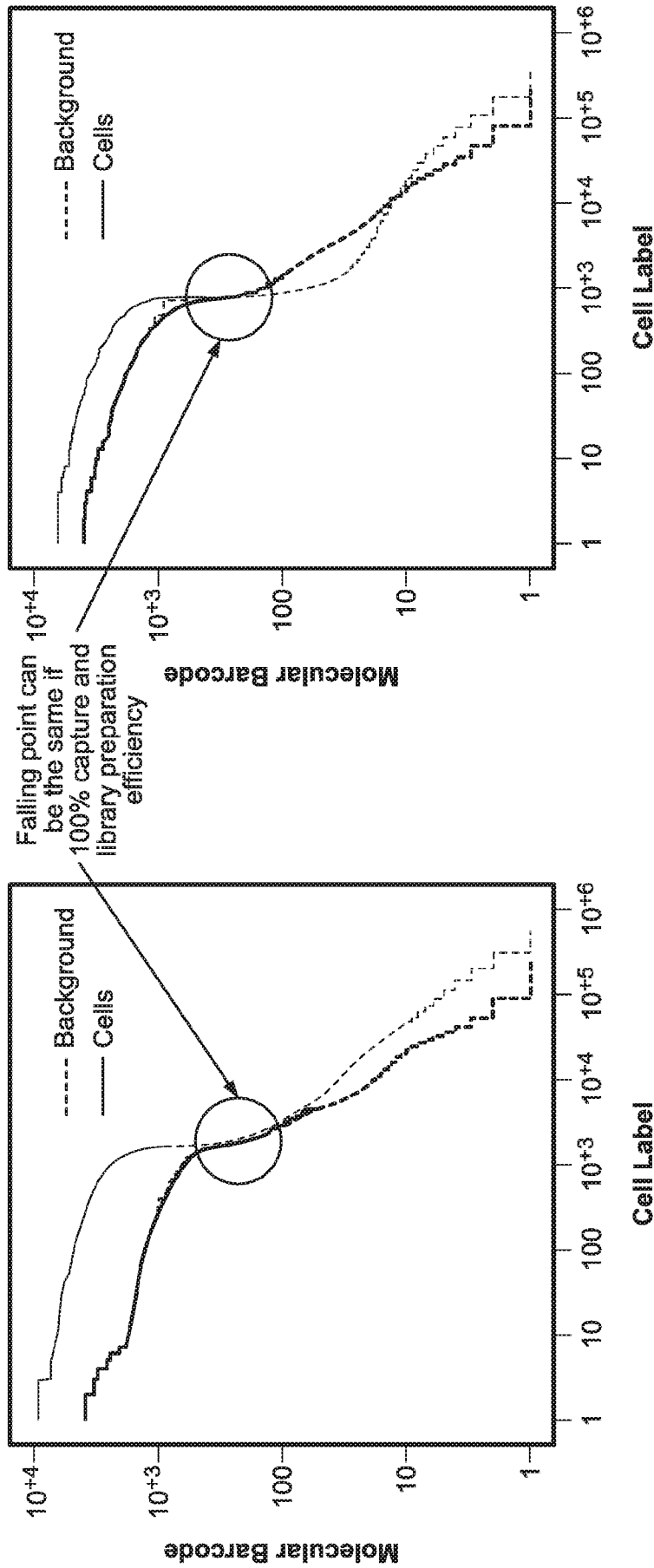
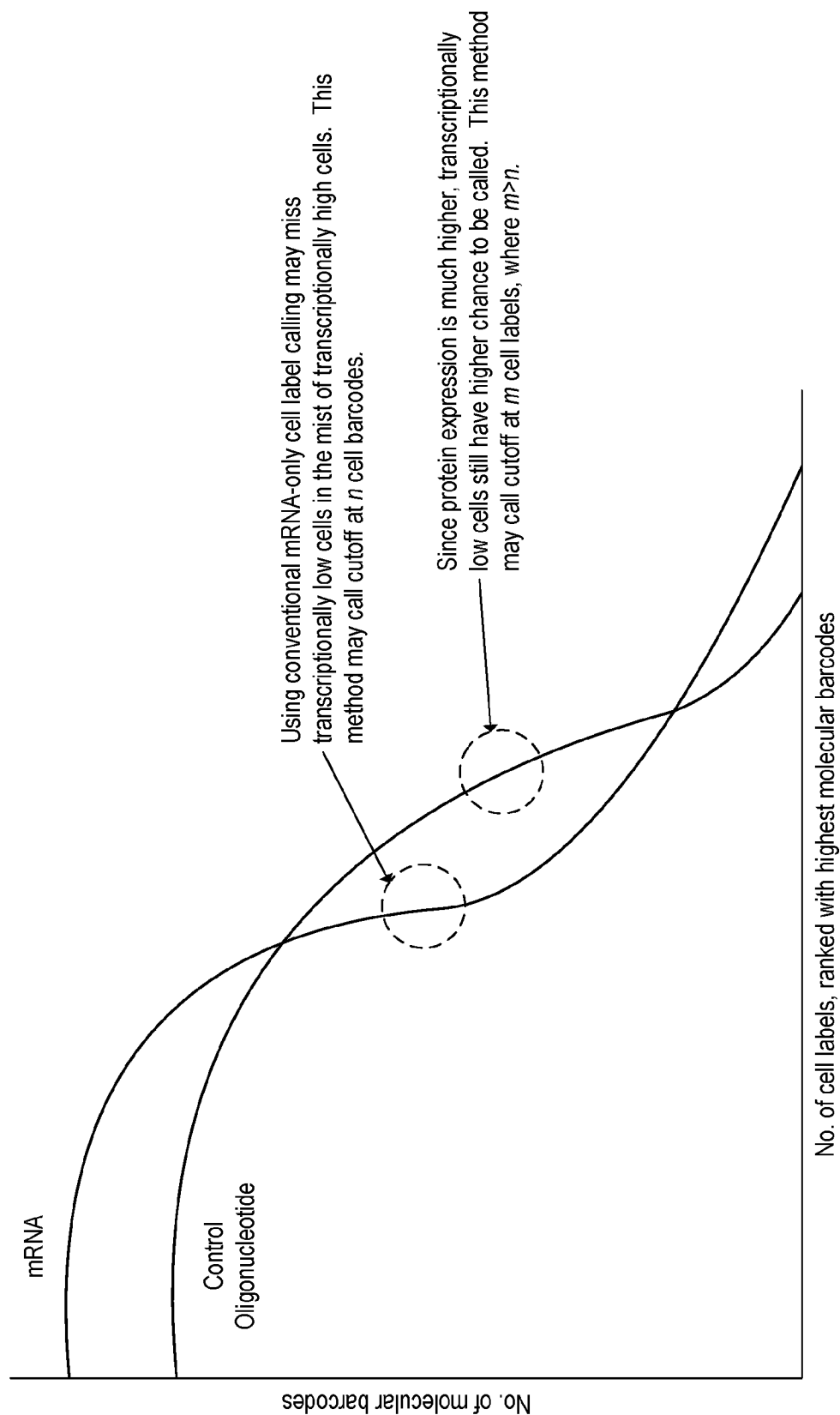


FIG. 13B

FIG. 13A

22/112

**FIG. 13C**

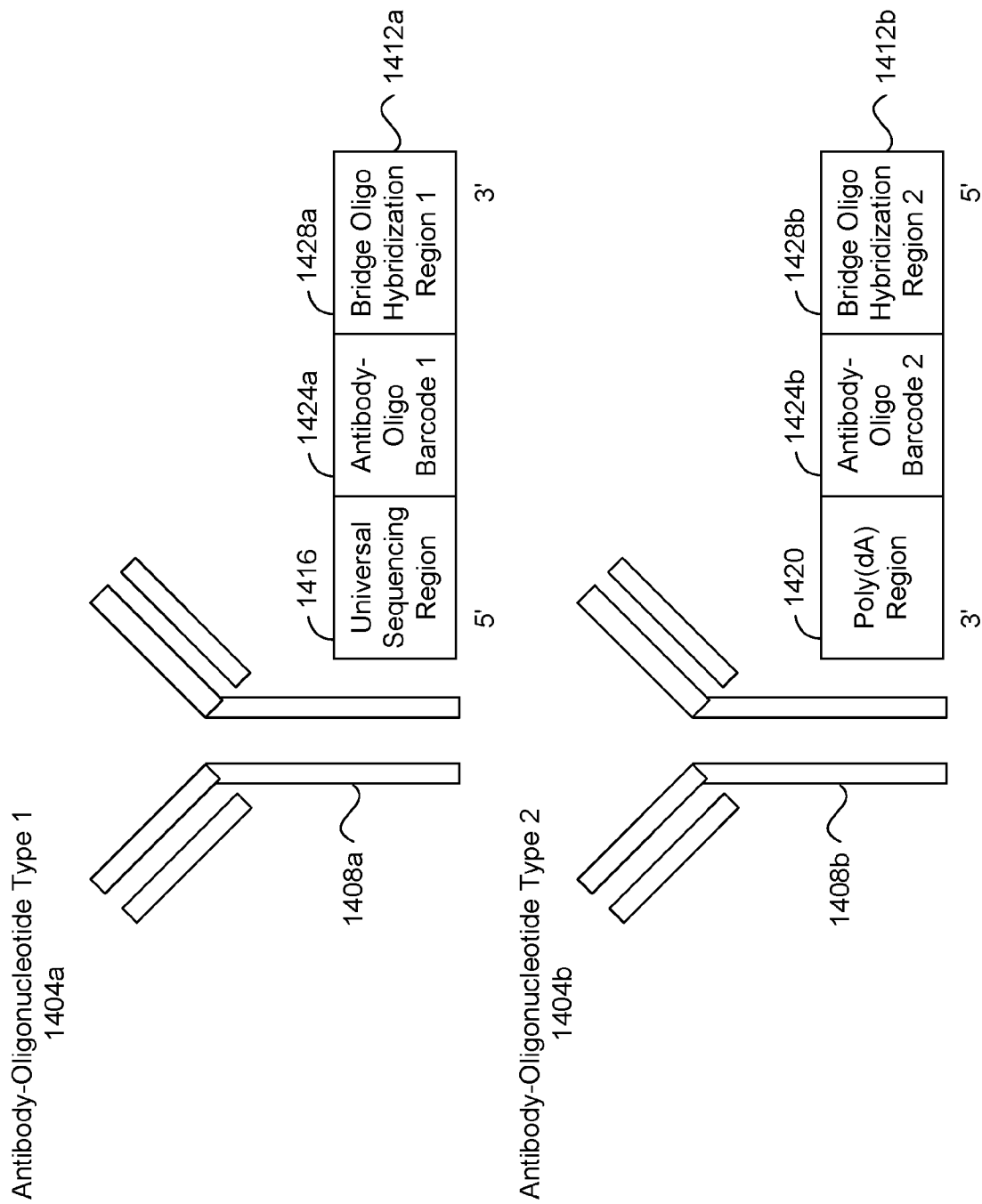
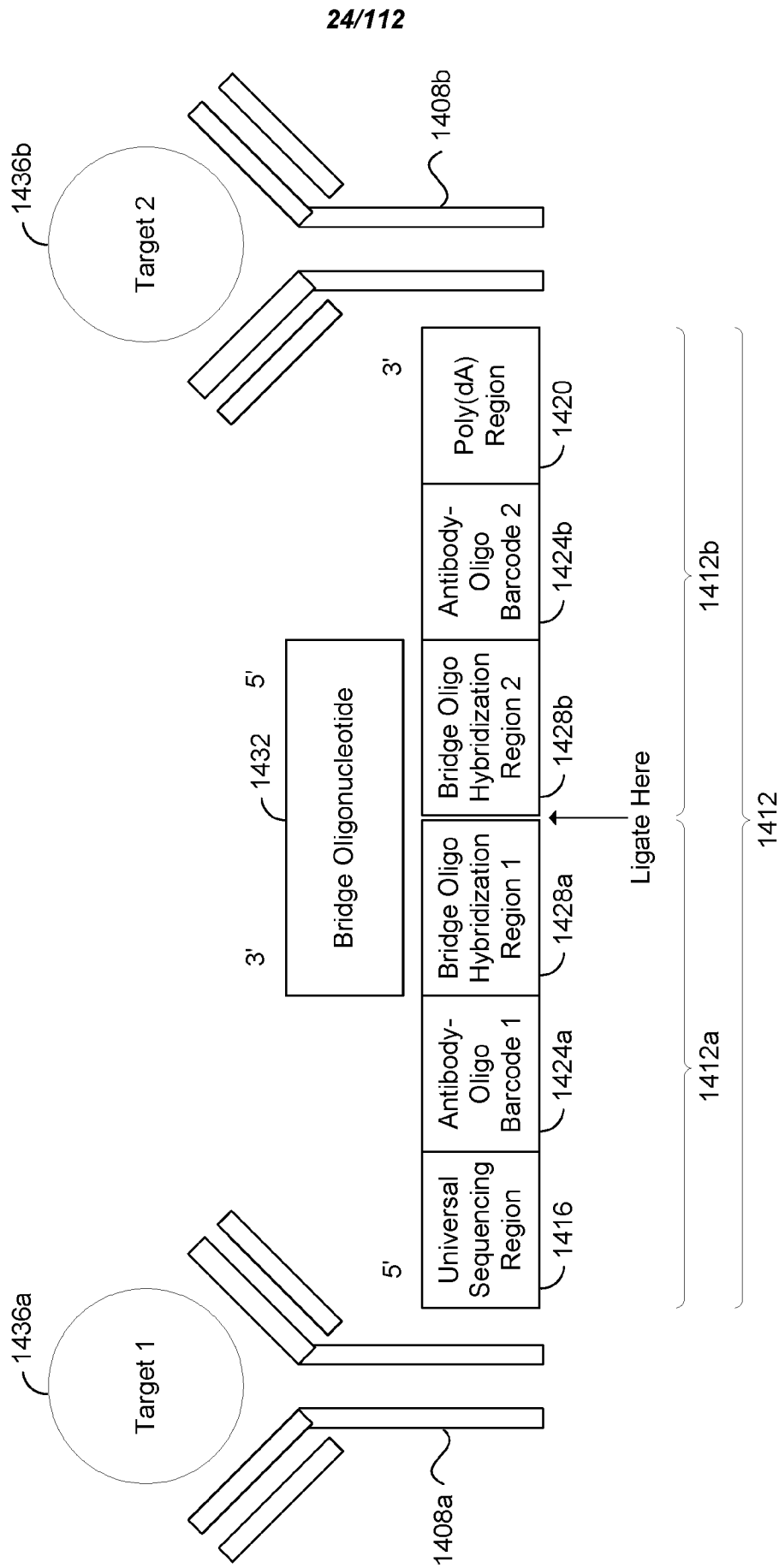


FIG. 14A



**FIG. 14B**



**FIG. 14C**

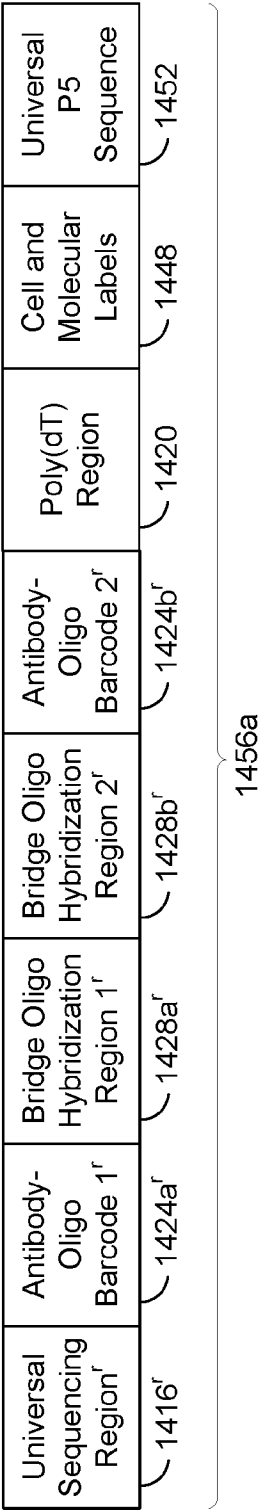


FIG. 14D

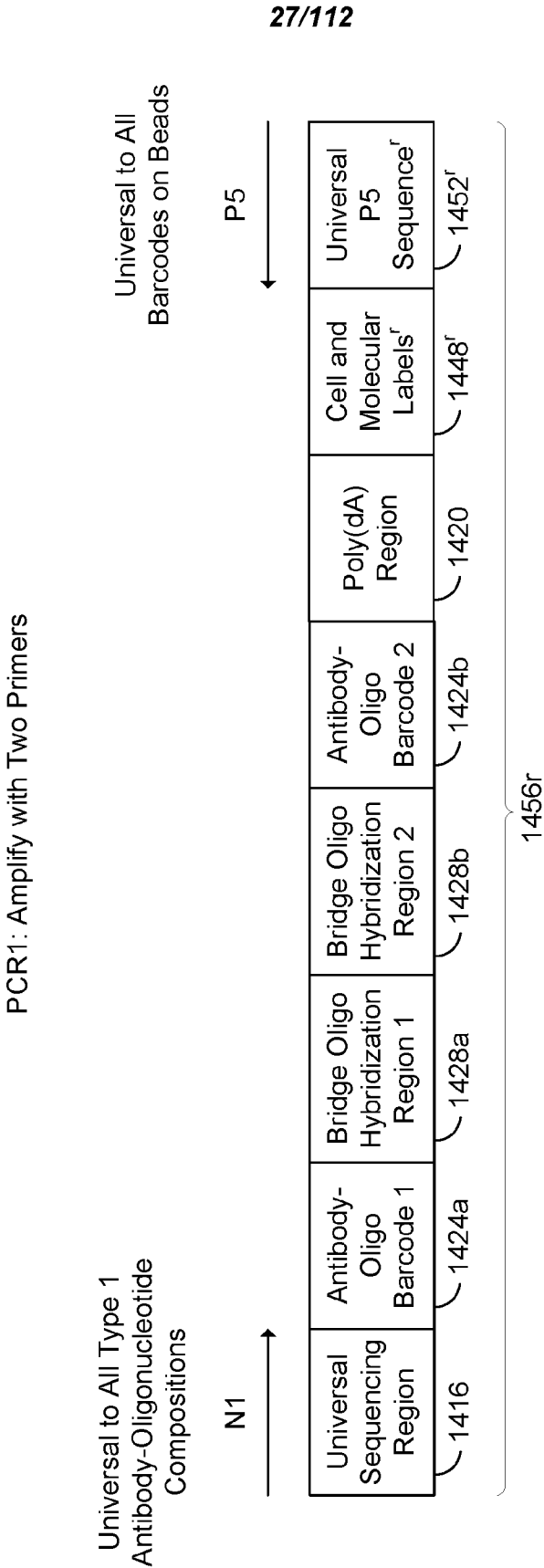


FIG. 14E

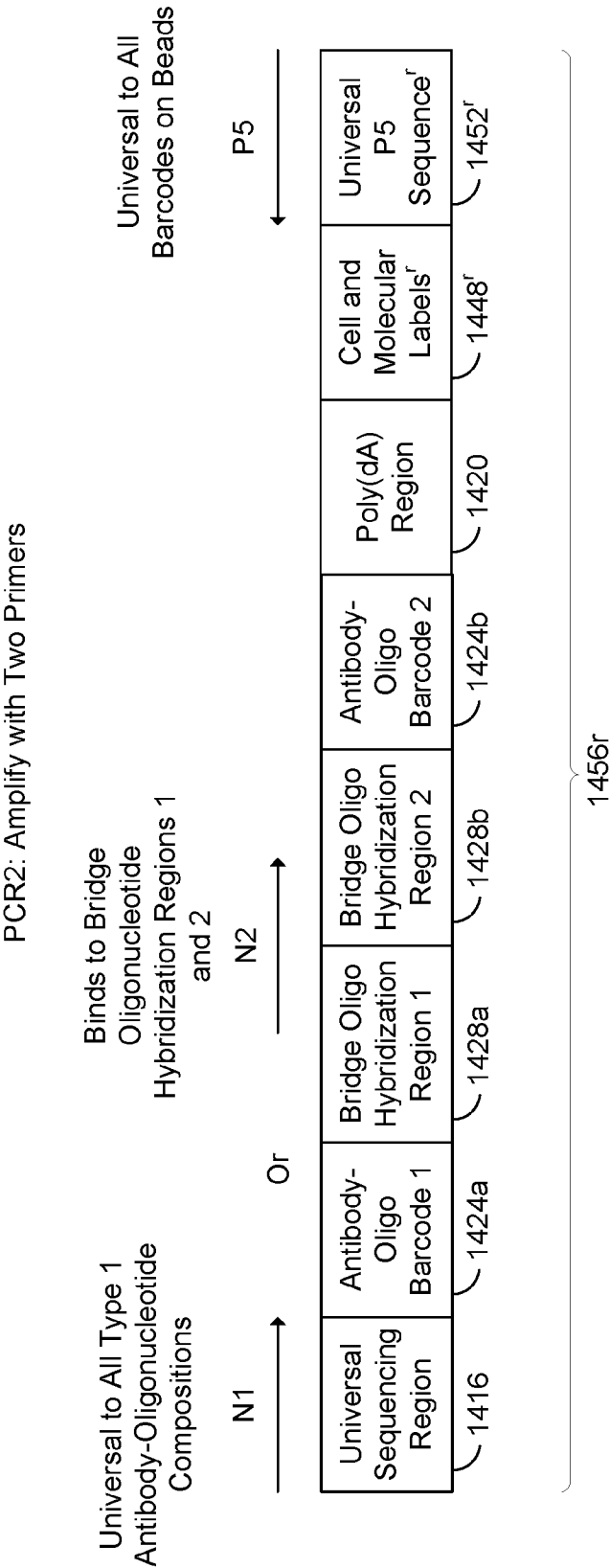
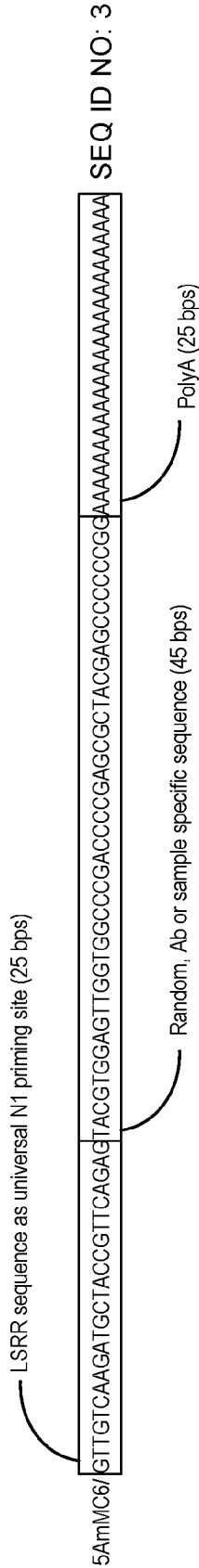


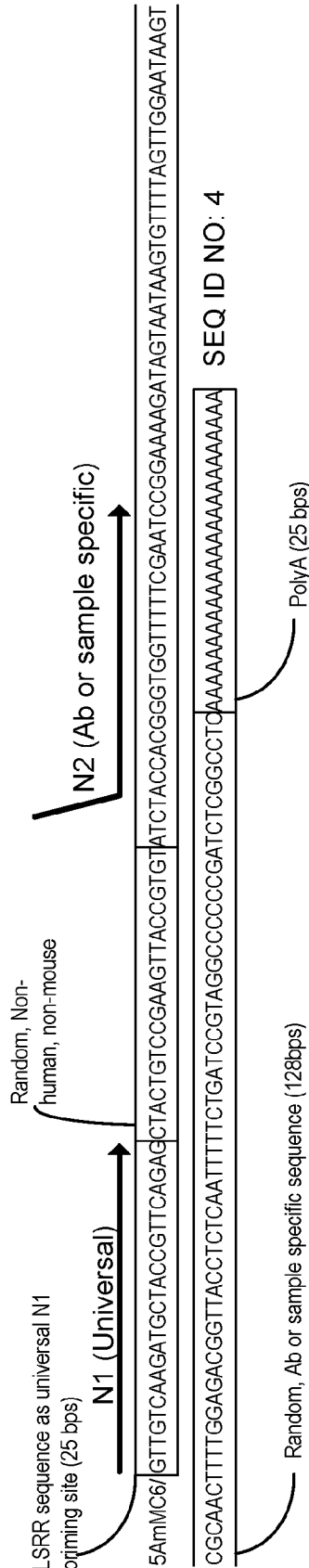
FIG. 14F



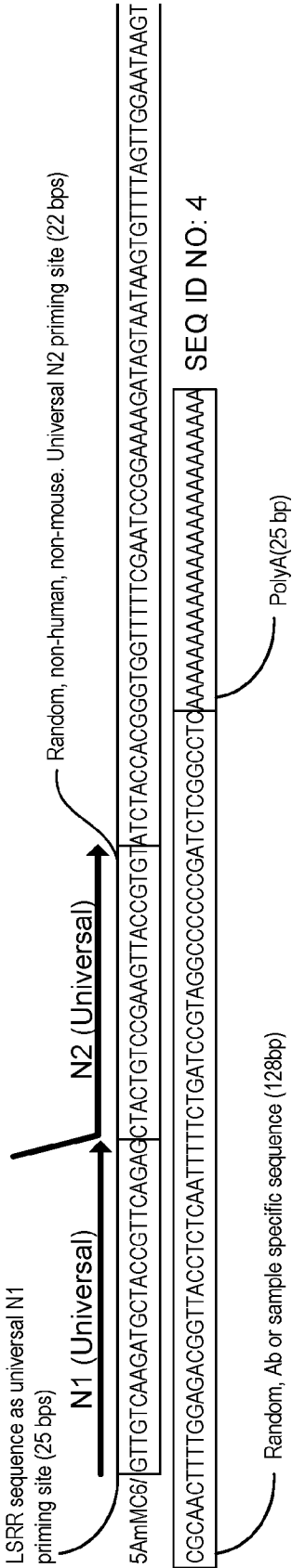
**FIG. 15A** 95mer oligo design.



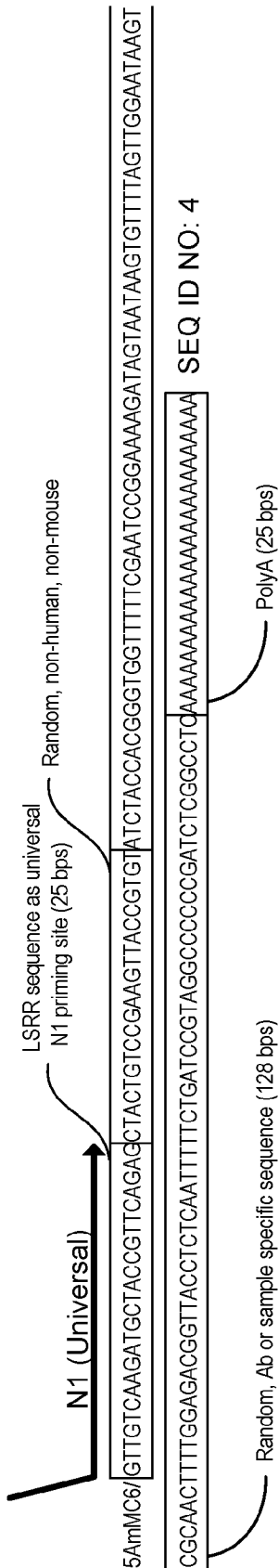
**FIG. 15B** 200mer oligo design. Targeted amplification scheme (nested specific)



**FIG. 15C** 200mer oligo design. Targeted amplification scheme (nested universal)



**FIG. 15D** 200mer oligo design. Targeted amplification scheme (one step)



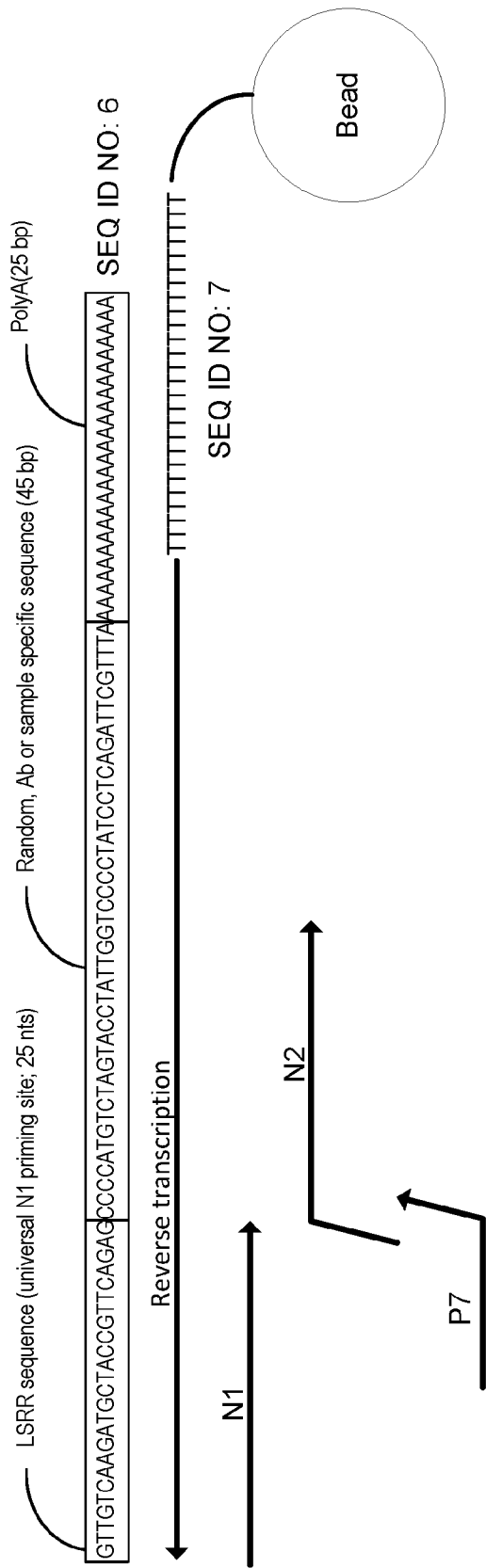
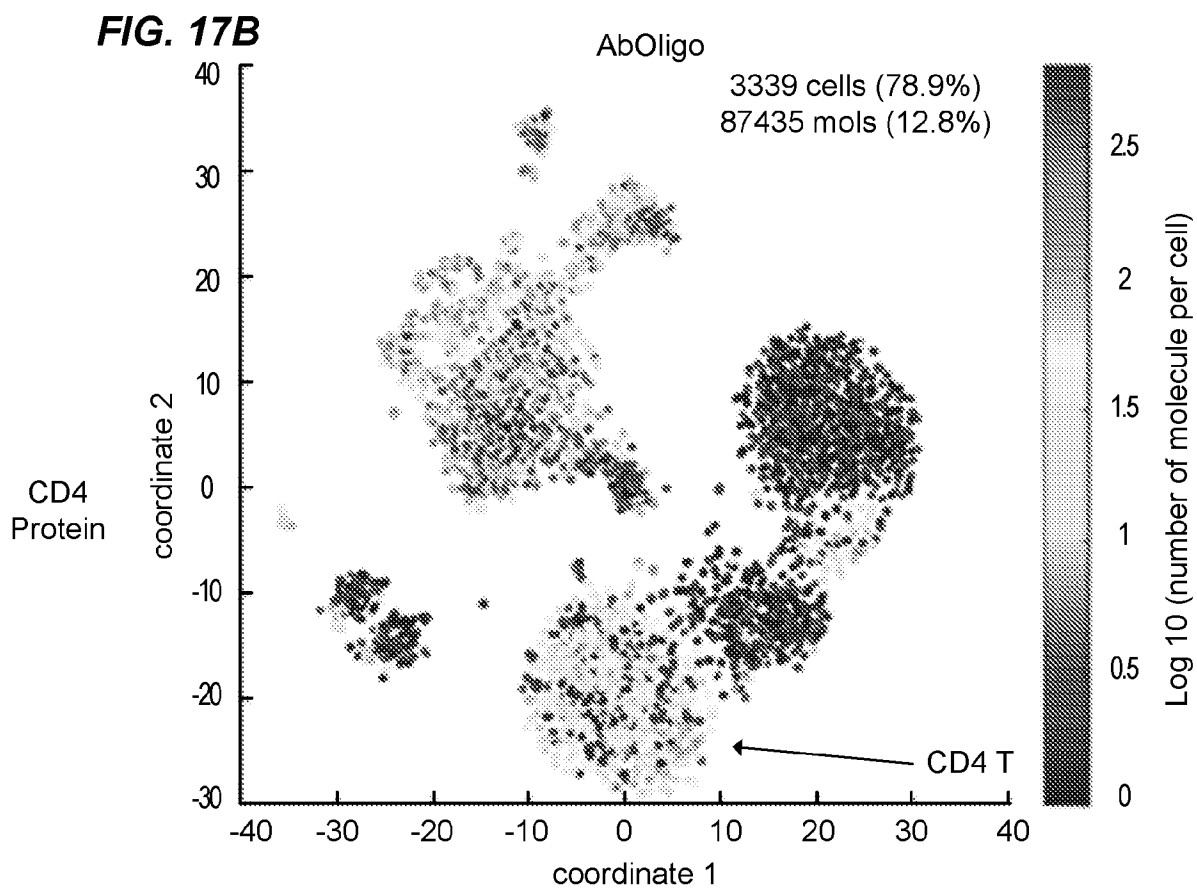
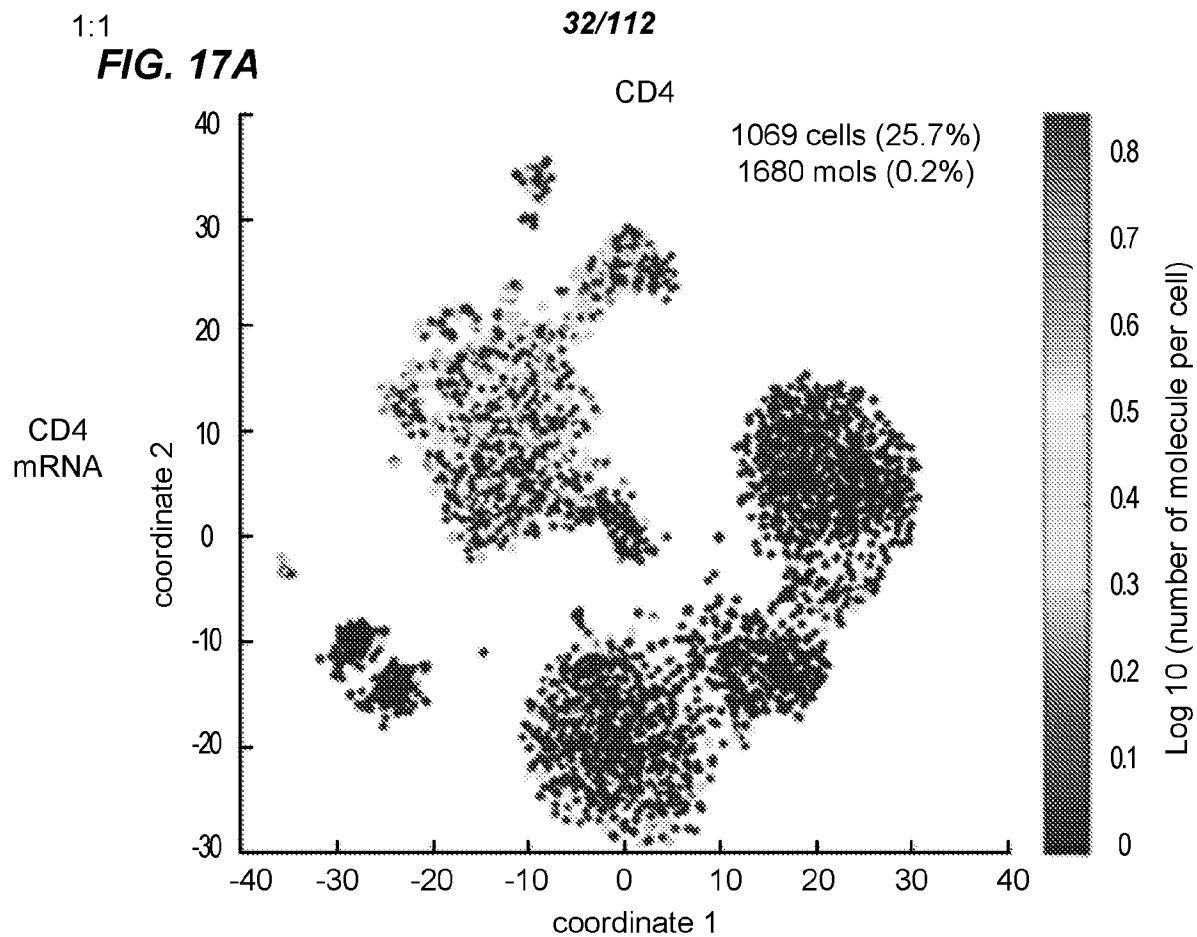
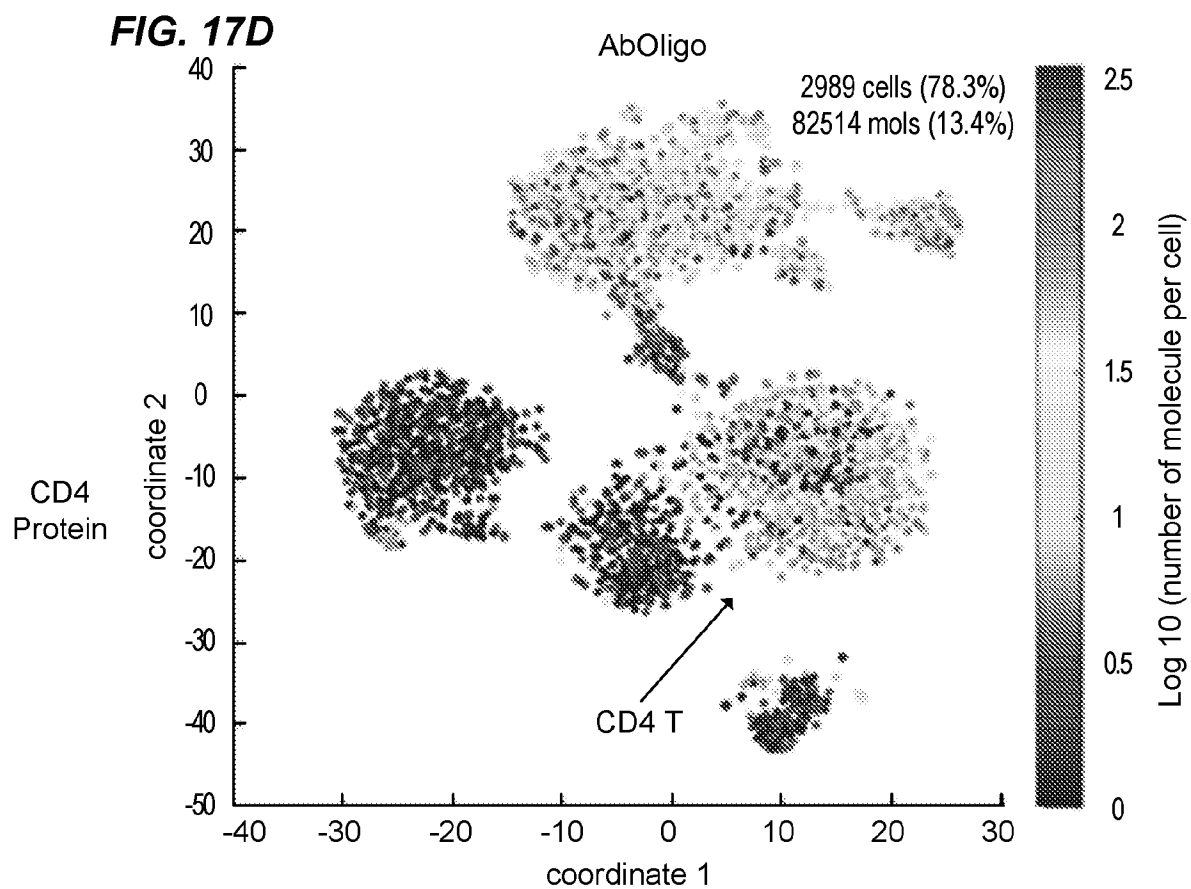
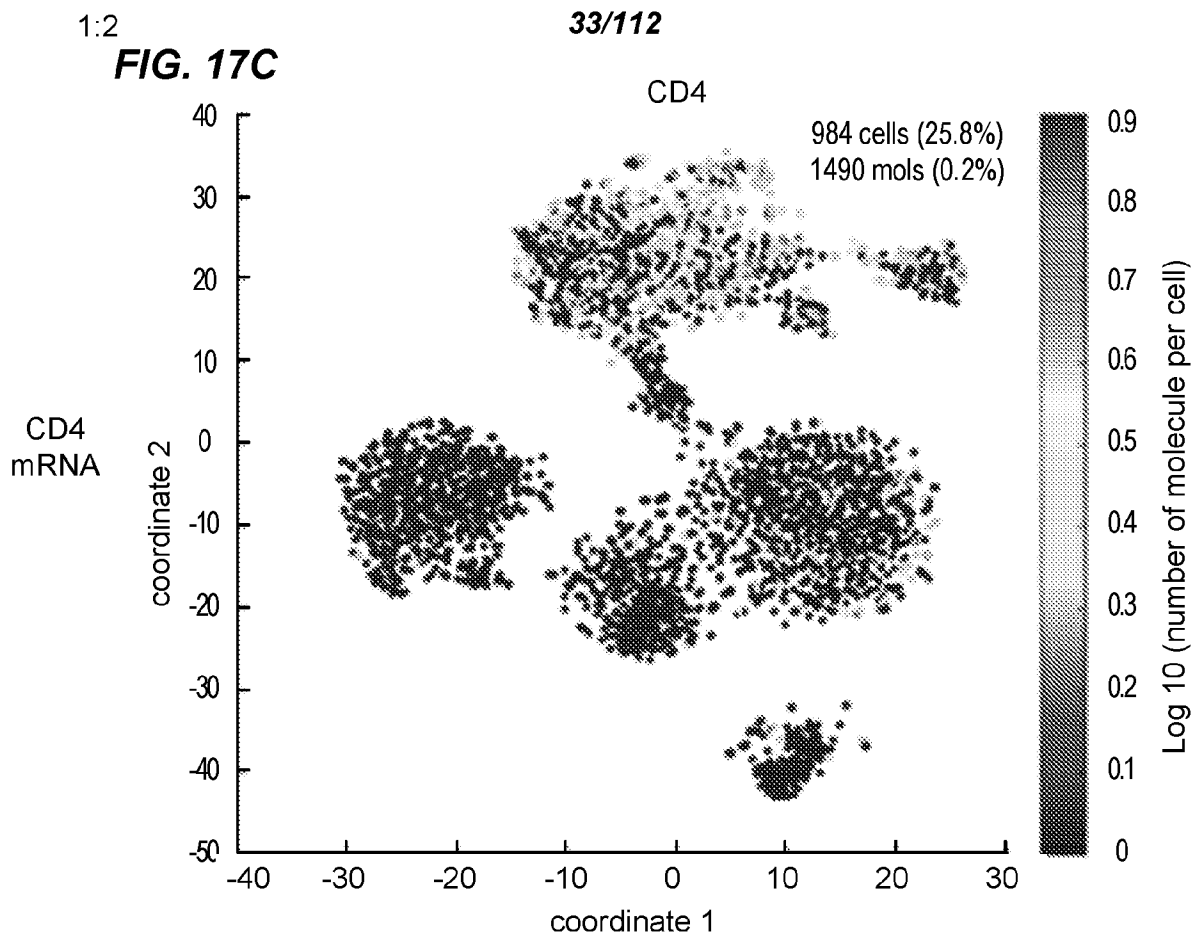
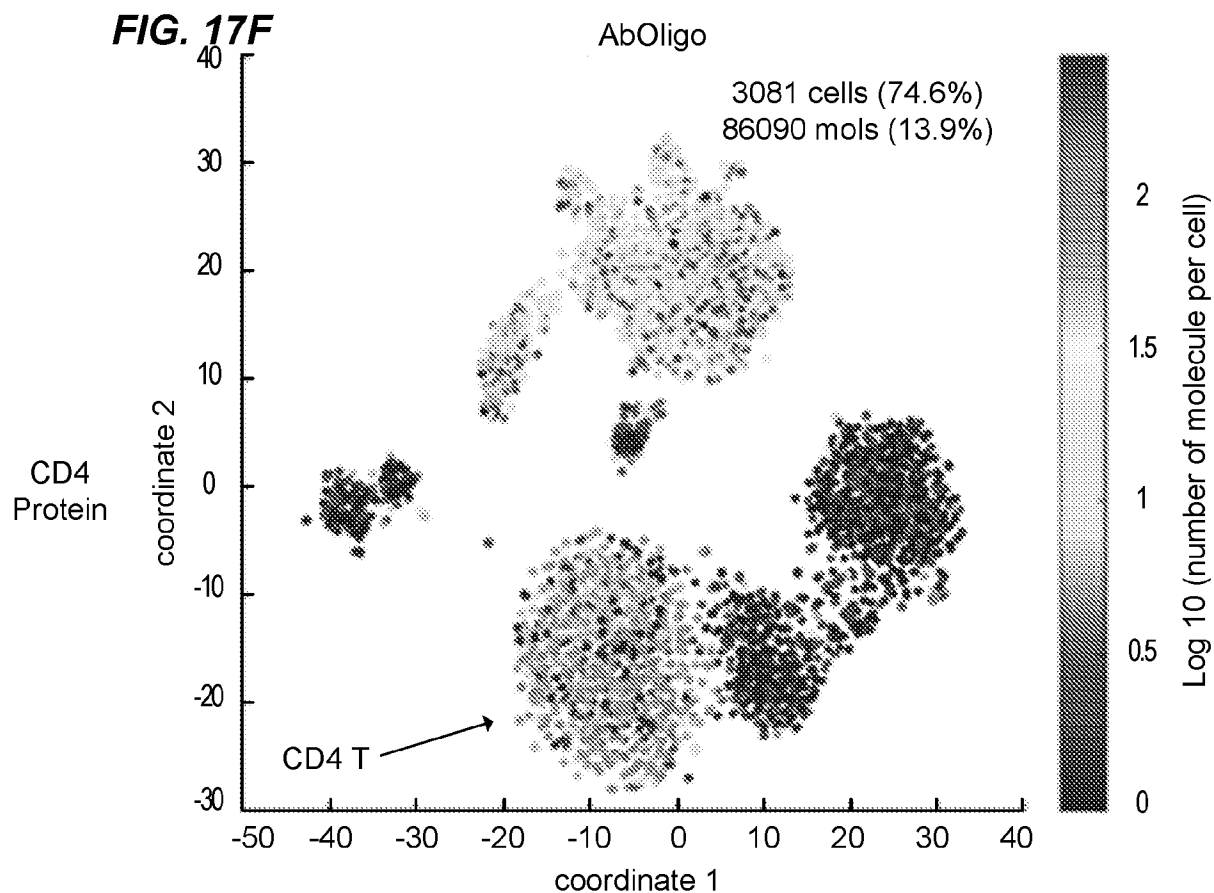
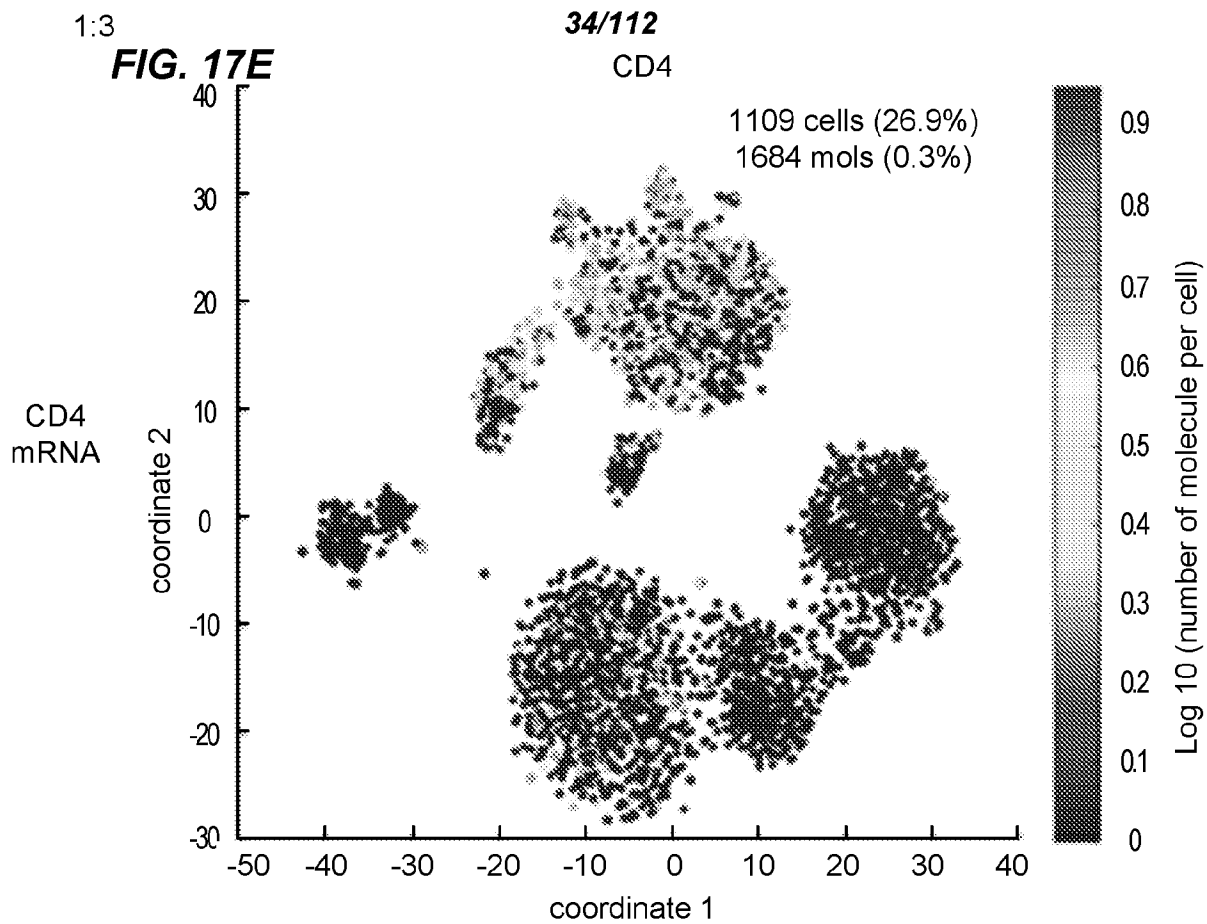


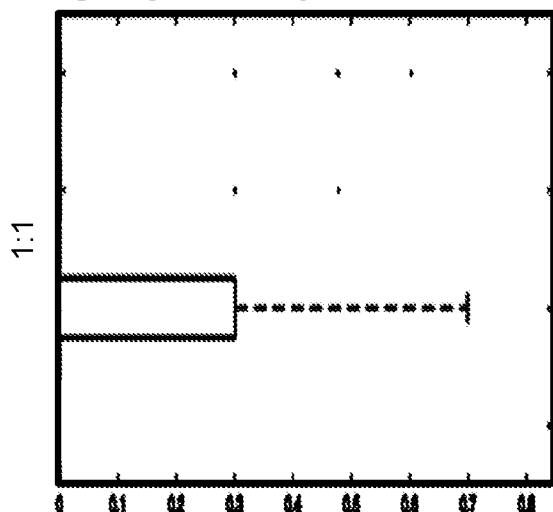
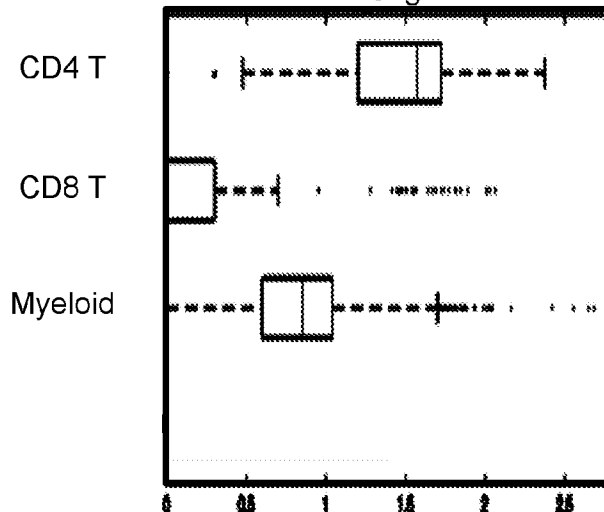
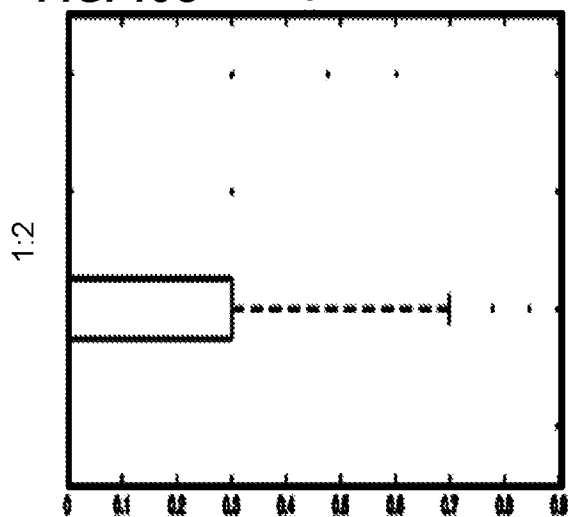
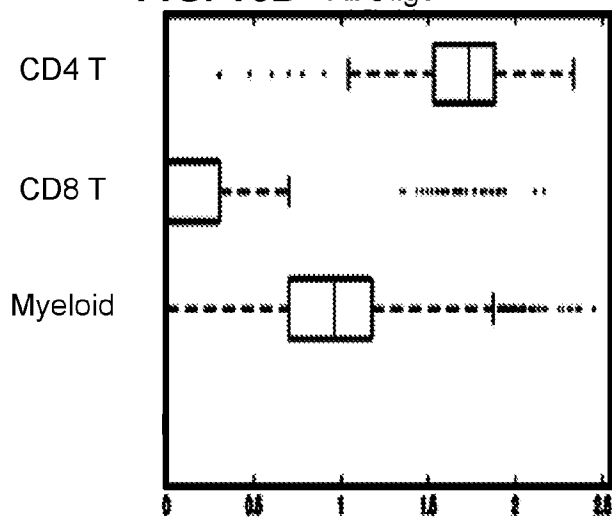
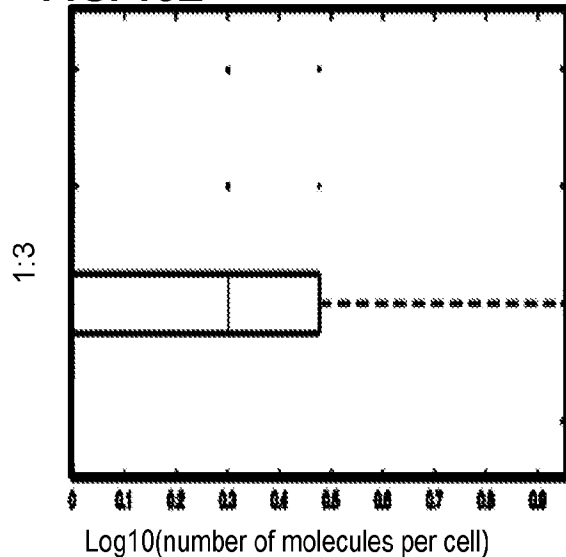
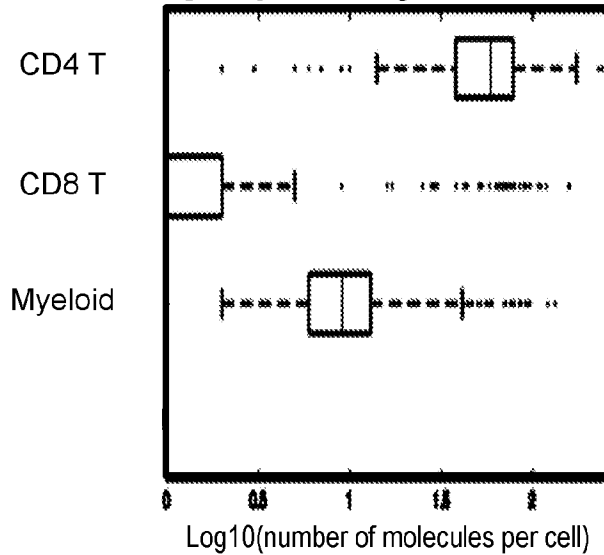
FIG. 16







35/112

**FIG. 18A** CD4 mRNA  
CD4**FIG. 18B** CD4 protein  
AbOligo**FIG. 18C** CD4**FIG. 18D** AbOligo**FIG. 18E** CD4**FIG. 18F** AbOligo

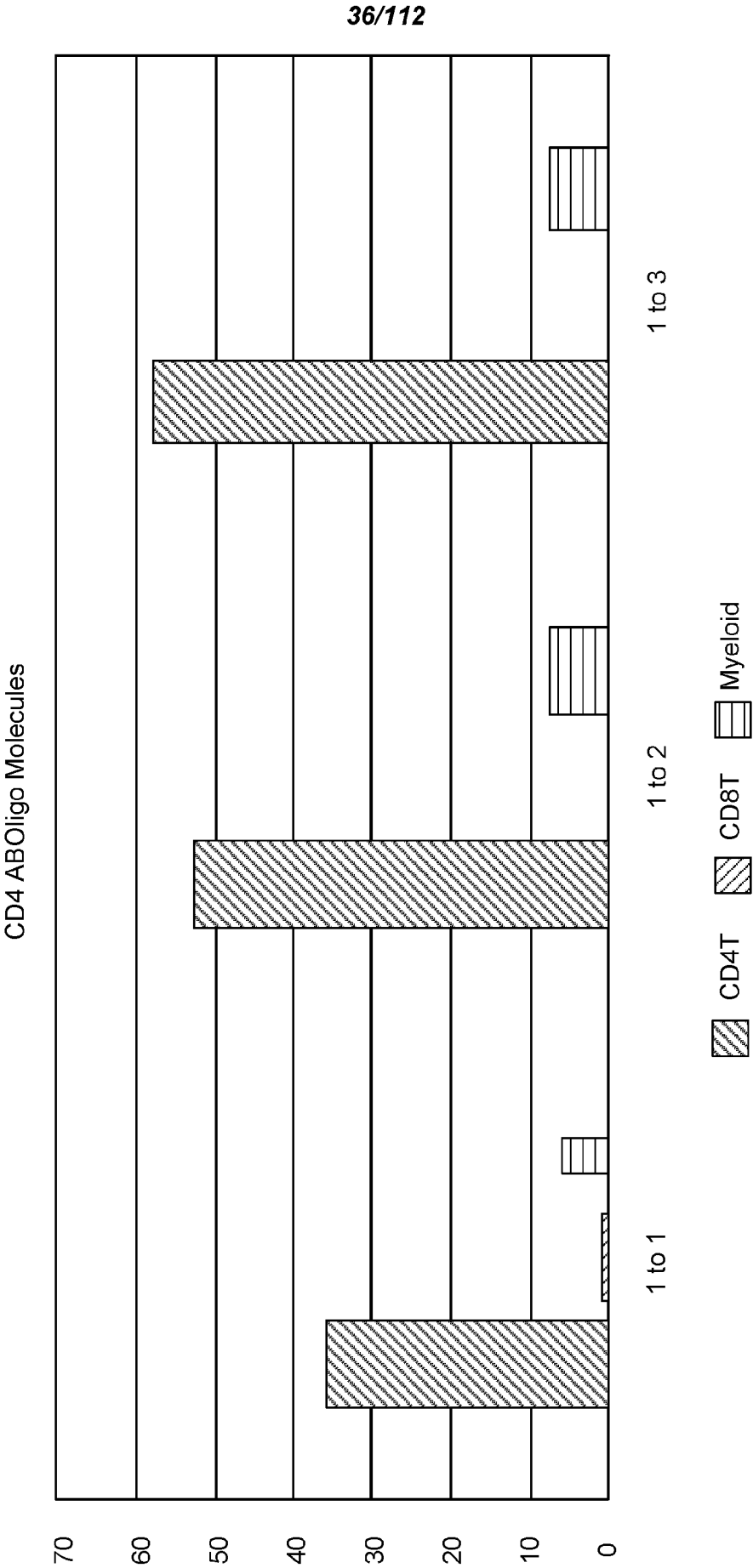
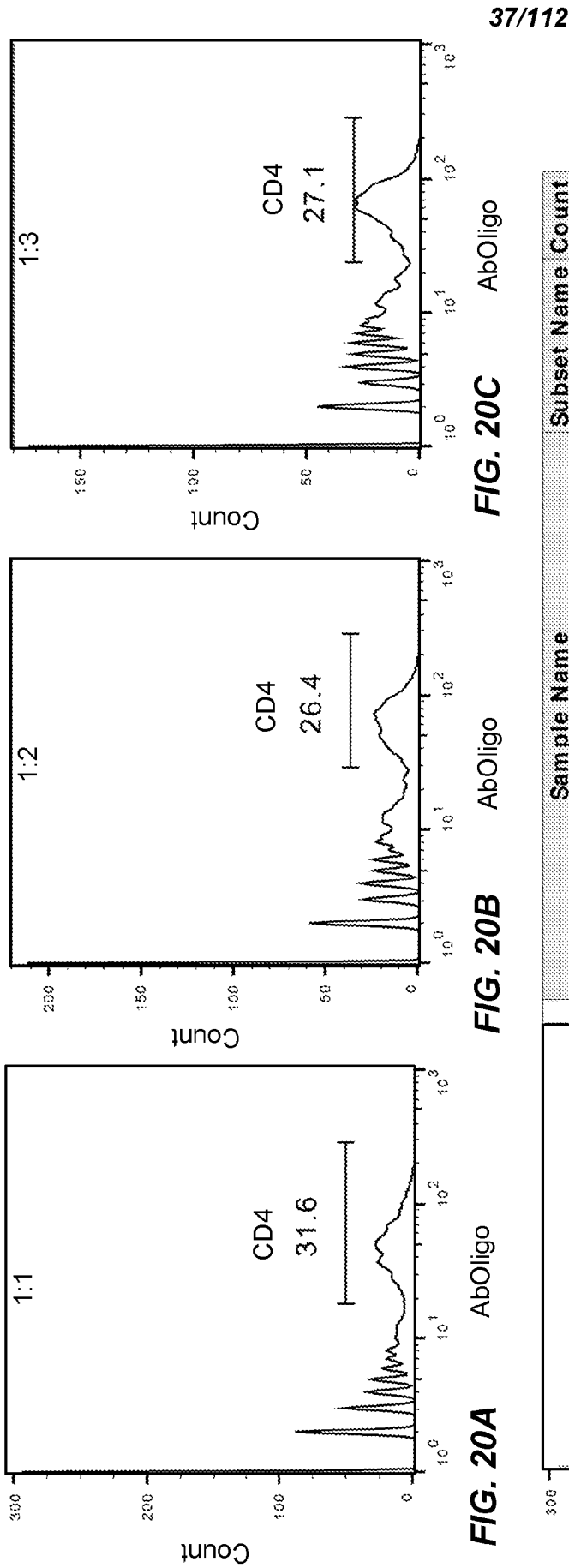
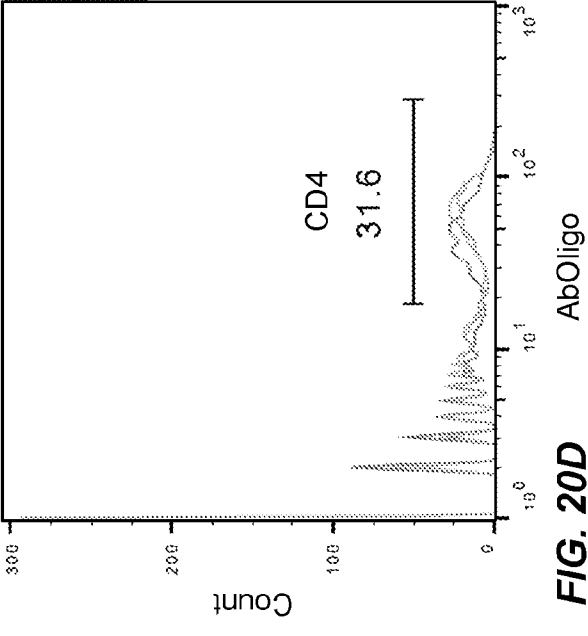


FIG. 19



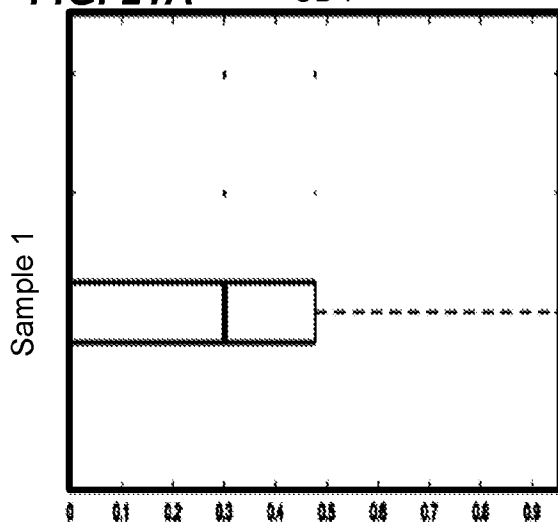


| Sample Name  | Subset Name | Count |
|--|-------------|-------|
| <input type="checkbox"/> OligoOnly1to3_Corrected_AbOligo_MolsPerCell.fcs | Ungated     | 4128  |
| <input type="checkbox"/> OligoOnly1to2_Corrected_AbOligo_MolsPerCell.fcs | Ungated     | 3819  |
| <input type="checkbox"/> OligoOnly1to1_Corrected_AbOligo_MolsPerCell.fcs | Ungated     | 4234  |

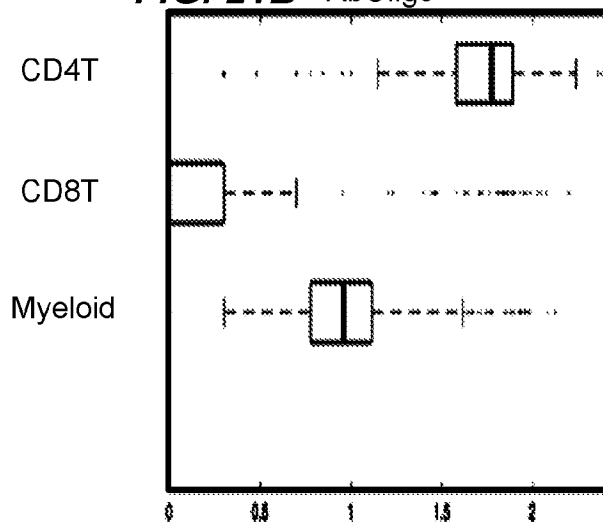


38/112

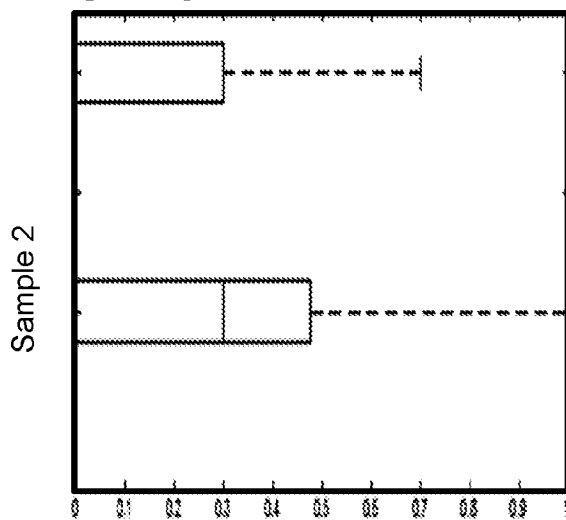
**FIG. 21A** CD4 mRNA  
CD4



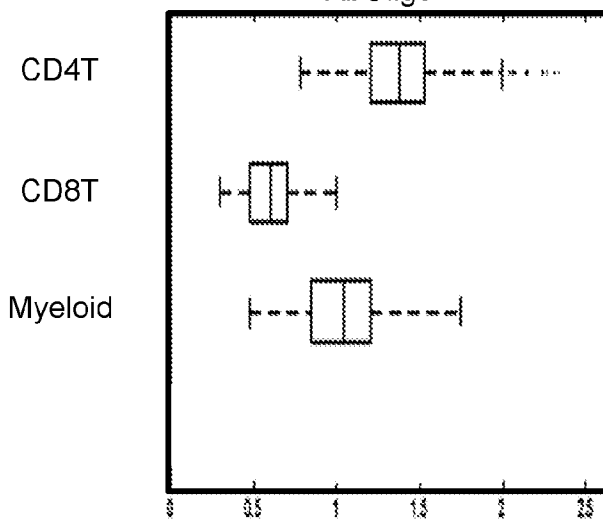
**FIG. 21B** CD4 protein  
AbOligo



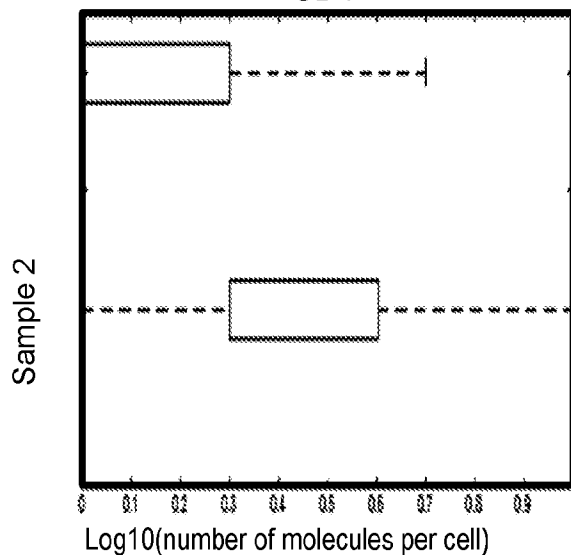
**FIG. 21C** CD4



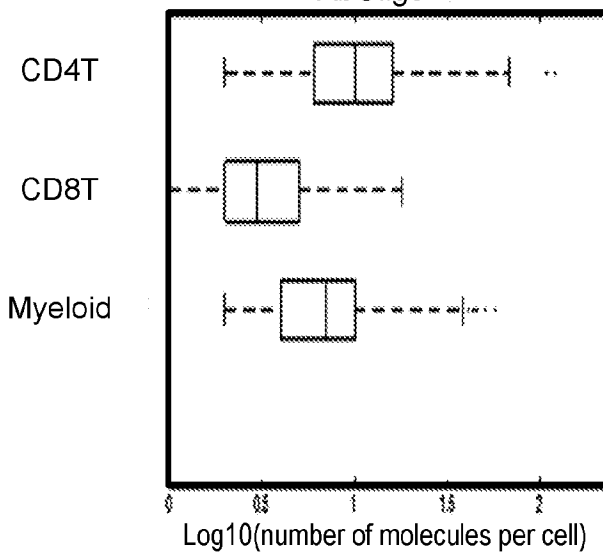
**FIG. 21D** AbOligo



**FIG. 21E** CD4



**FIG. 21F** AbOligo



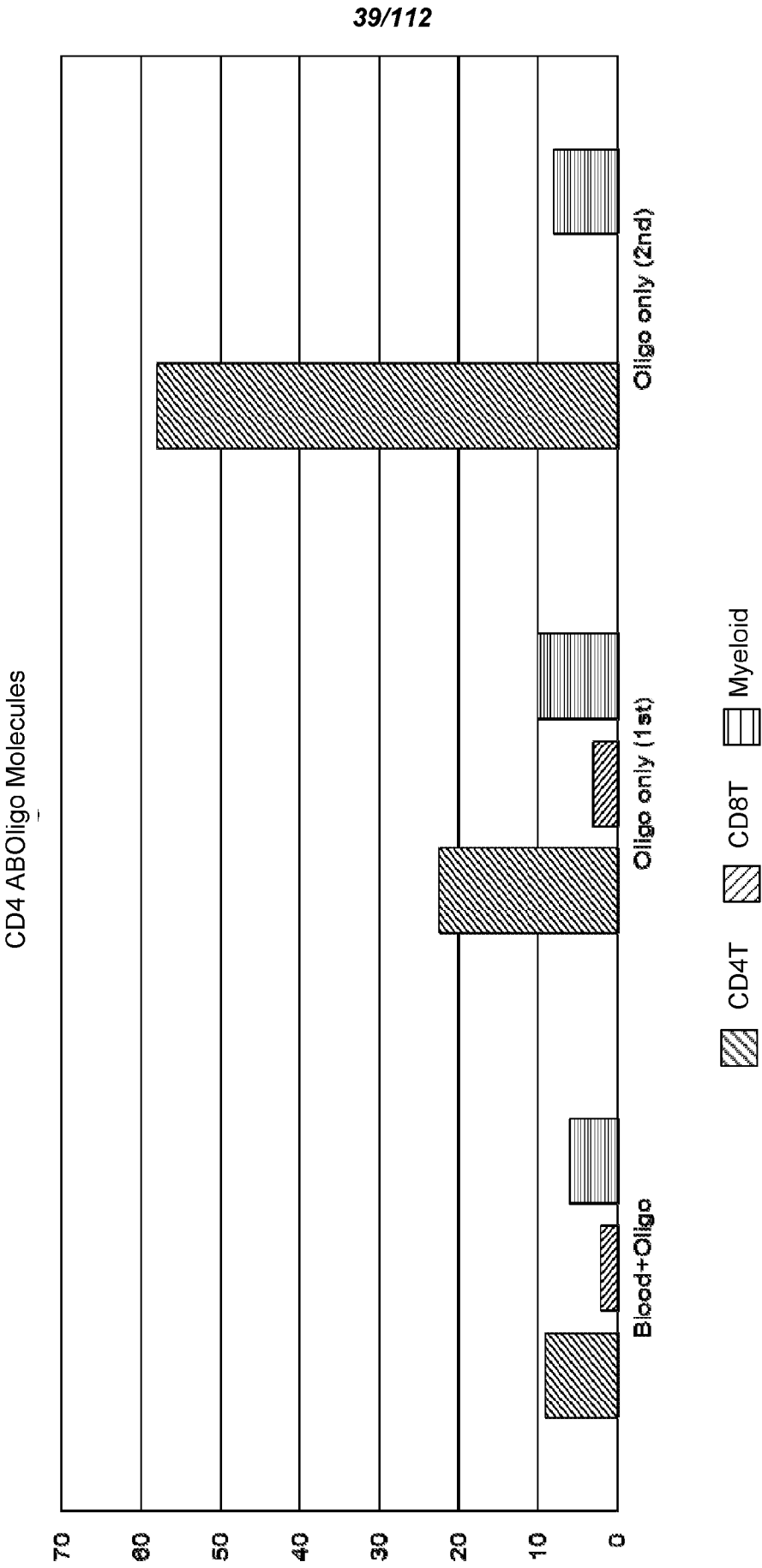


FIG. 22

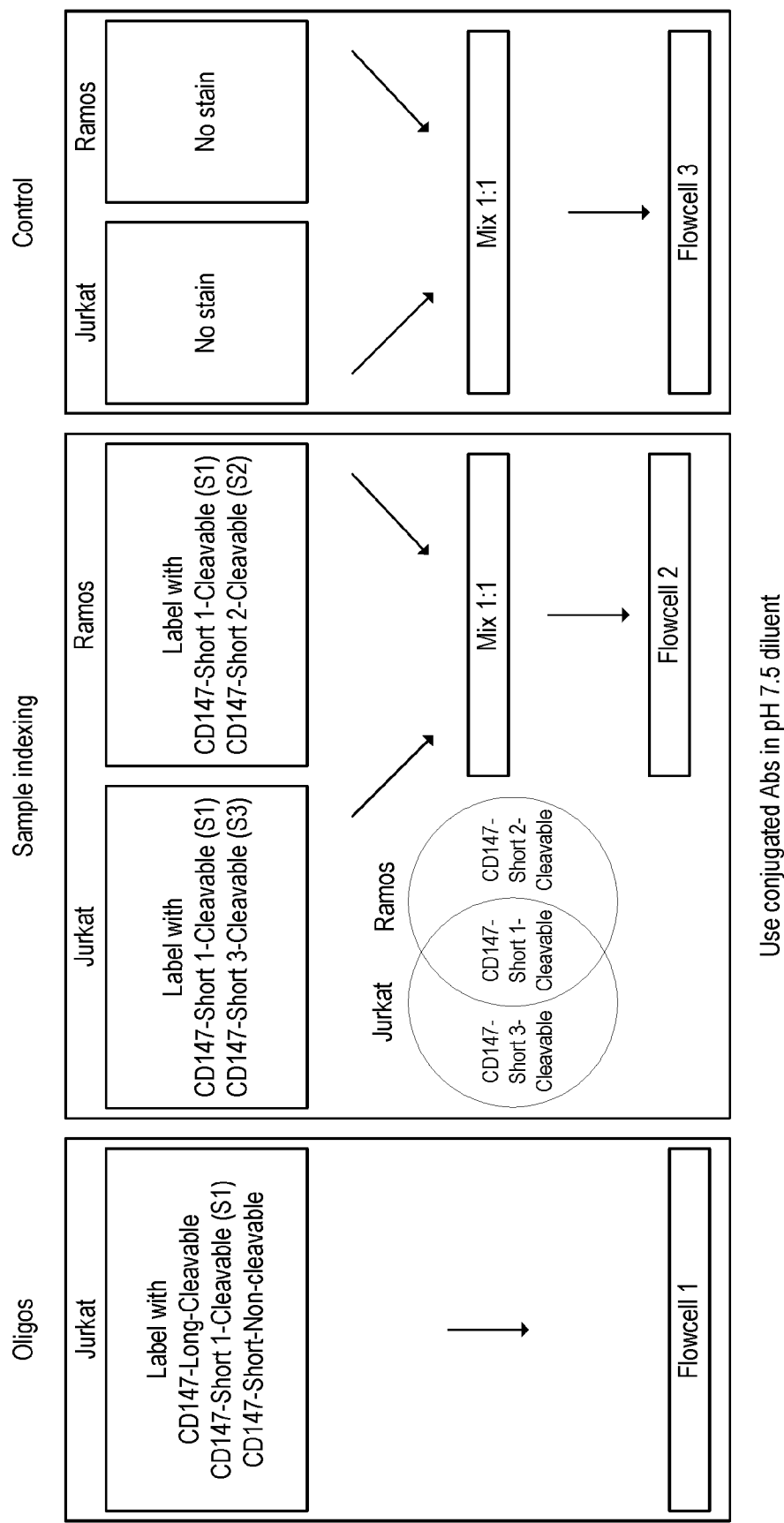
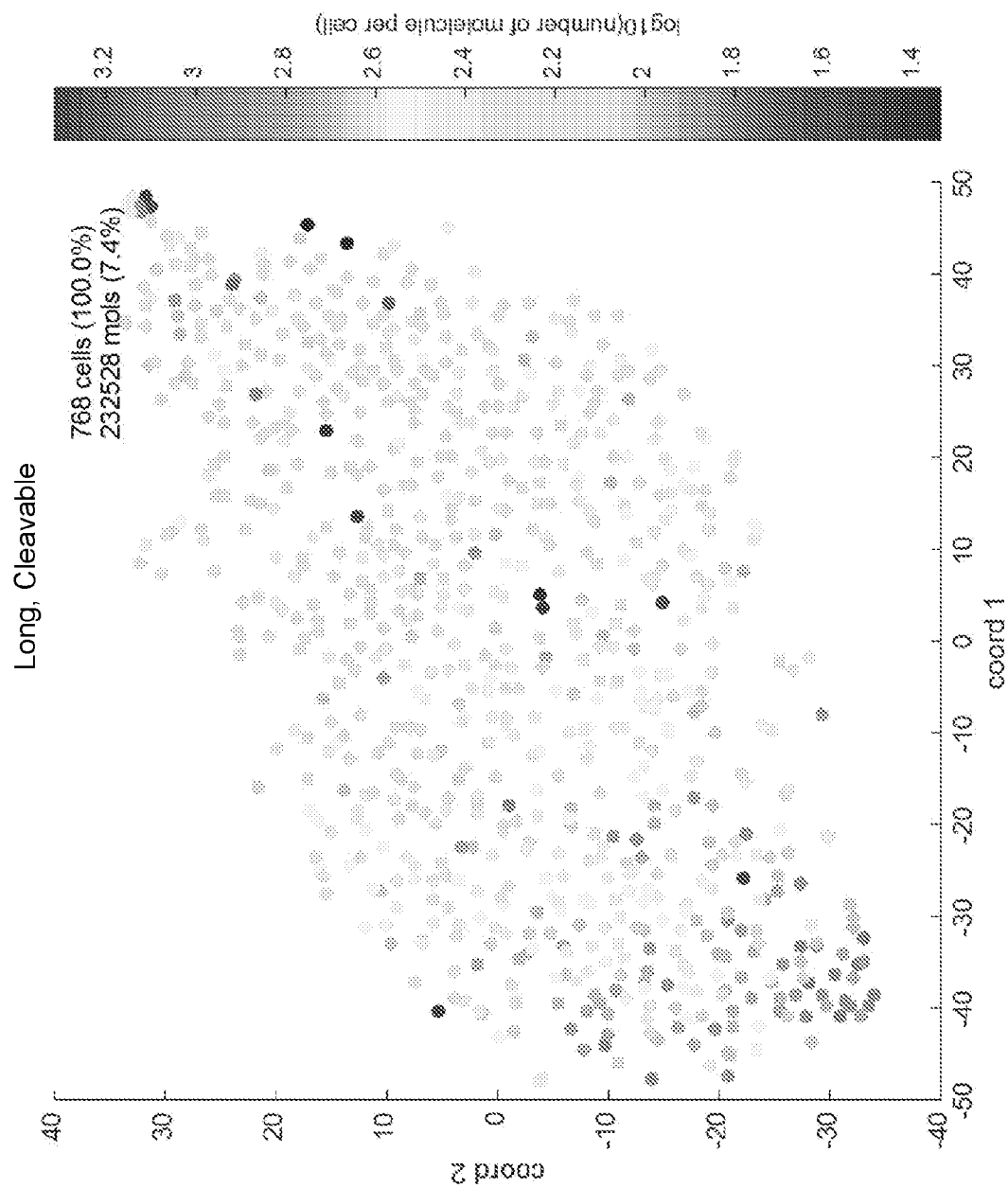


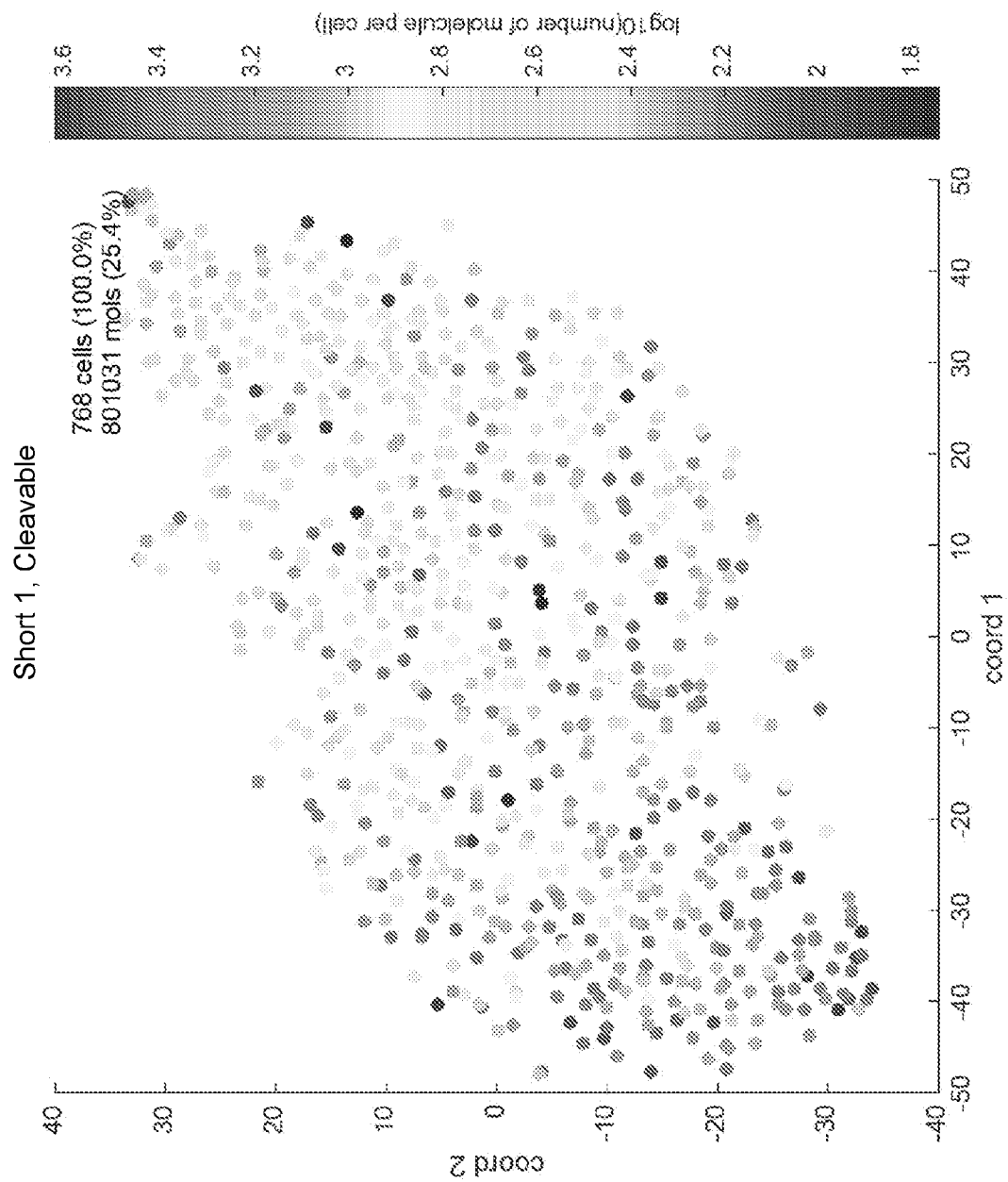
FIG. 23

41/112



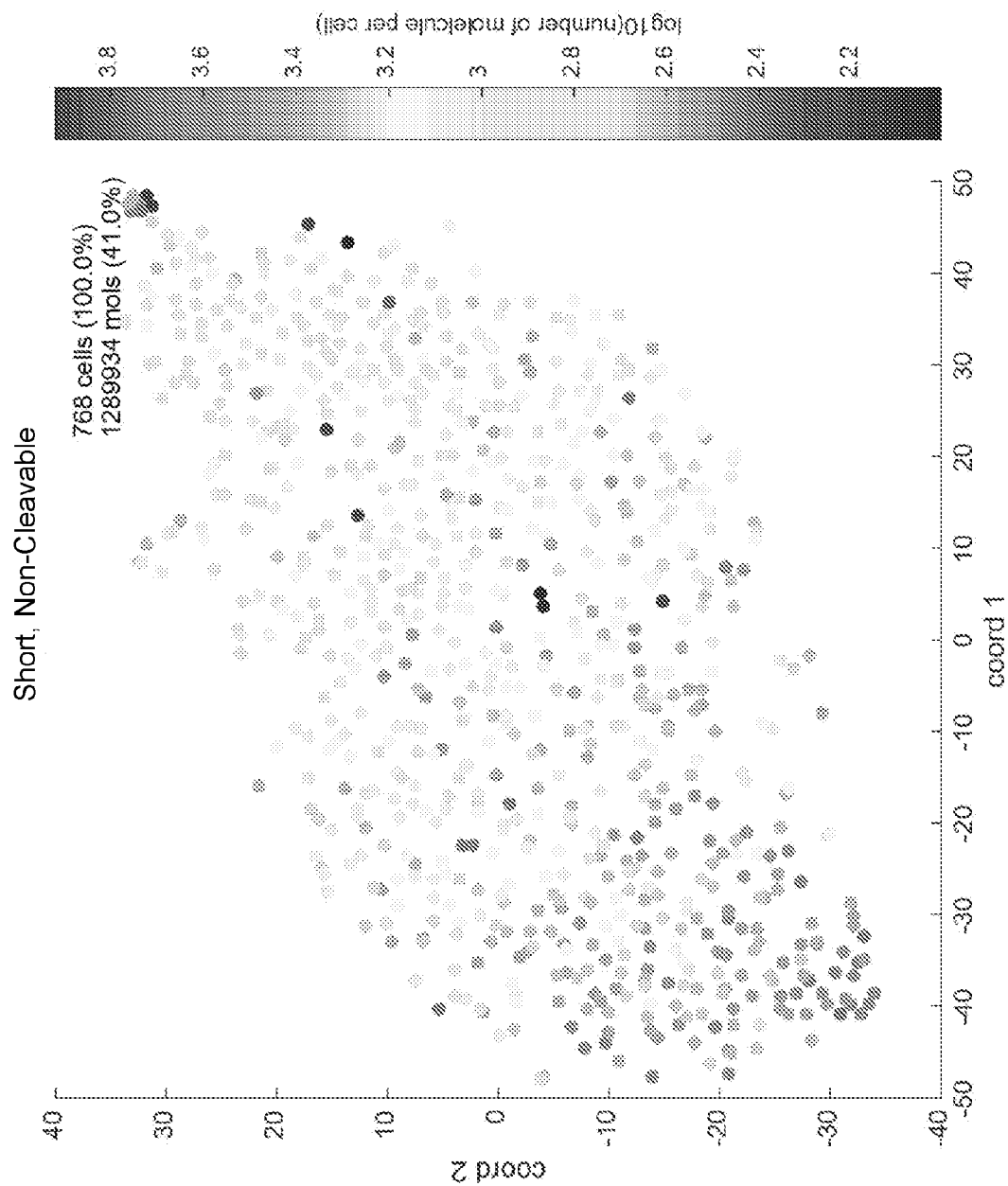
**FIG. 24A**

42/112



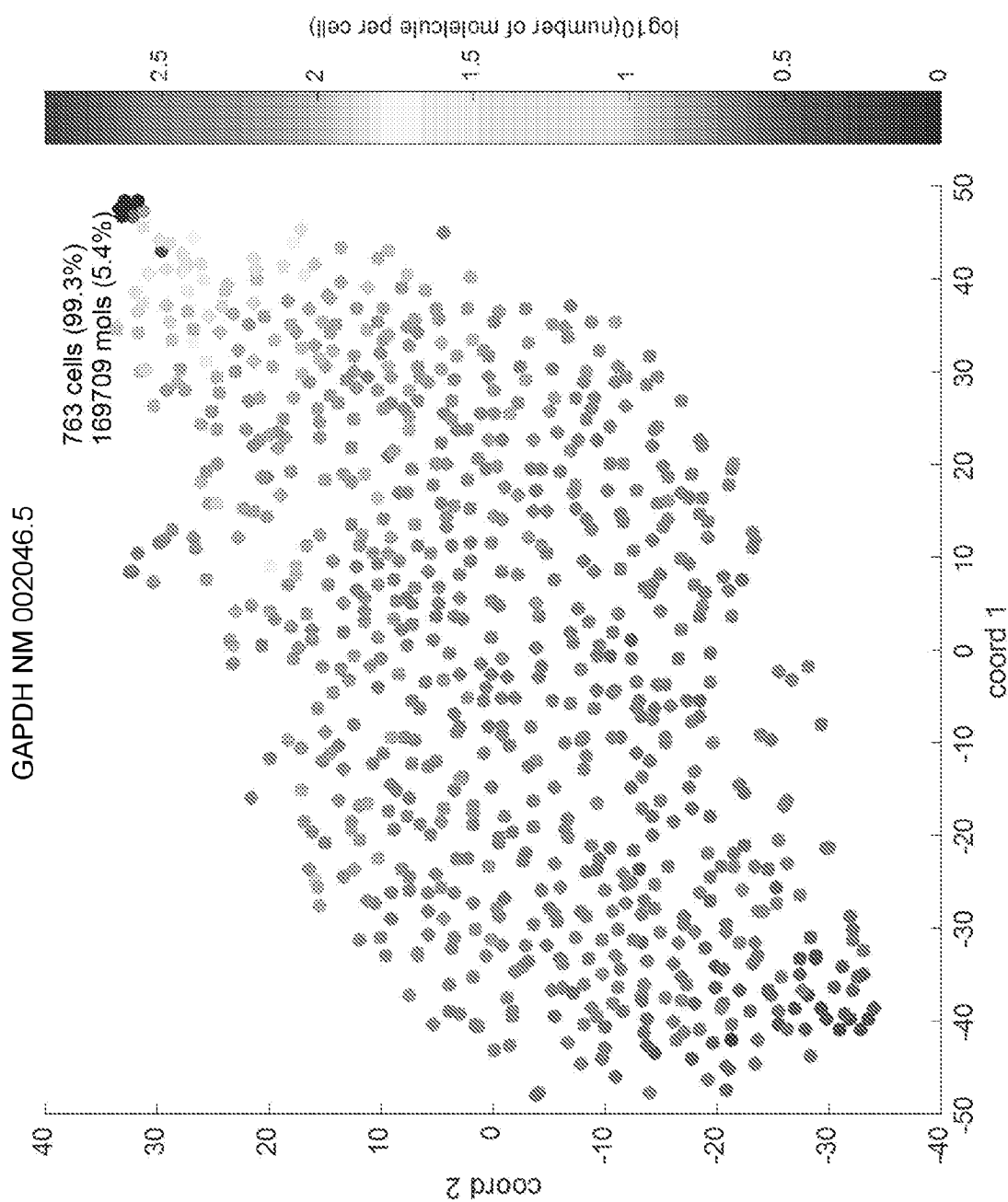
**FIG. 24B**

43/112



**FIG. 24C**

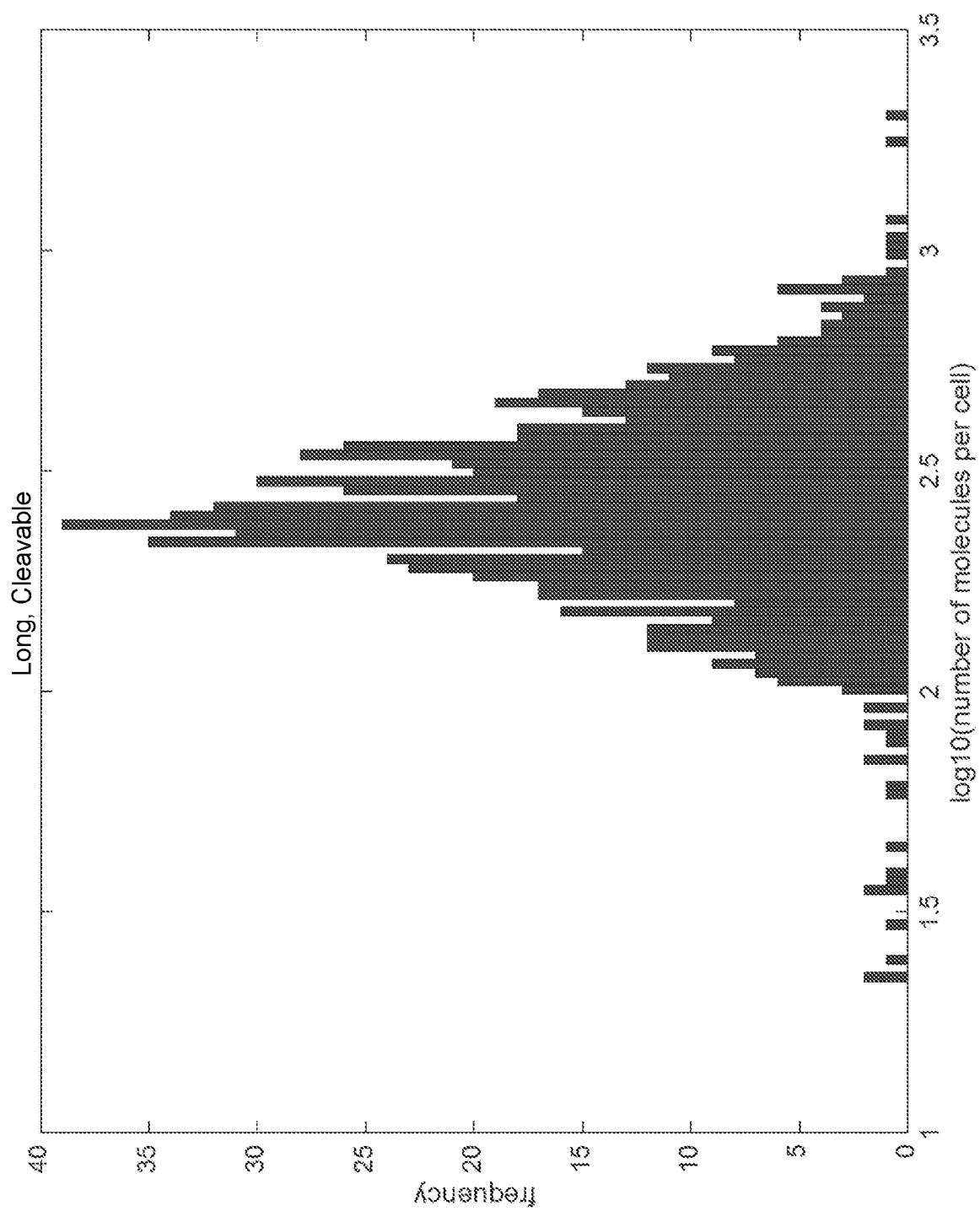
44/112



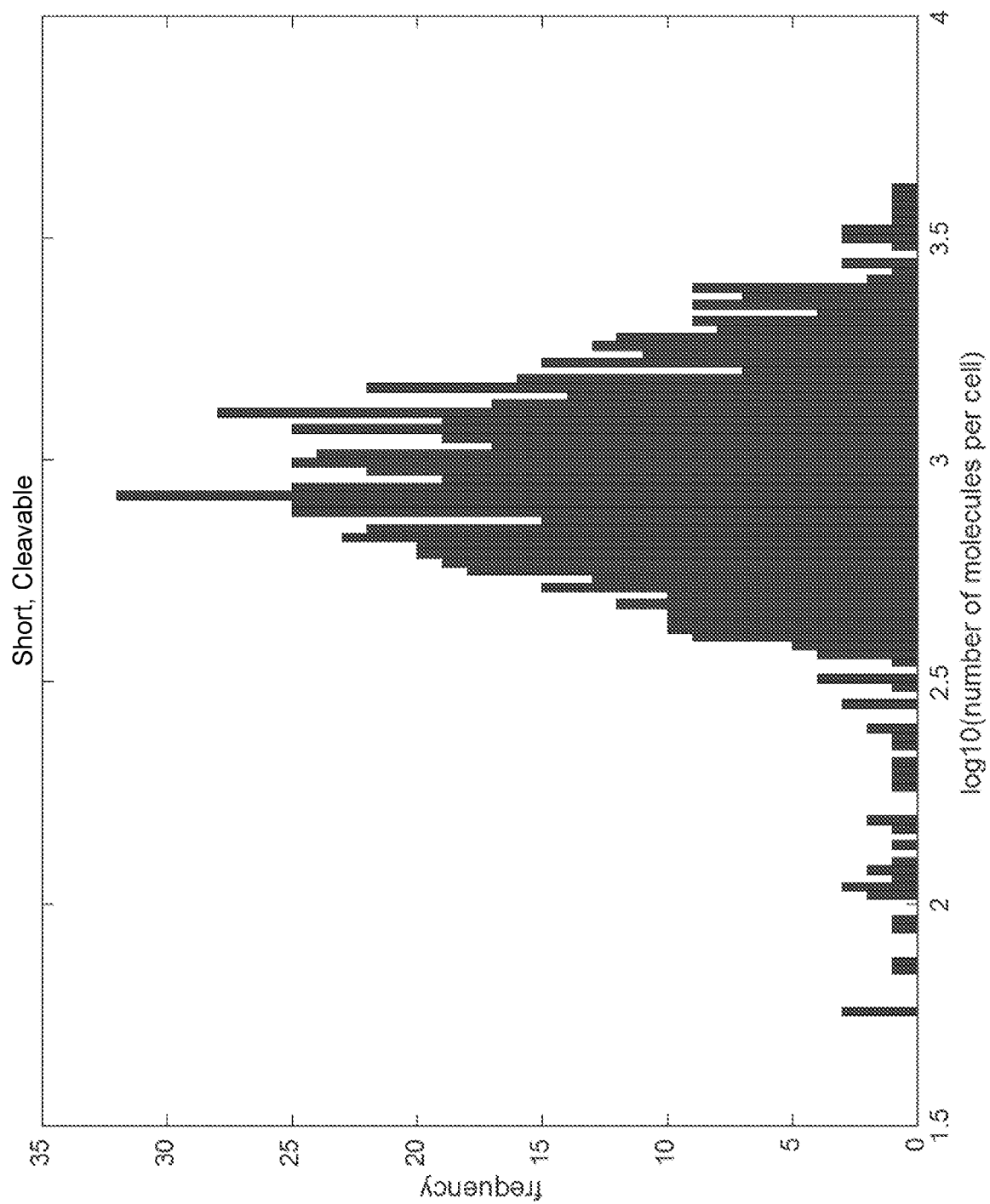
**FIG. 25**



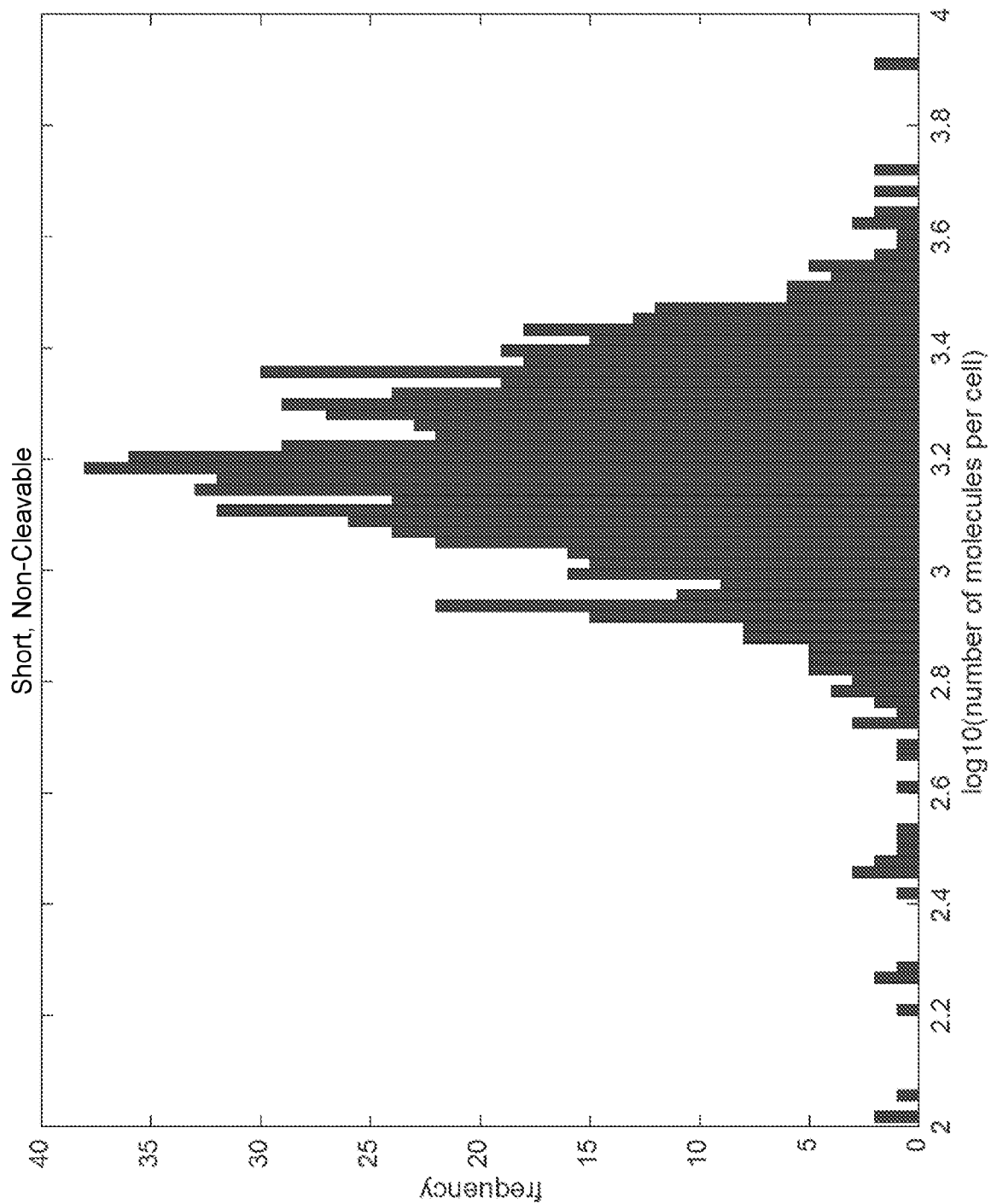
45/112

**FIG. 26A**

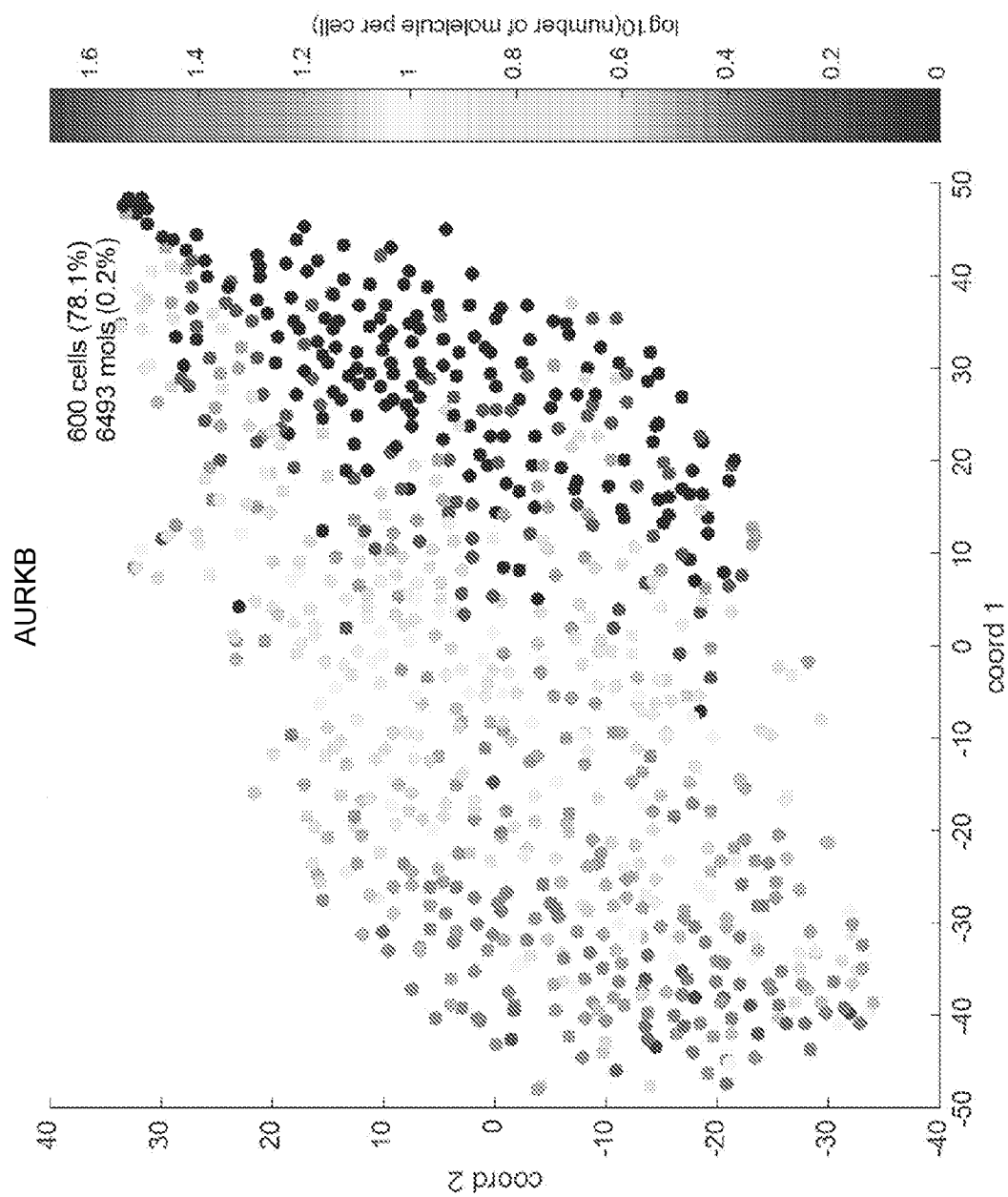
46/112

**FIG. 26B**

47/112

**FIG. 26C**

48/112

**FIG. 27A**

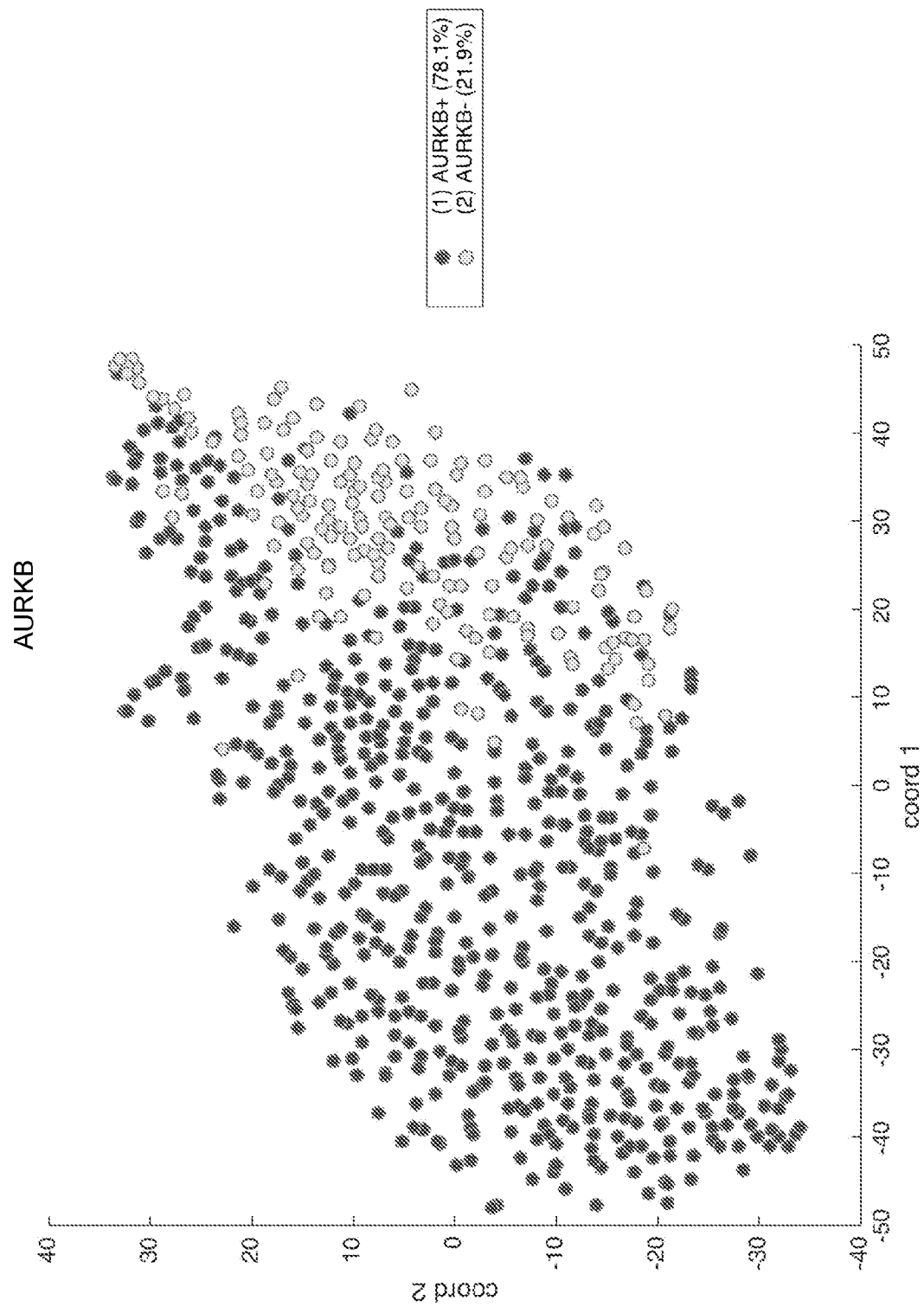
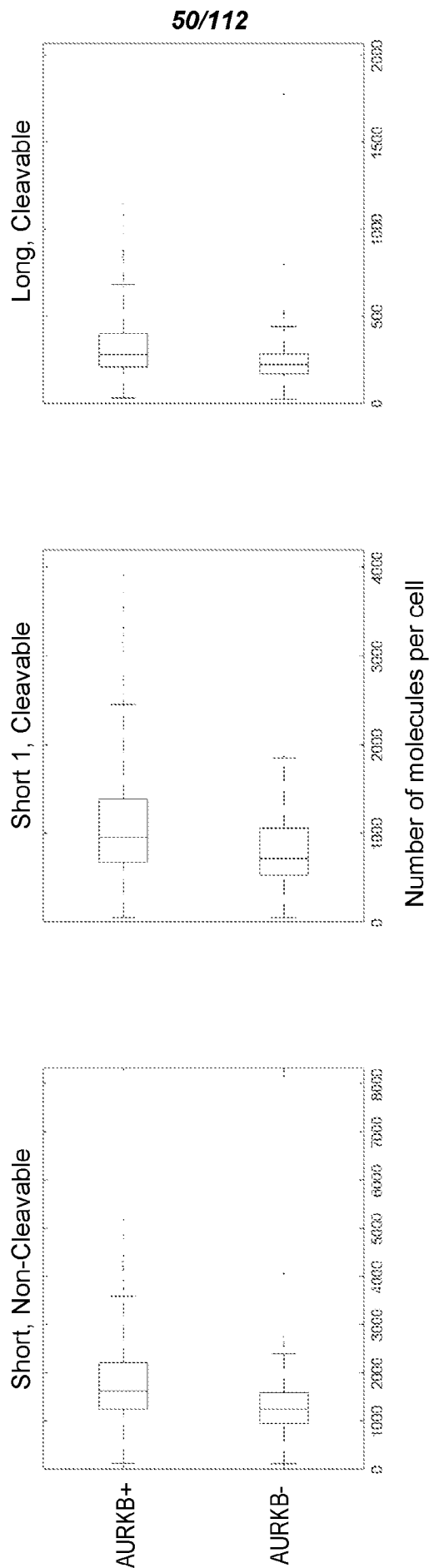
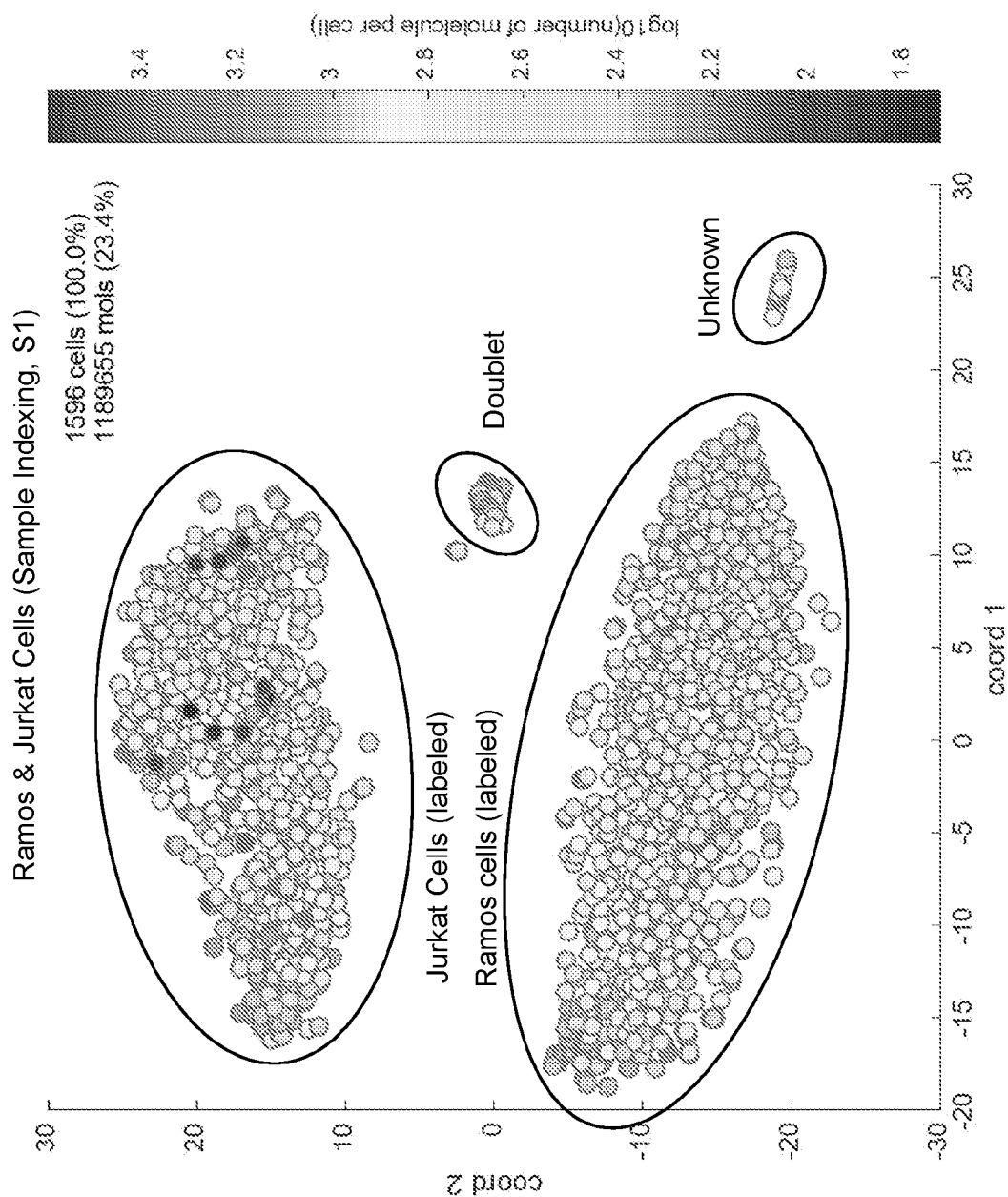


FIG. 27B



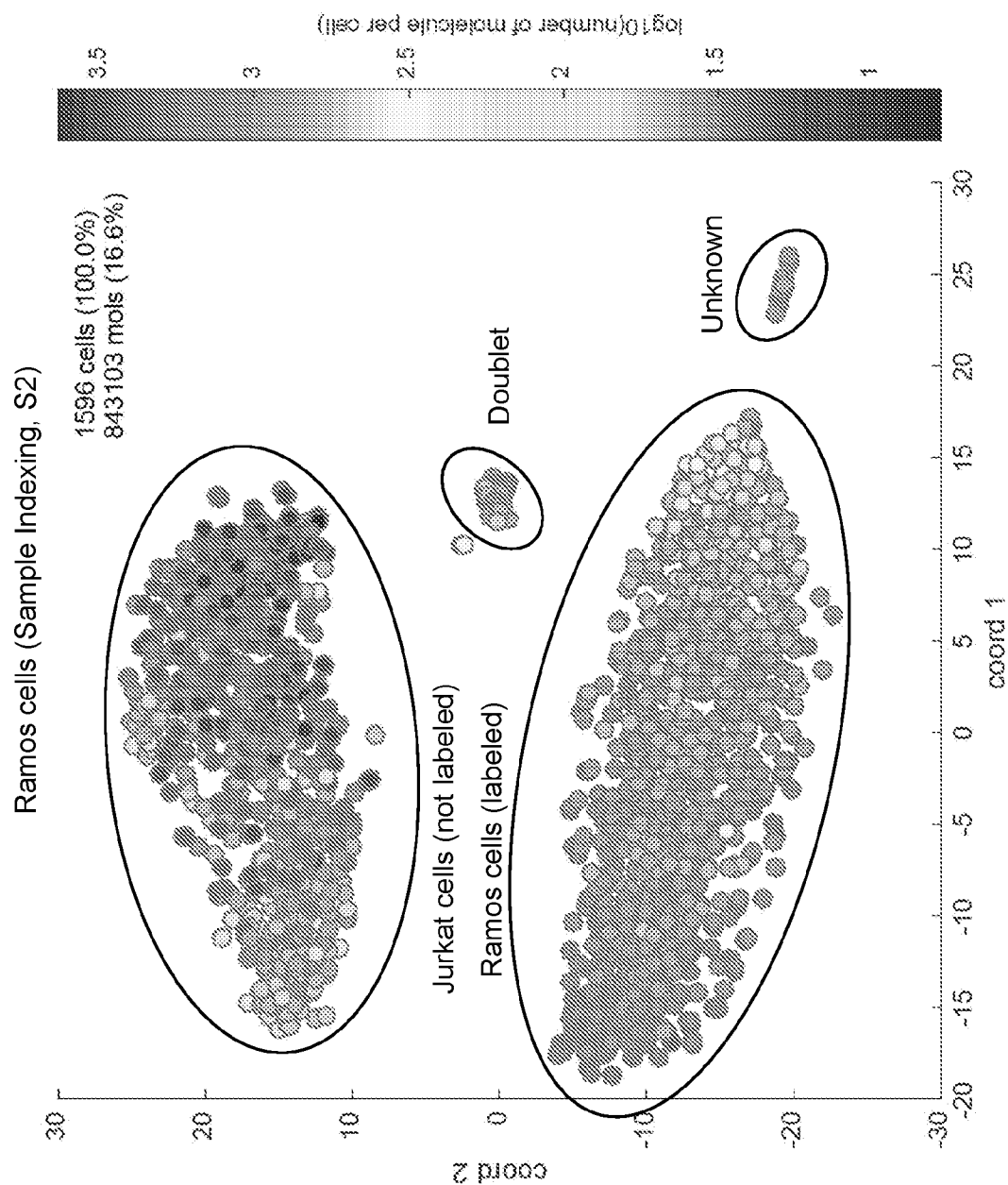
**FIG. 27C**

51/112



**FIG. 28A**

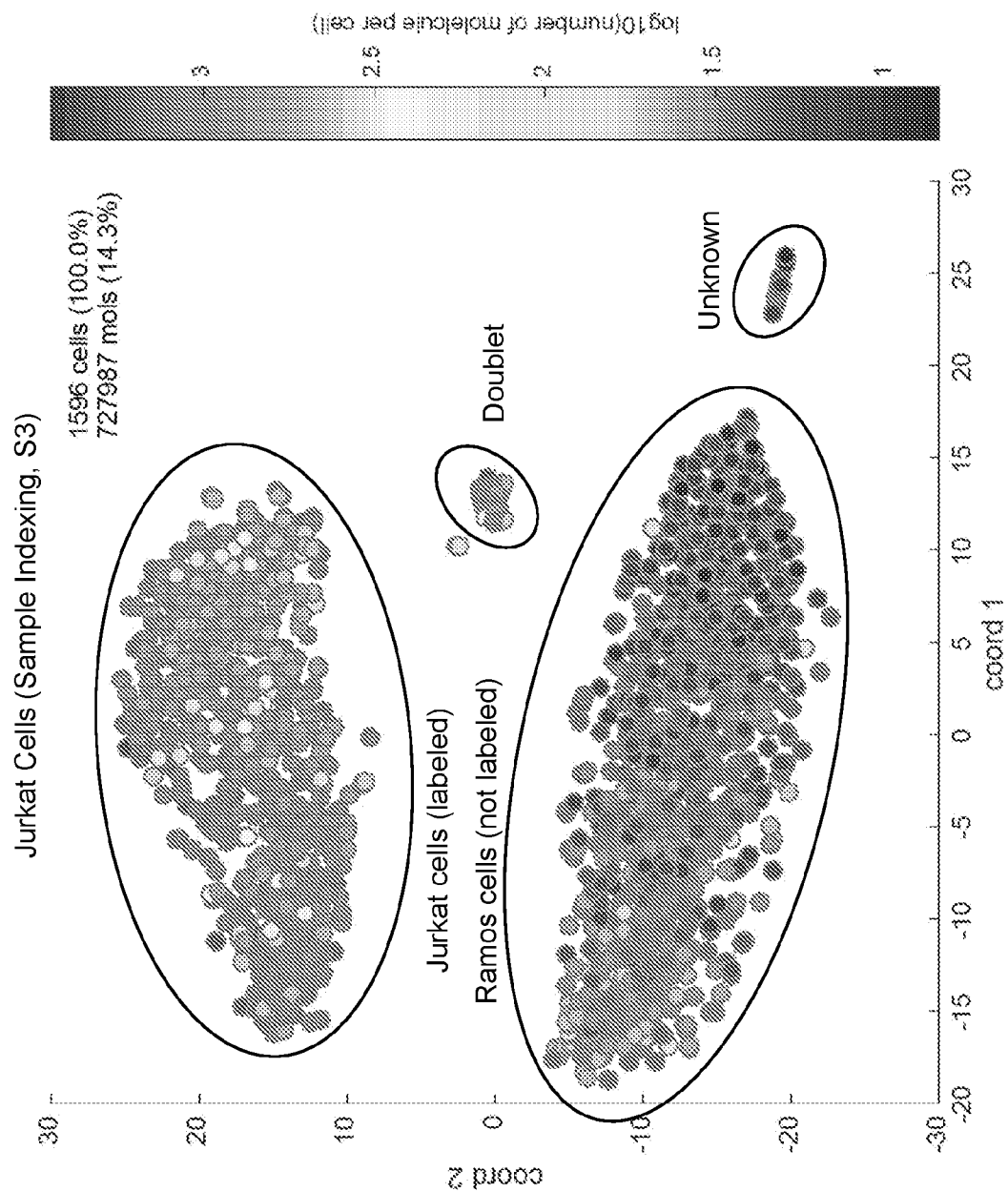
52/112



**FIG. 28B**

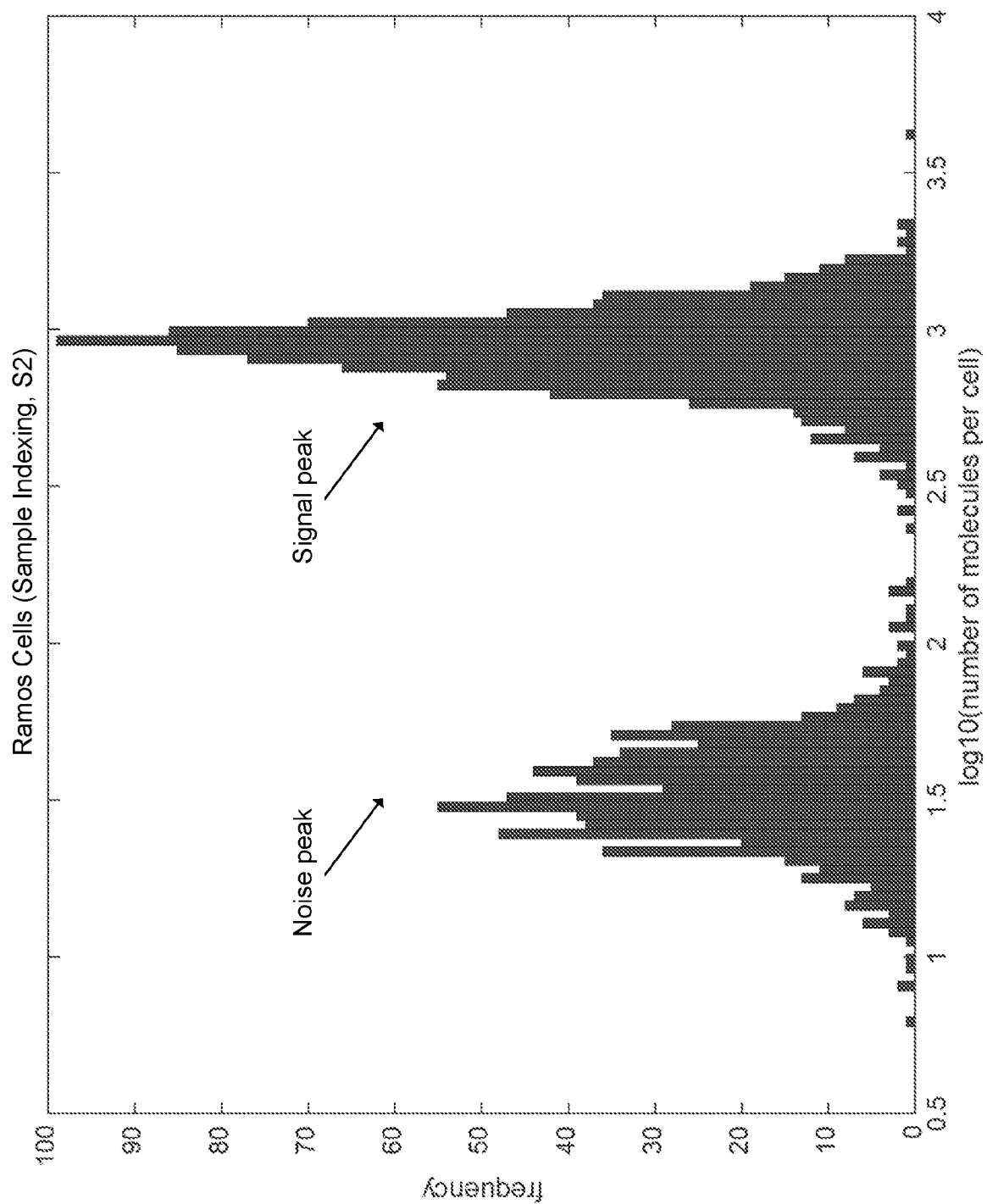


53/112

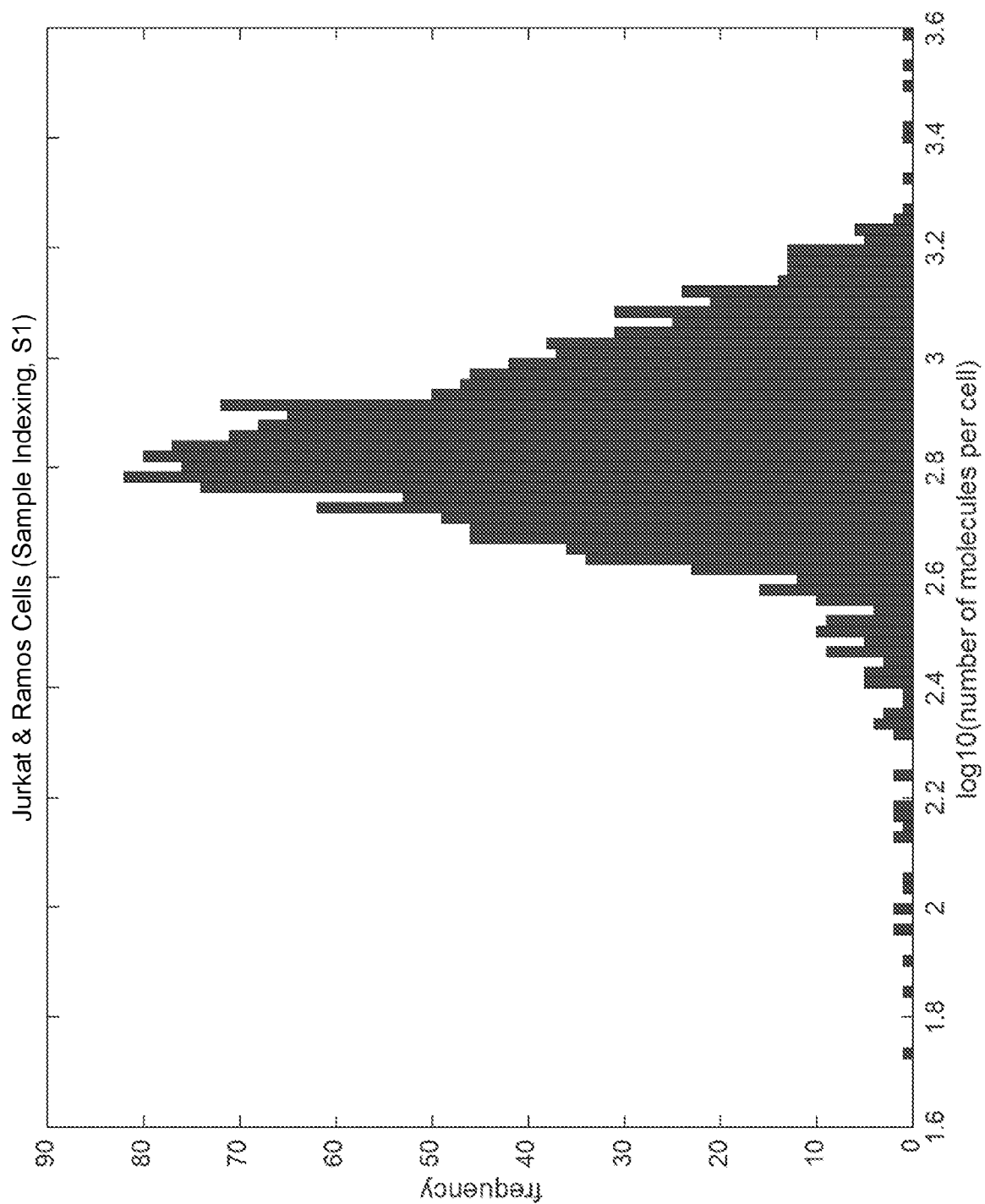


**FIG. 28C**

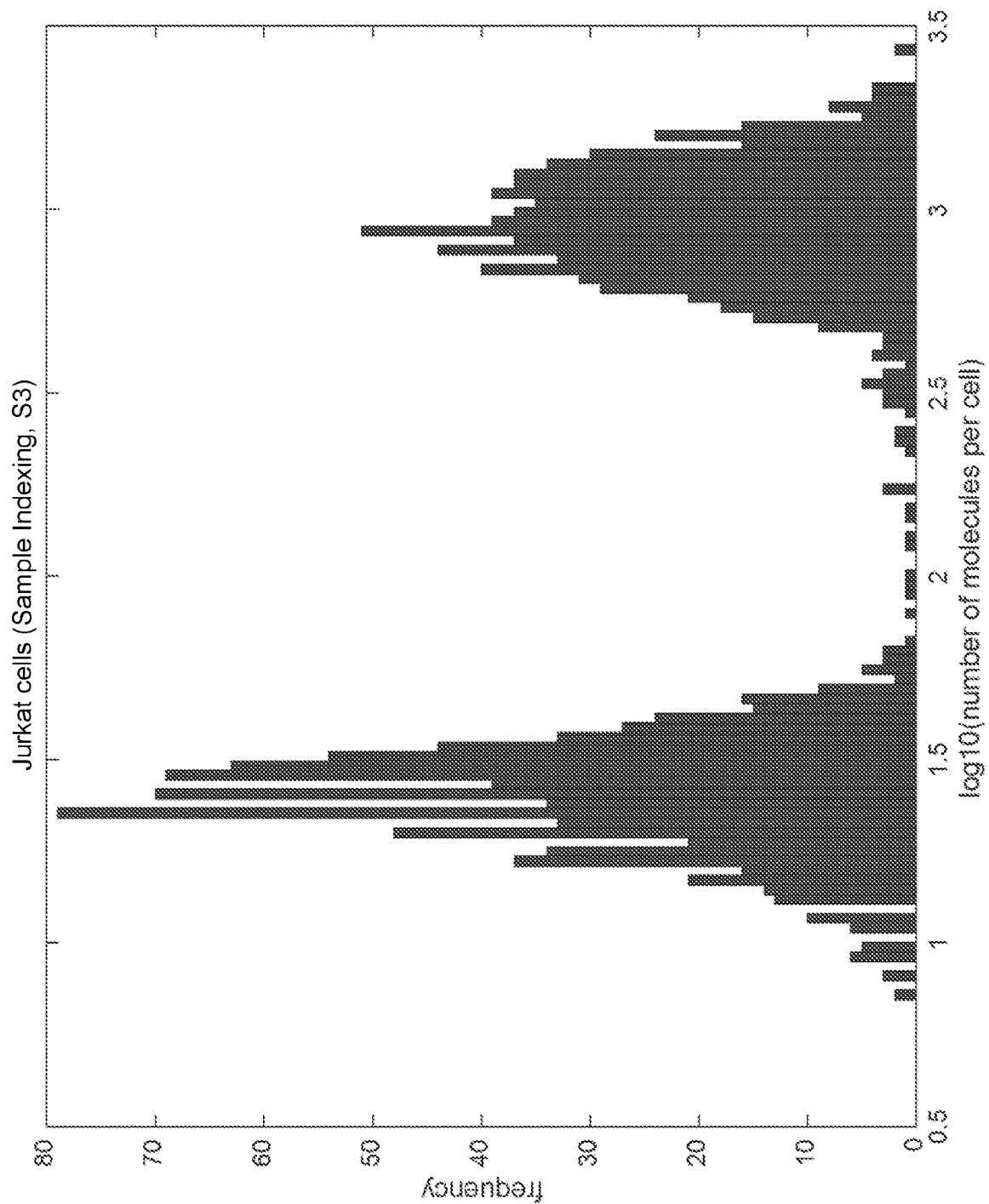
54/112

**FIG. 29A**

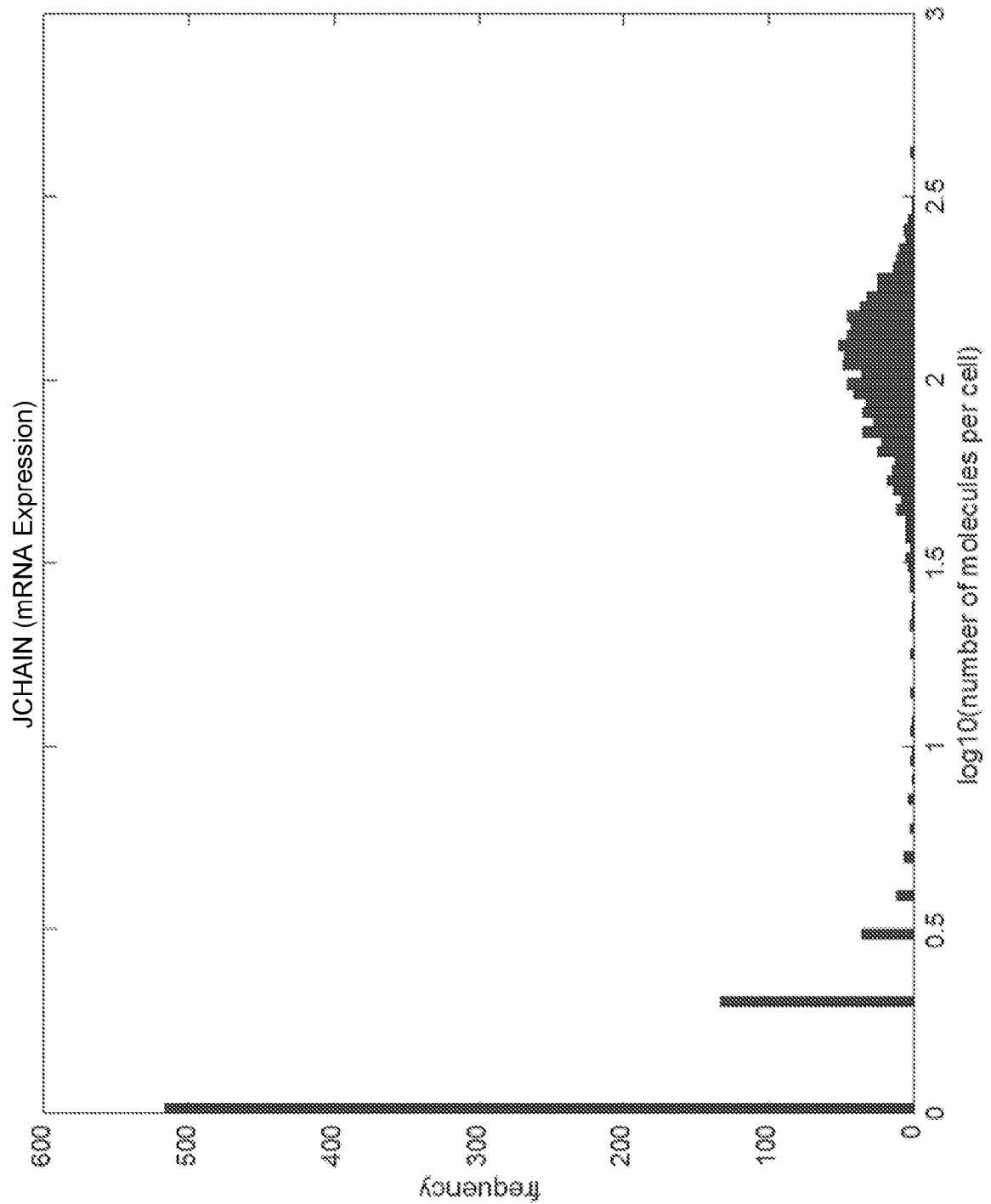
55/112

**FIG. 29B**

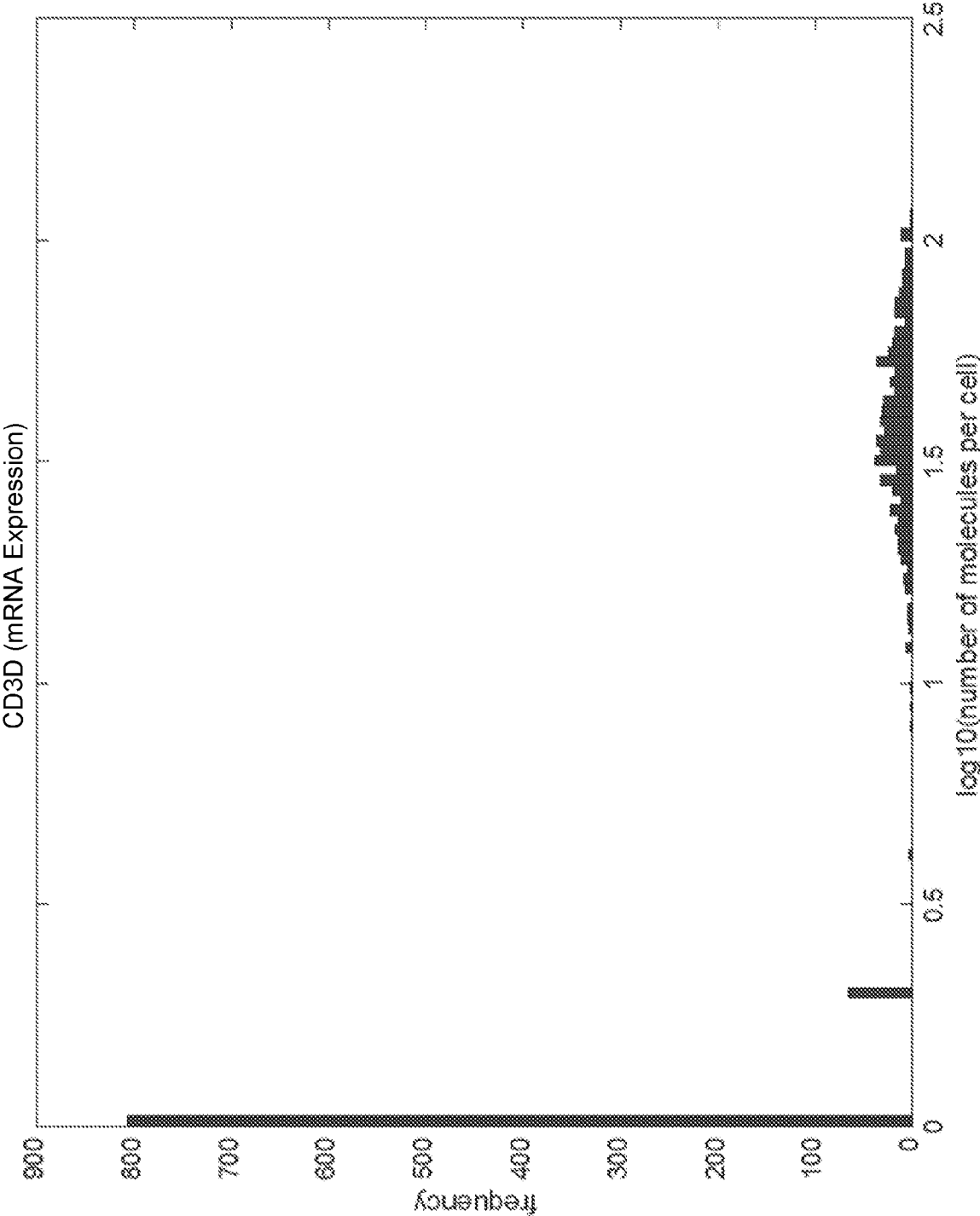
56/112

**FIG. 29C**

57/112

**FIG. 30A**

58/112



**FIG. 30B**

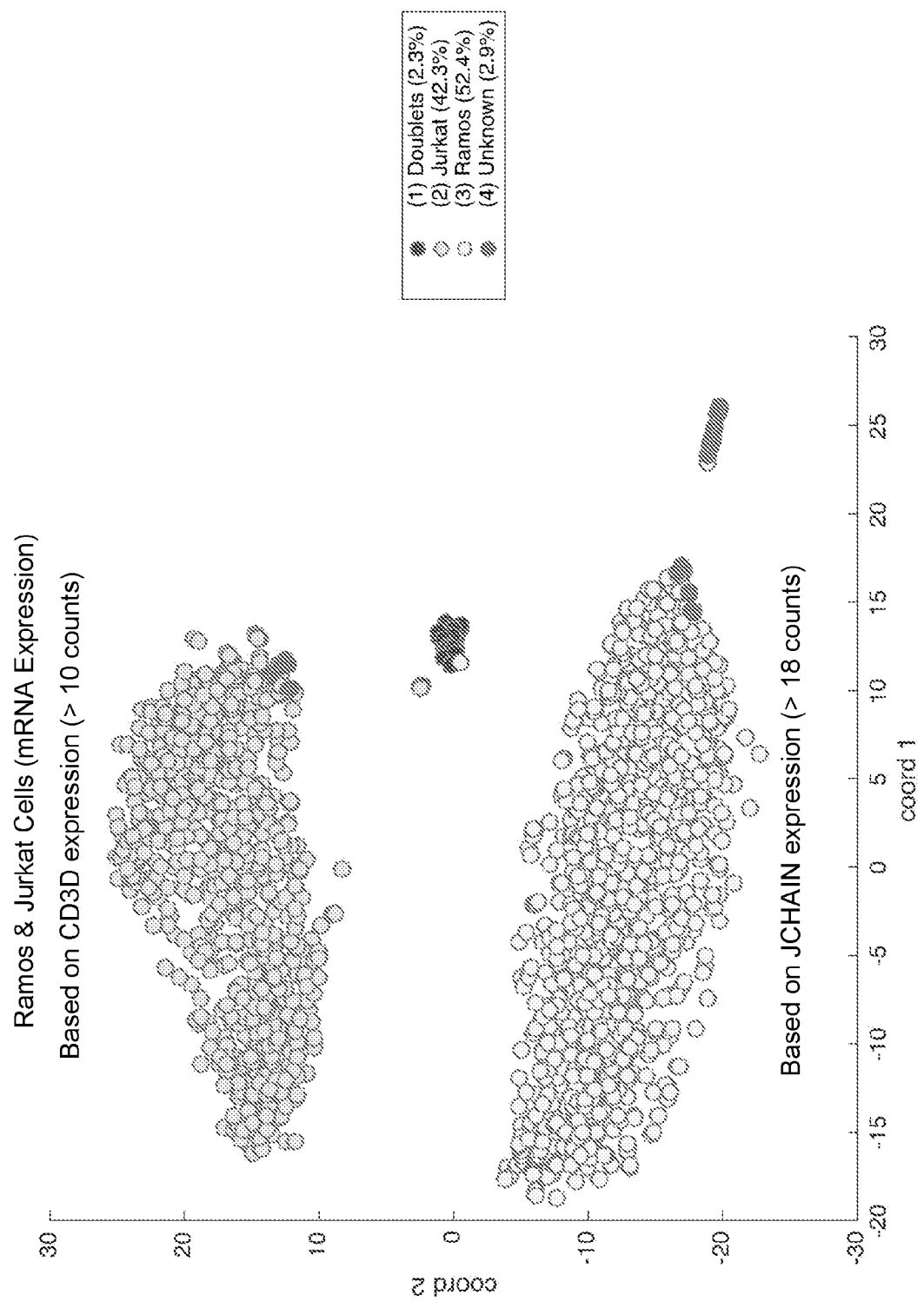


FIG. 30C

60/112

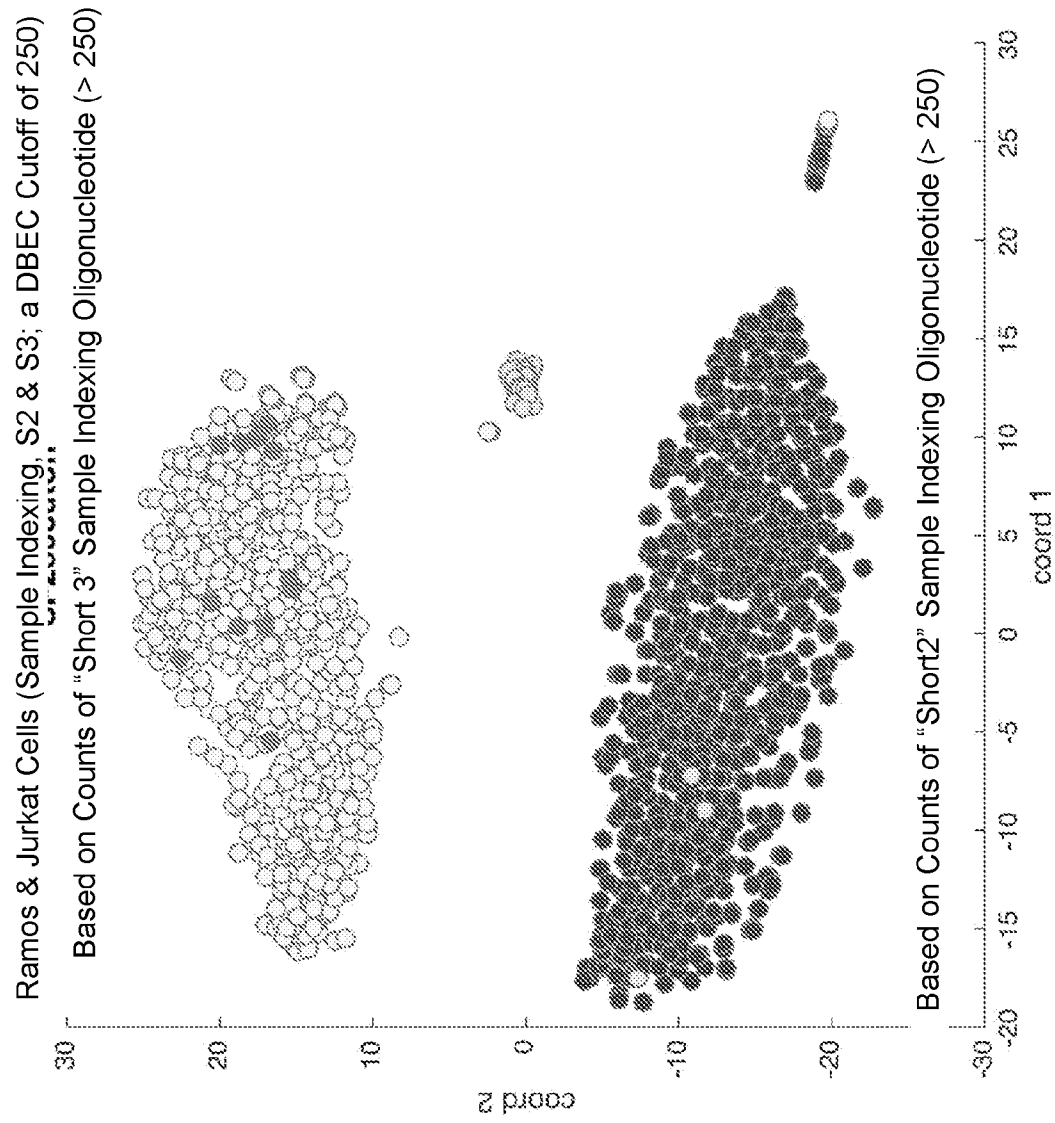
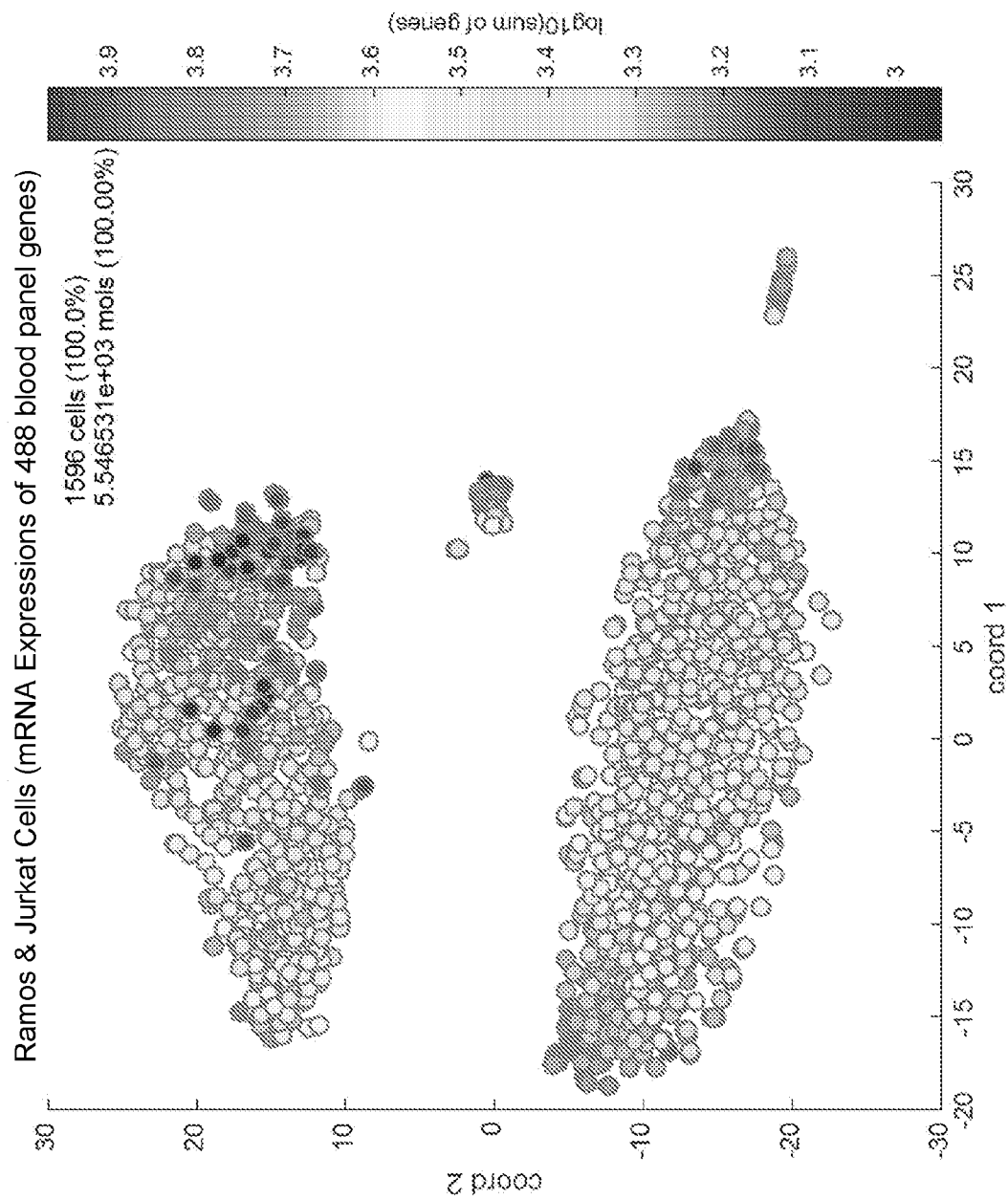


FIG. 30D

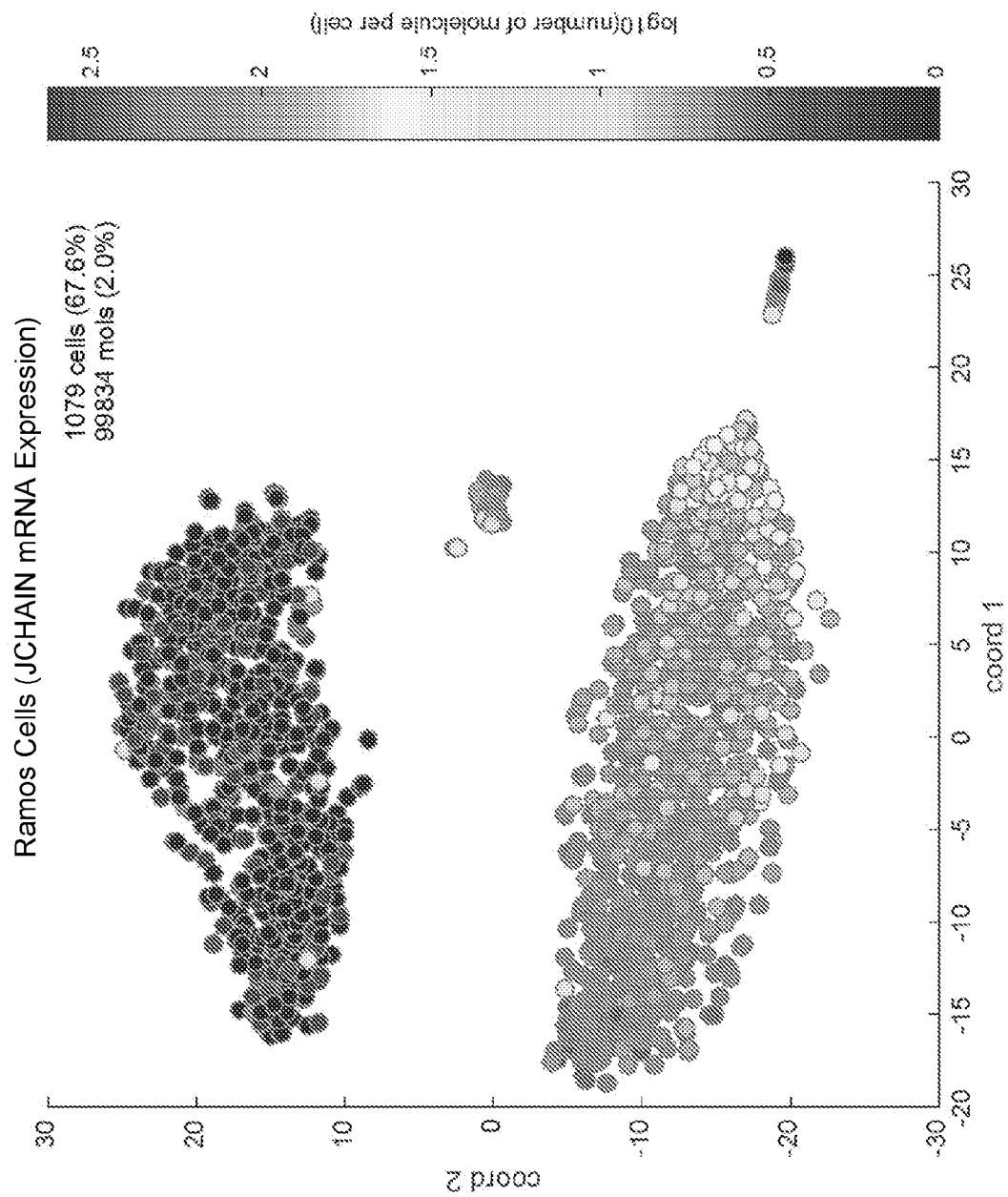


61/112



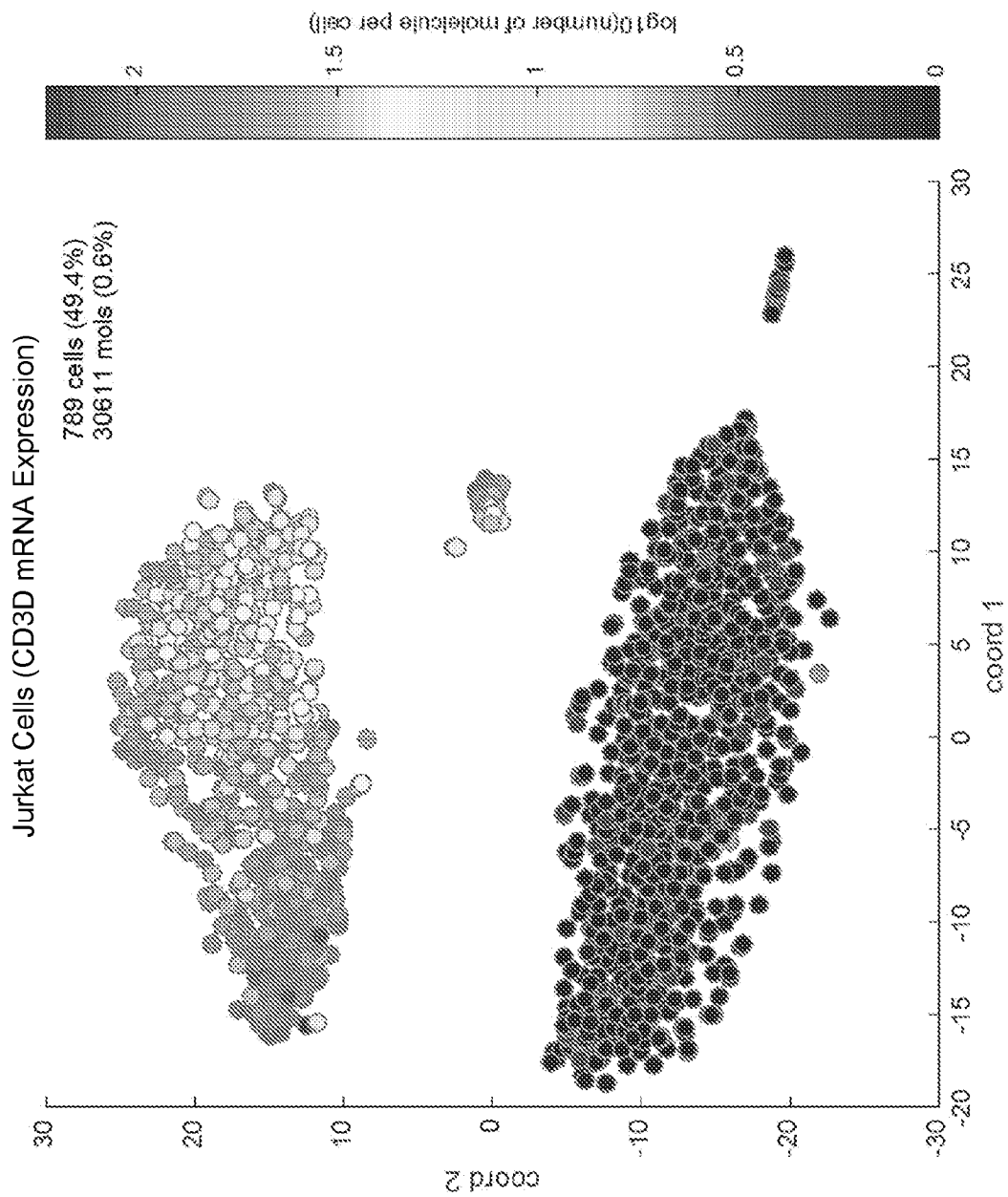
**FIG. 31A**

62/112

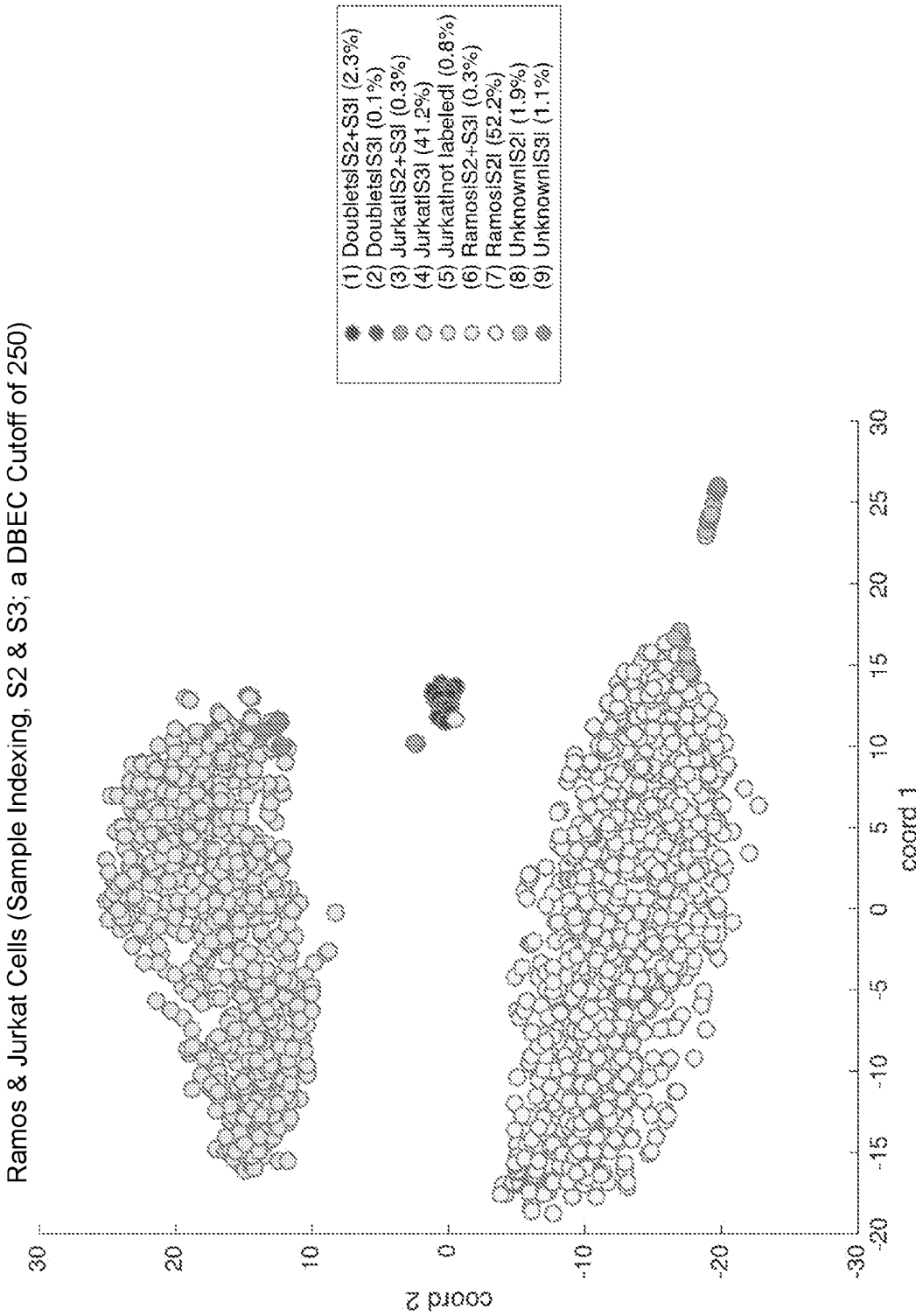


**FIG. 31B**

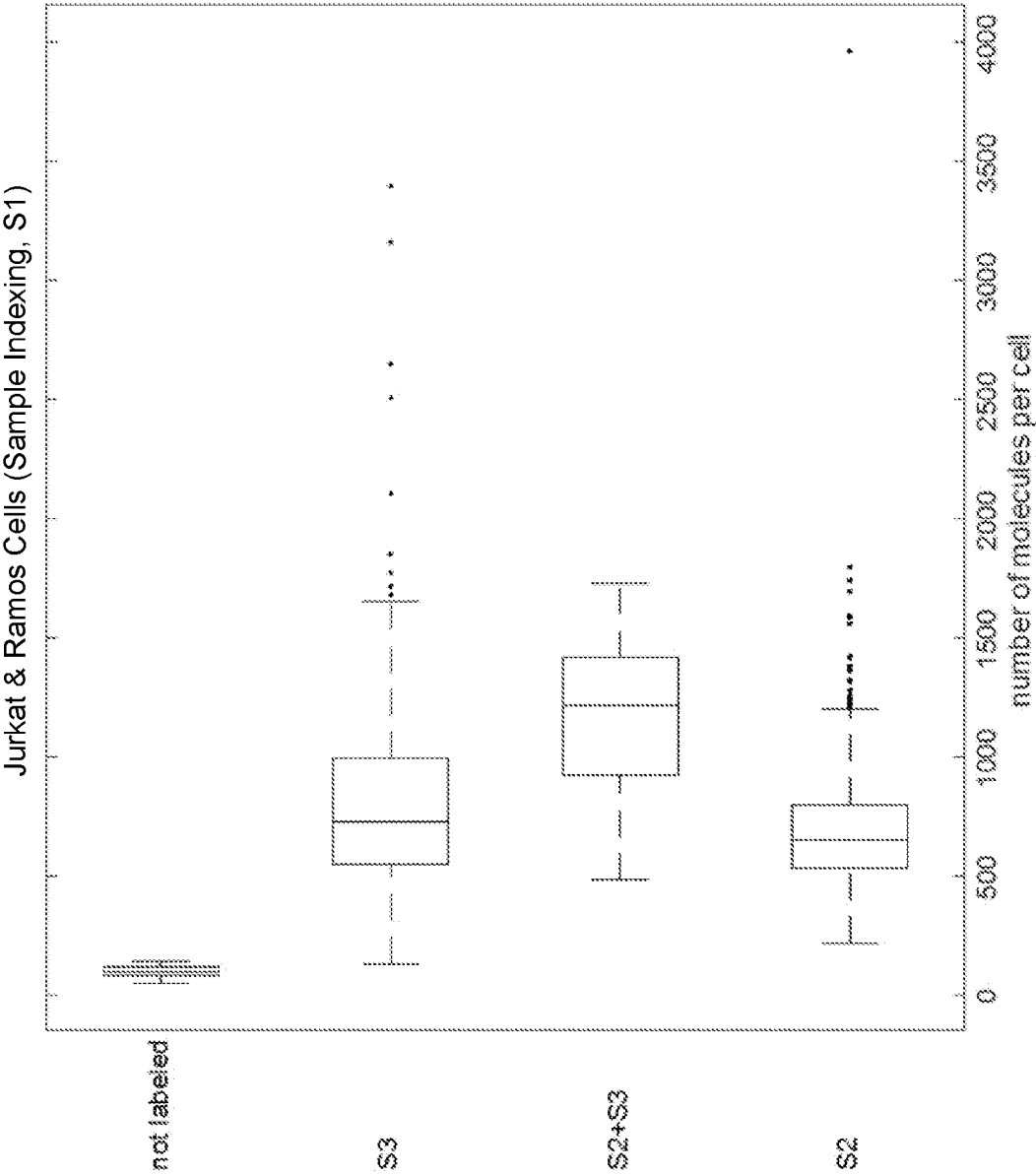
63/112



**FIG. 31C**

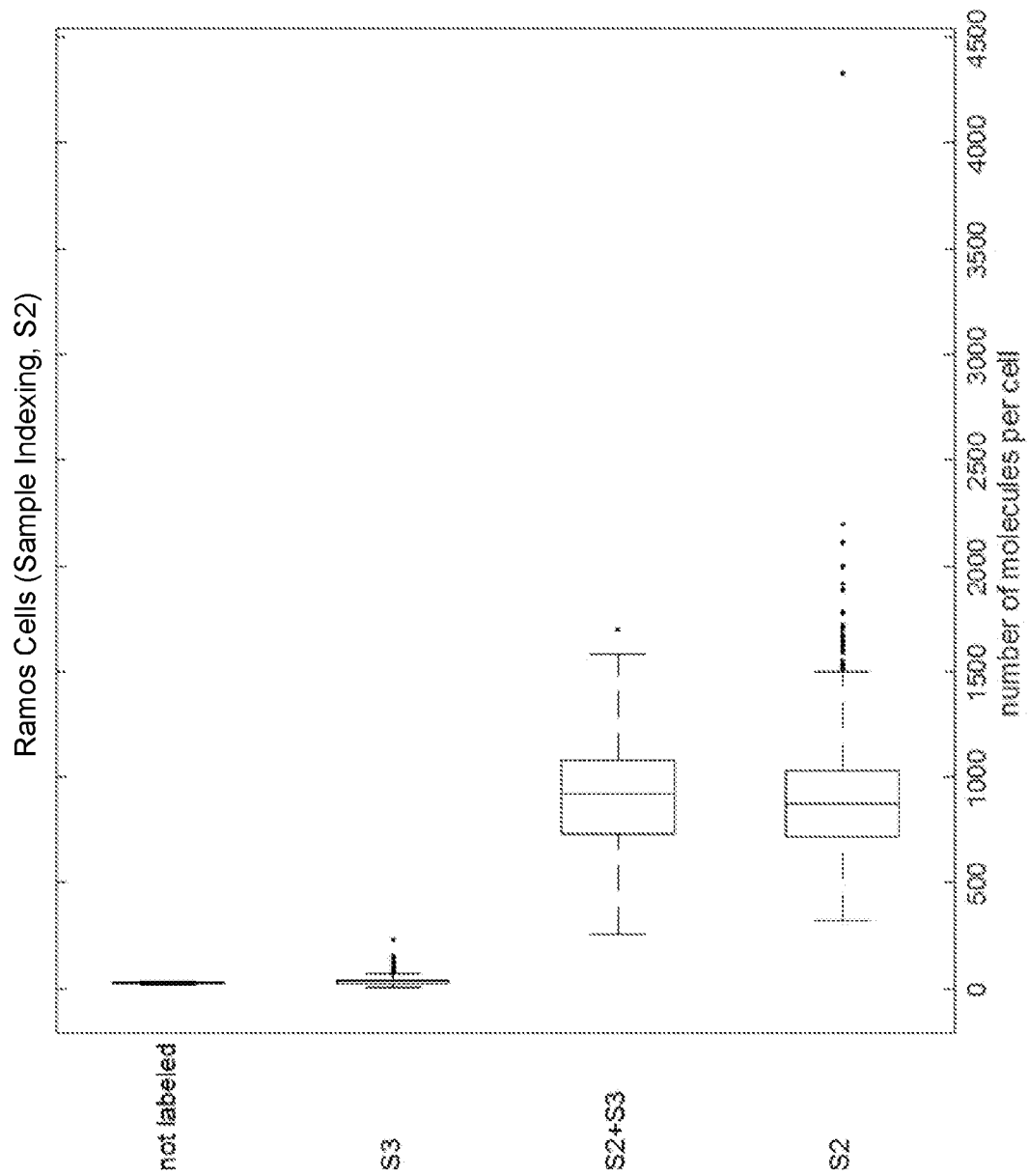


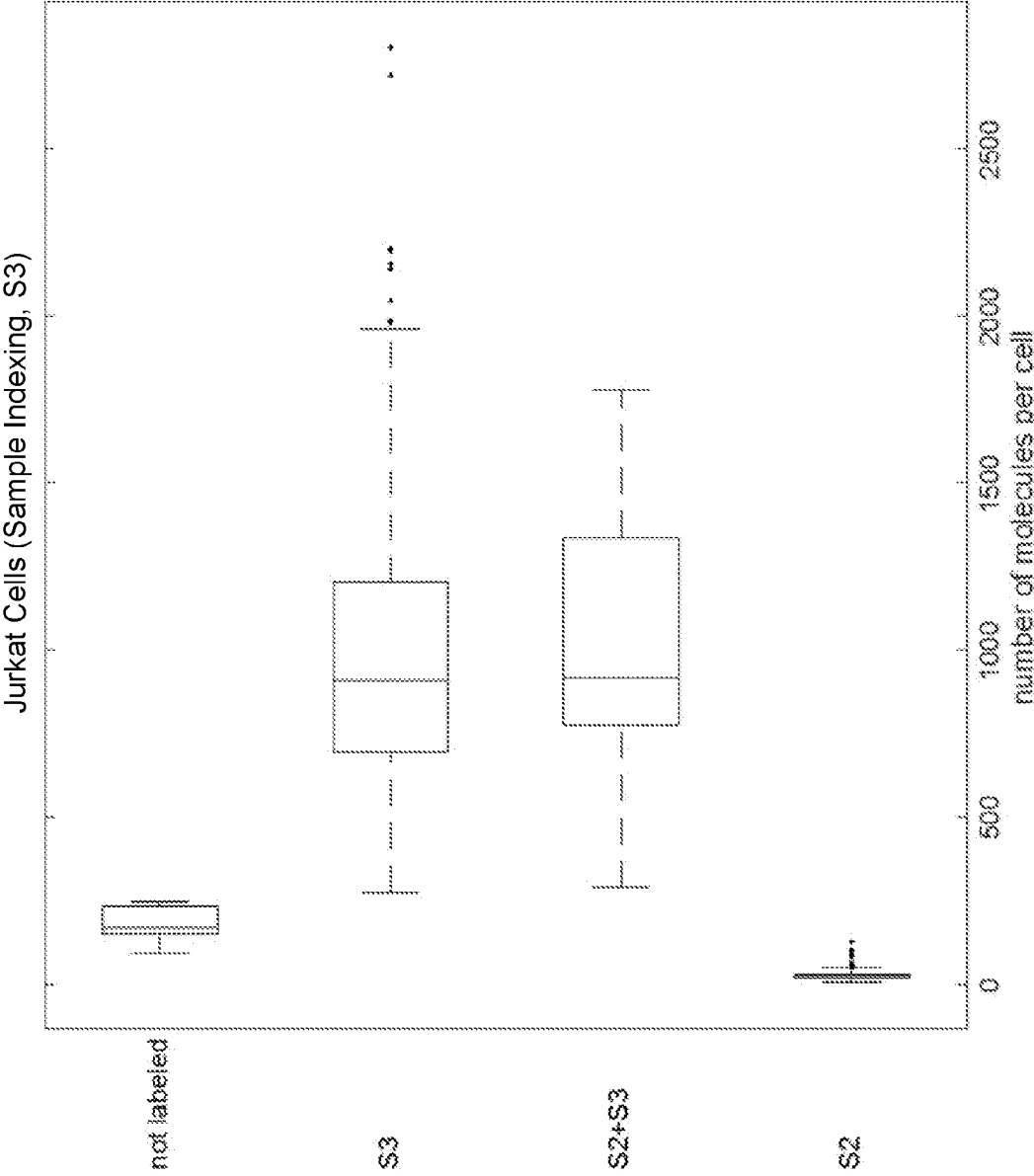
**FIG. 32**



**FIG. 33A**

66/112

**FIG. 33B**



**FIG. 33C**

68/112

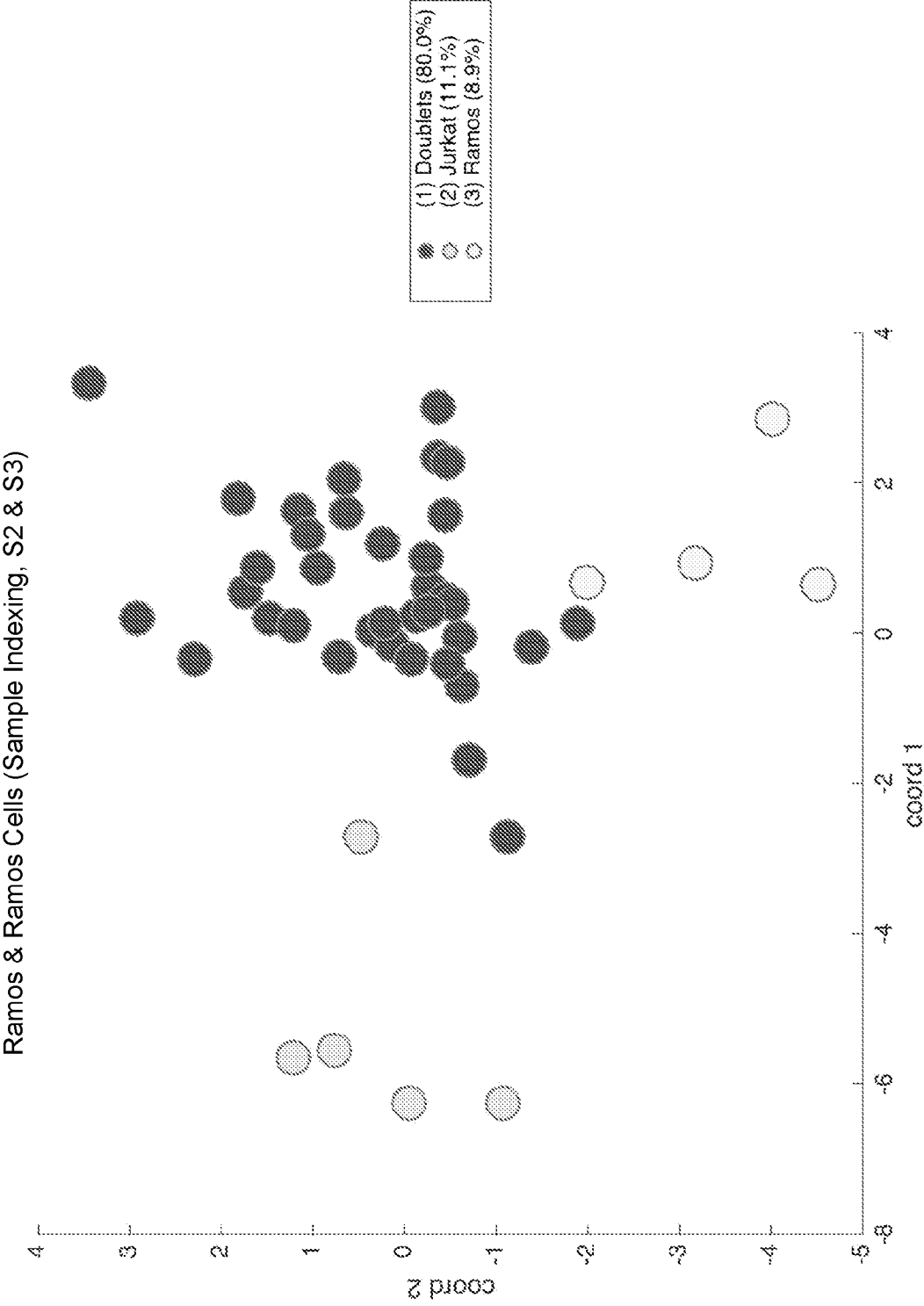
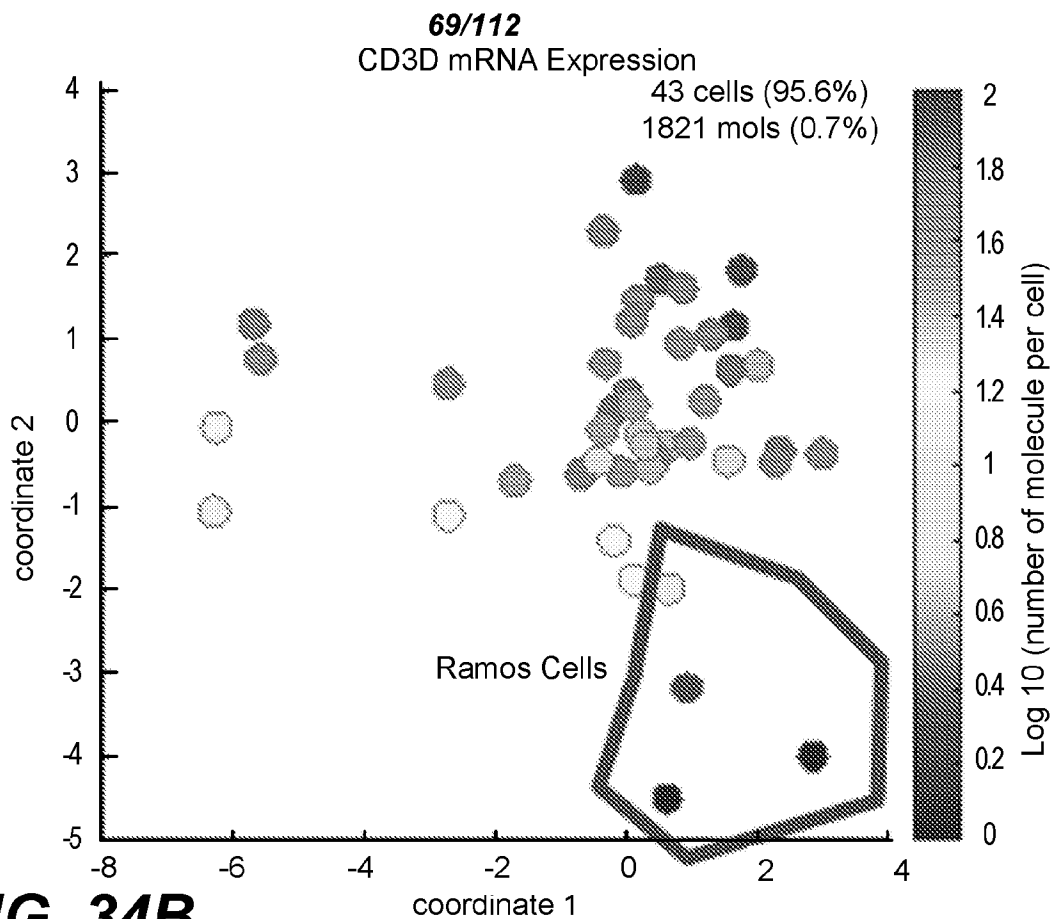
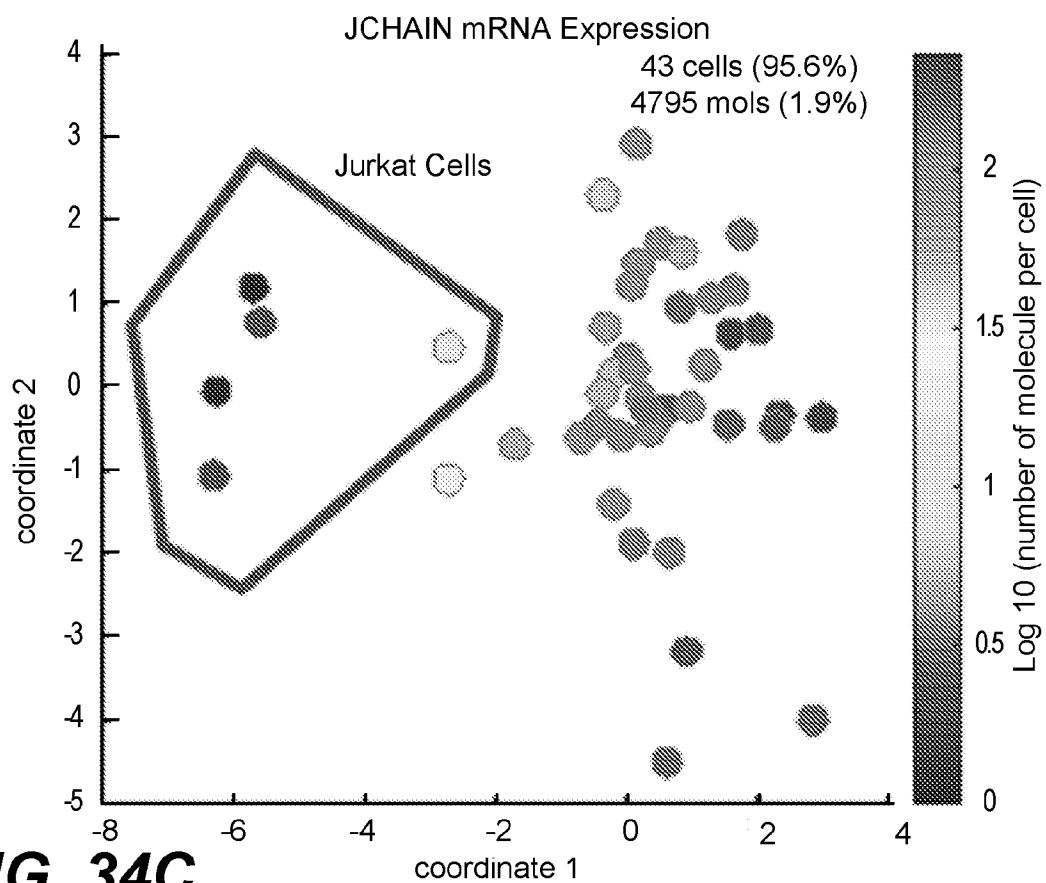
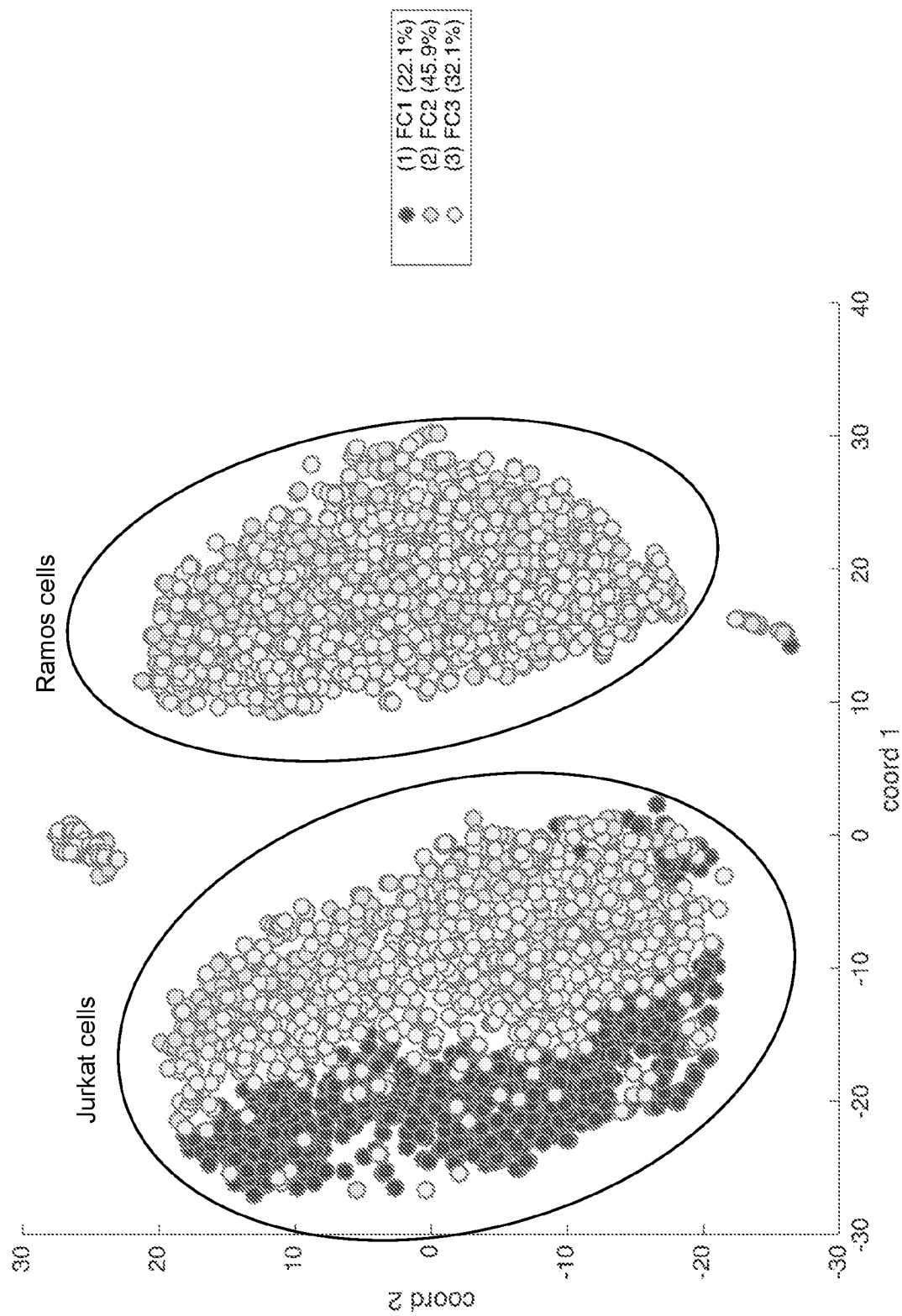


FIG. 34A



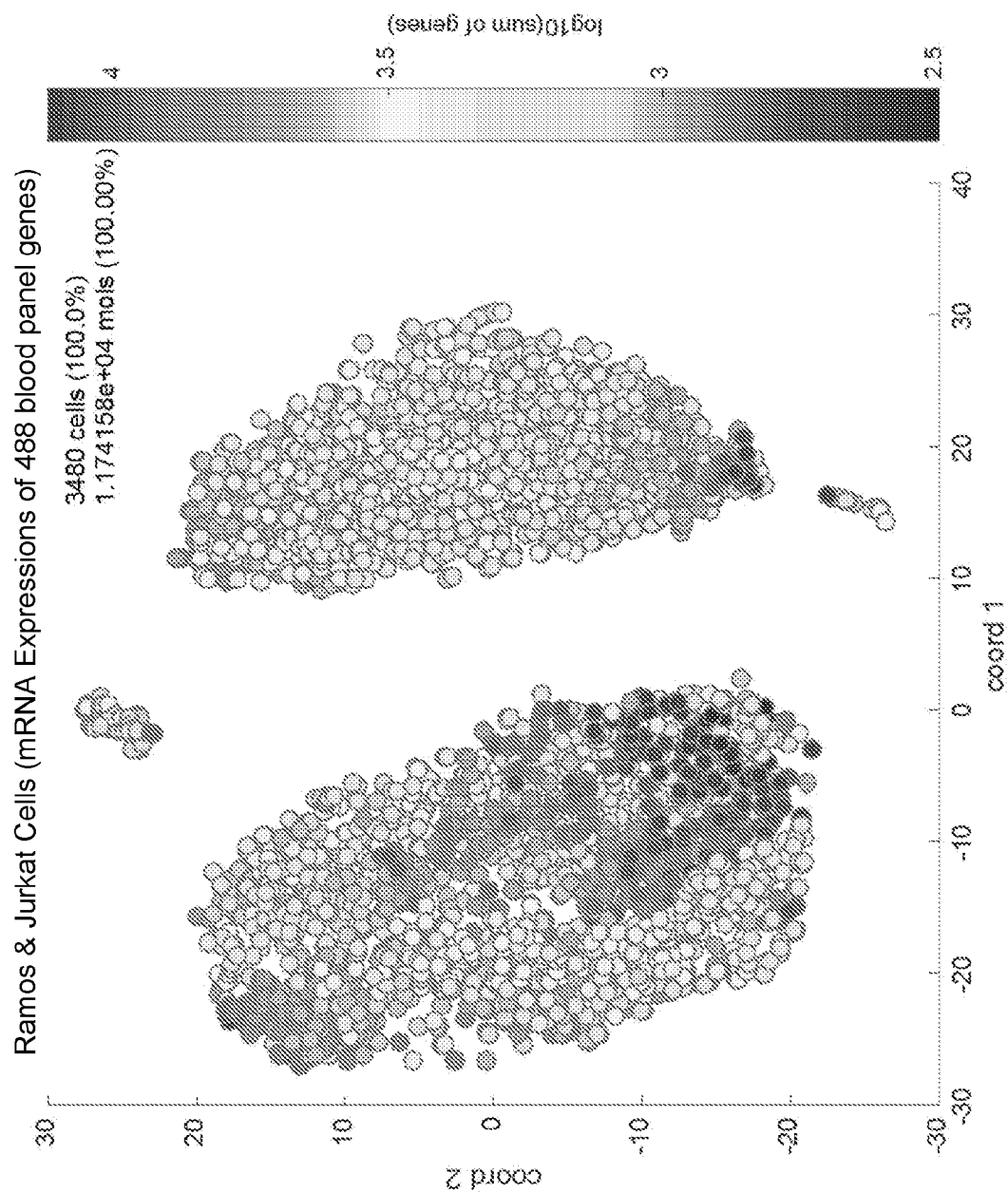
**FIG. 34B****FIG. 34C**

70/112



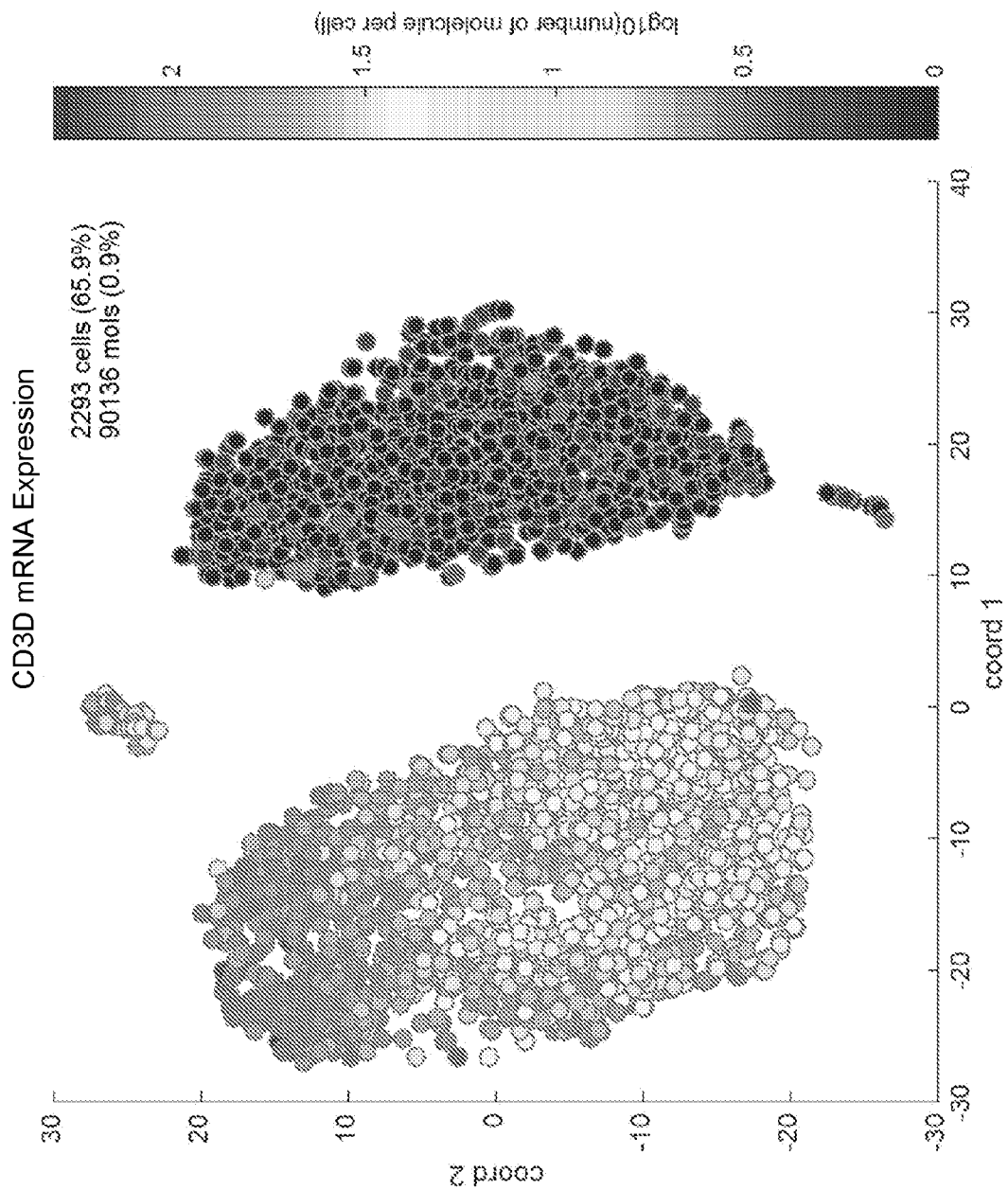
**FIG. 35A**

71/112

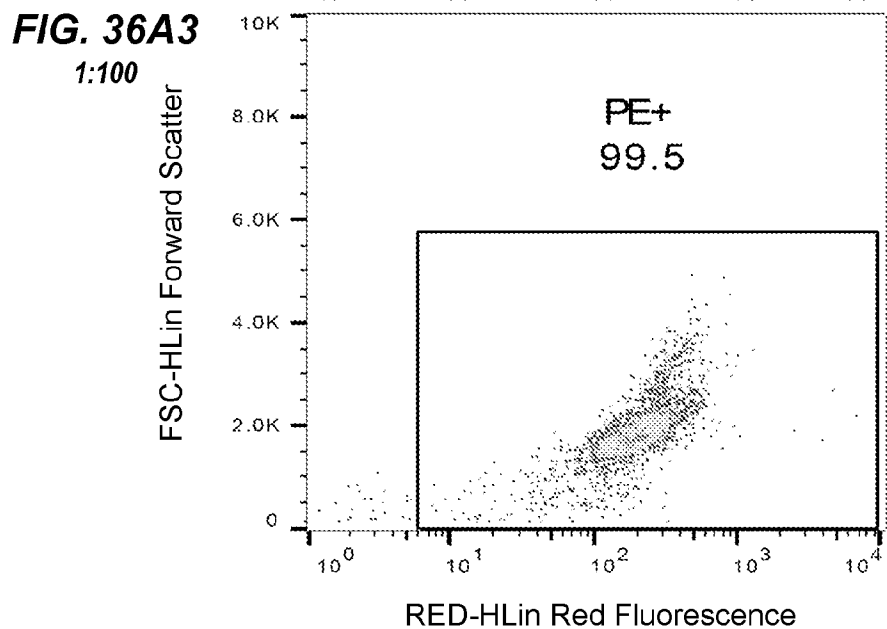
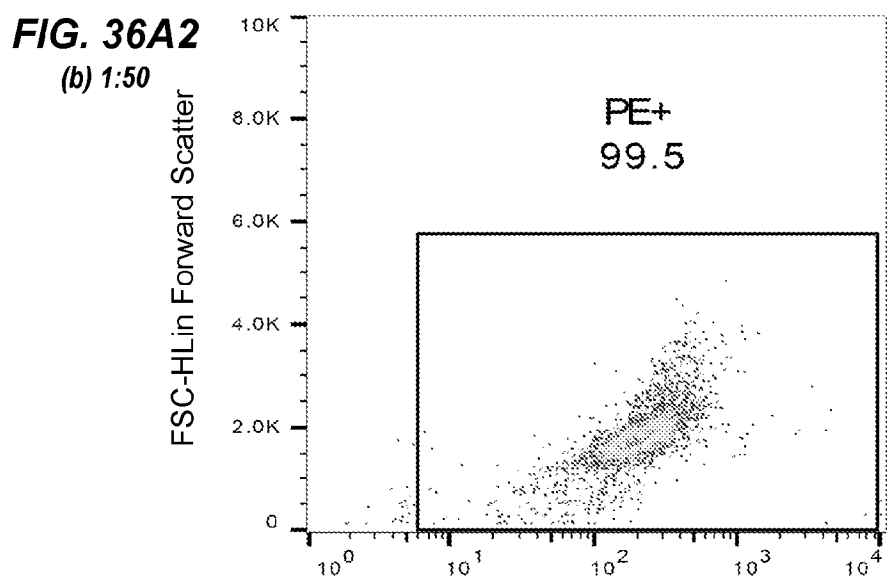
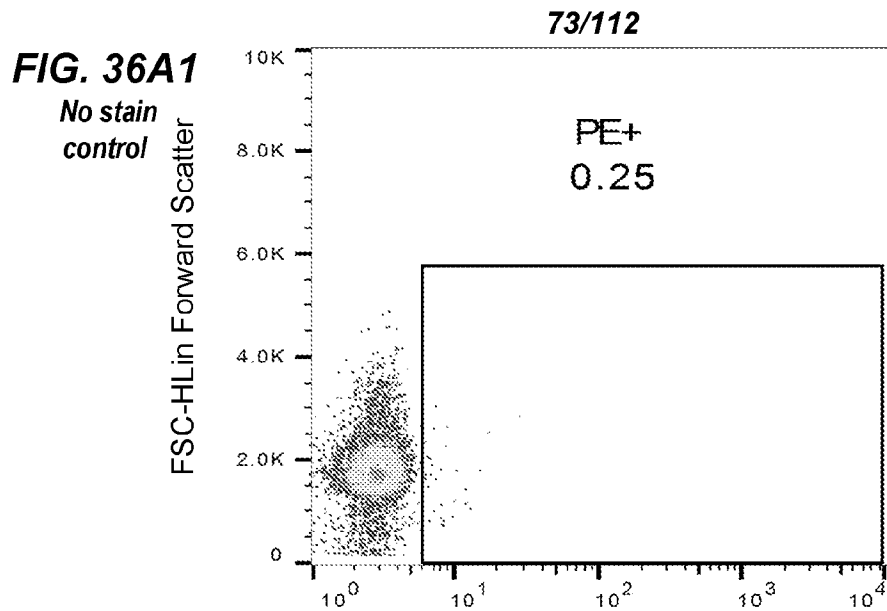


**FIG. 35B**

72/112



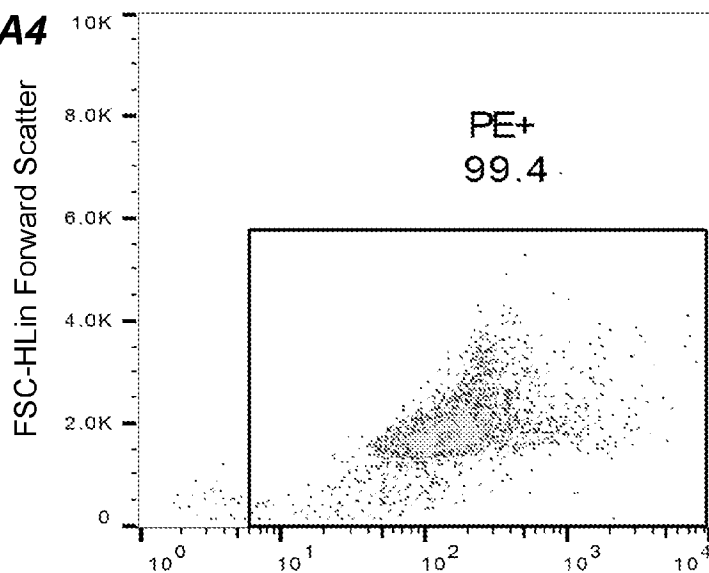
**FIG. 35C**



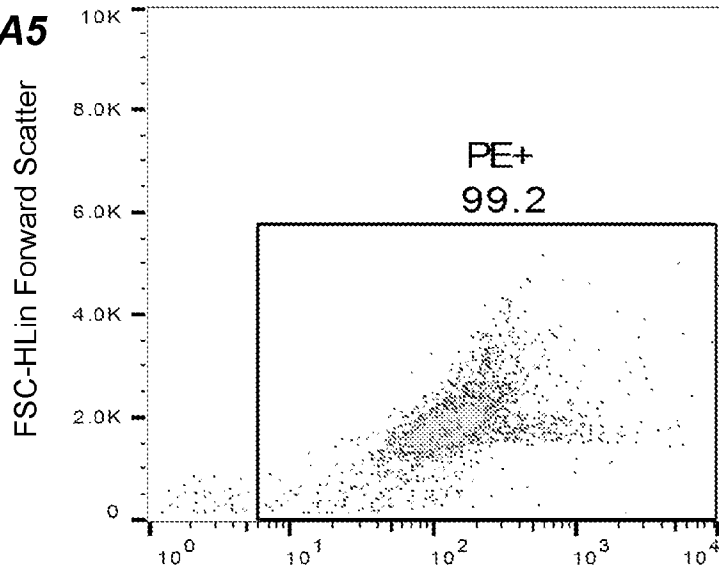
74/112

**FIG. 36A4**

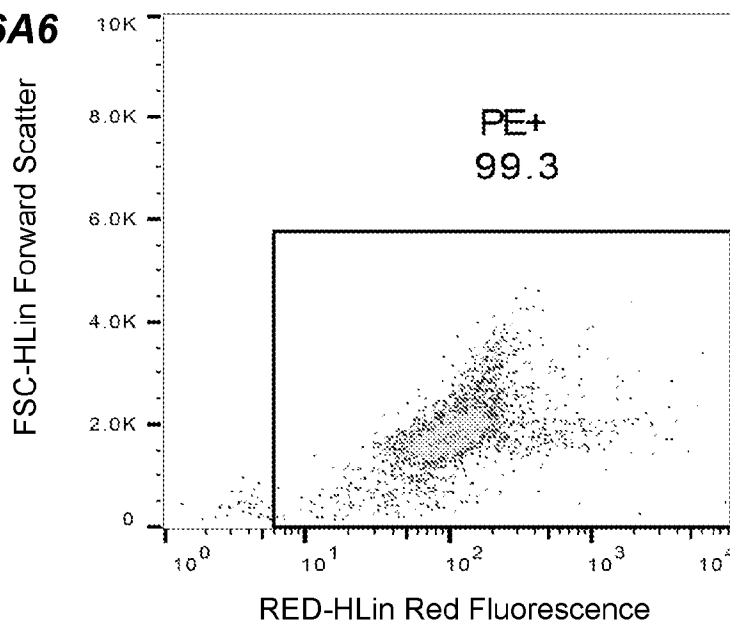
1:200

**FIG. 36A5**

1:300

**FIG. 36A6**

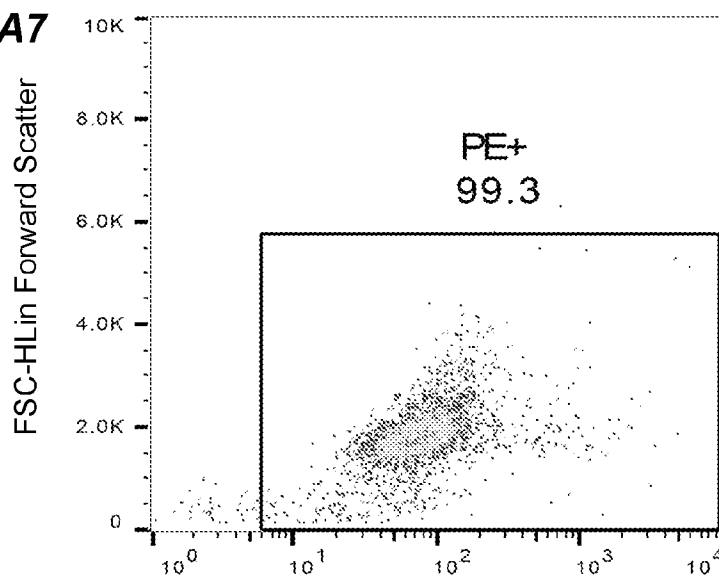
1:400



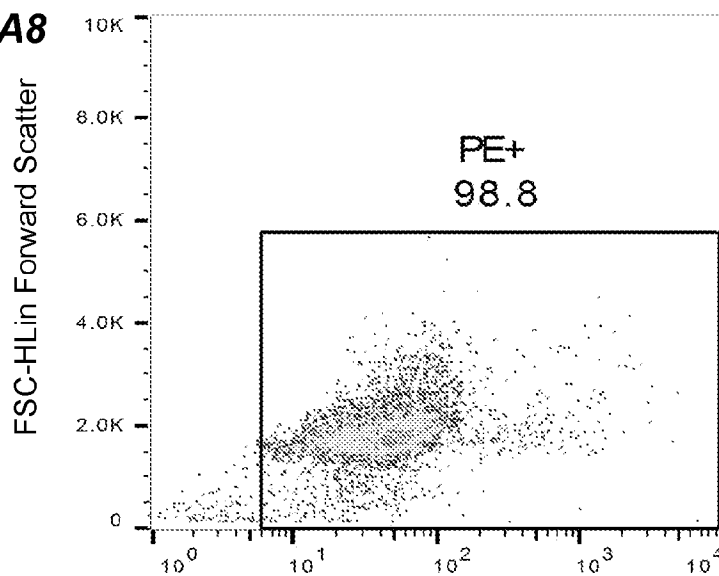
75/112

**FIG. 36A7**

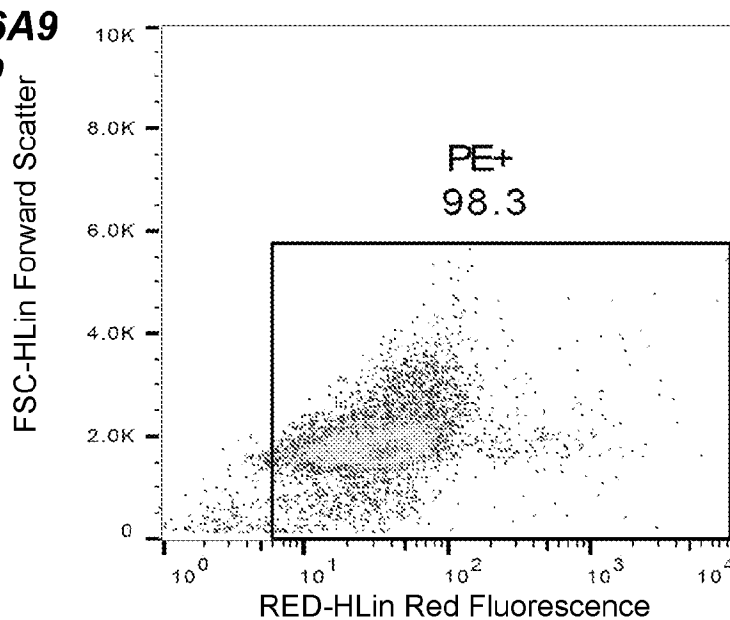
1:600

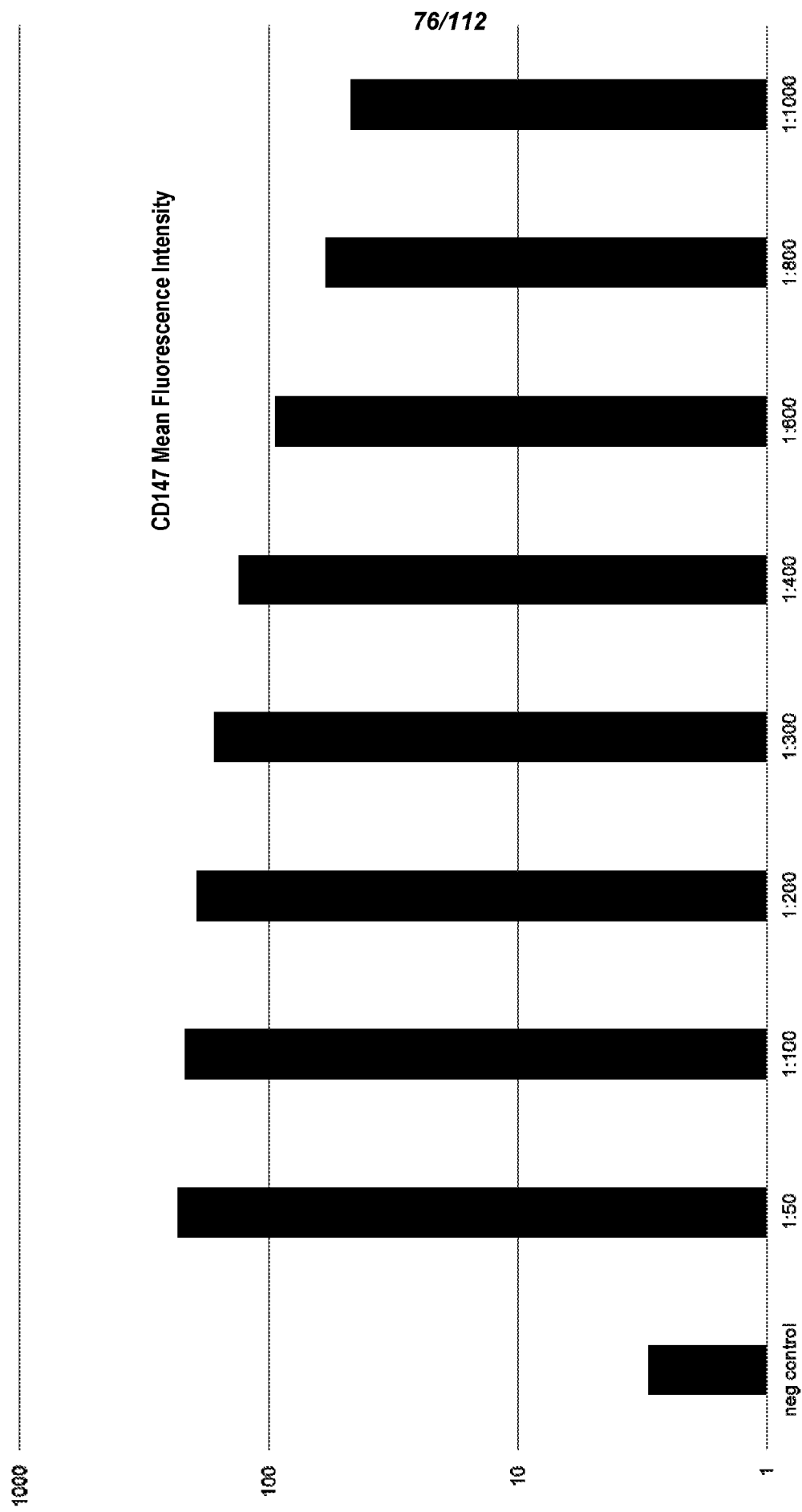
**FIG. 36A8**

1:800

**FIG. 36A9**

1:1000

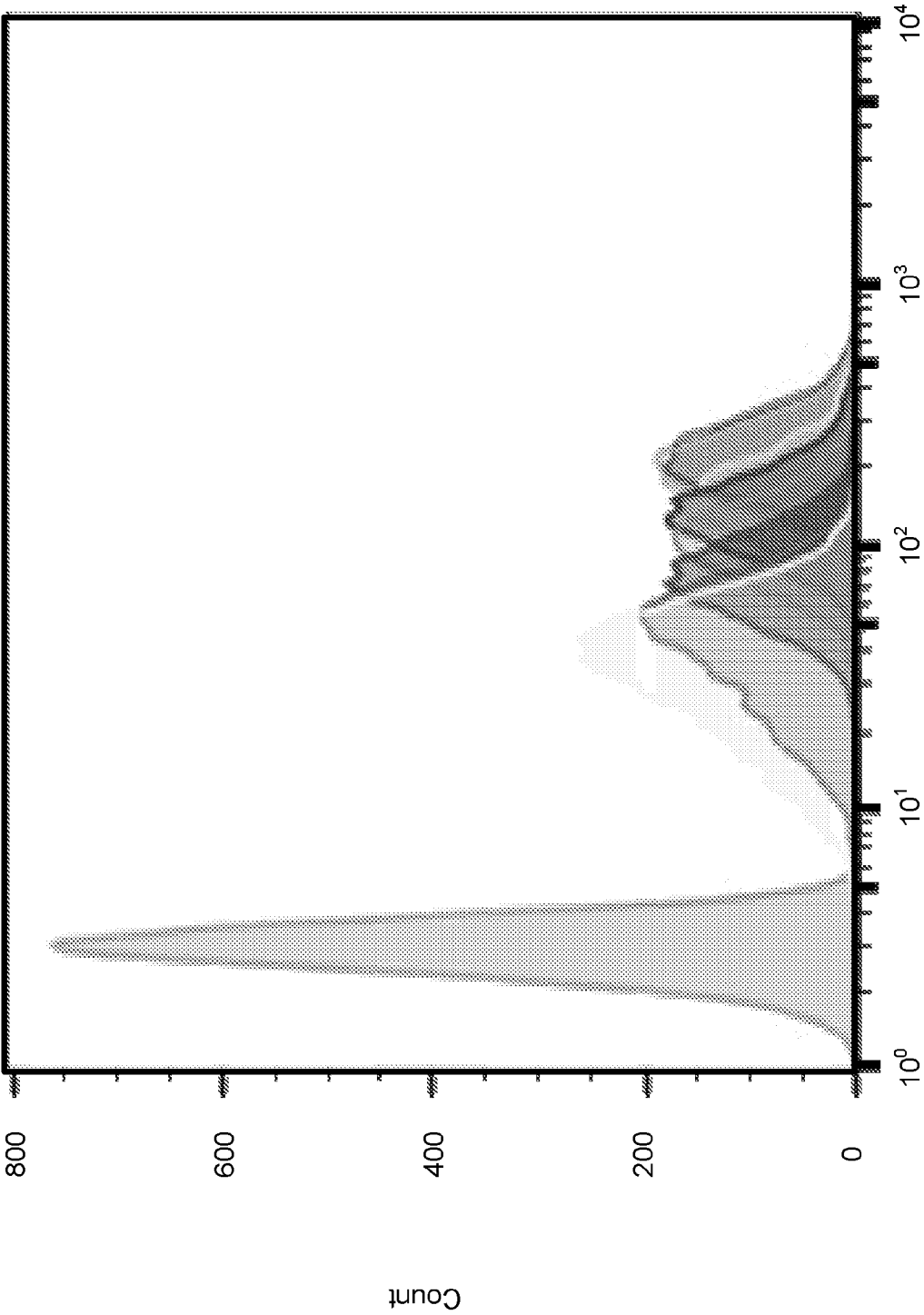




**FIG. 36B**



77/112

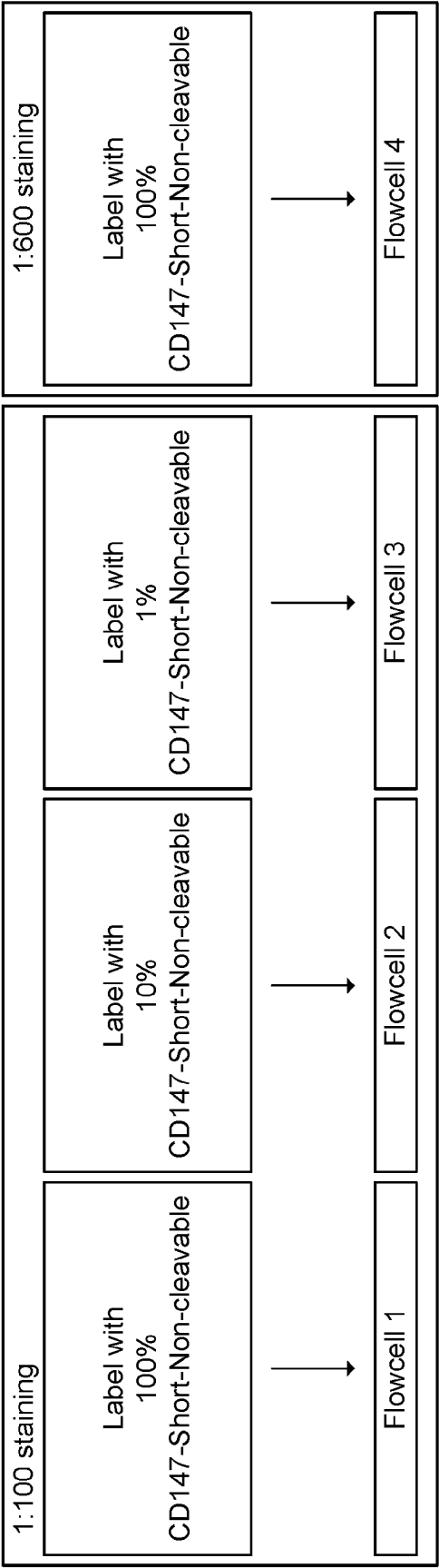


RED-HL in Red Fluorescence

**FIG. 36C**

78/112

PBMCs

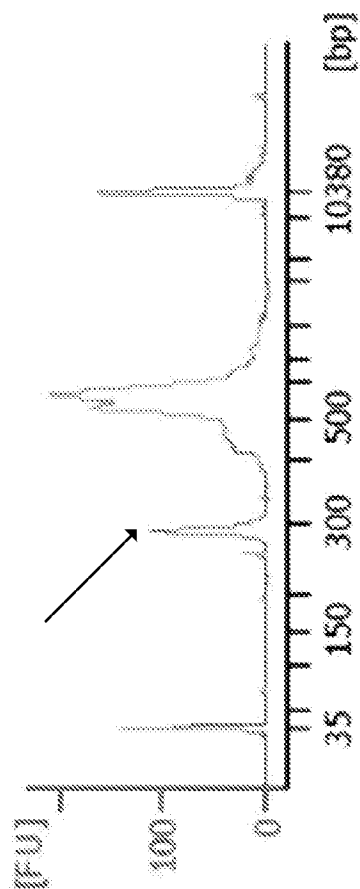


Use conjugated Abs in pH 7.5 diluent

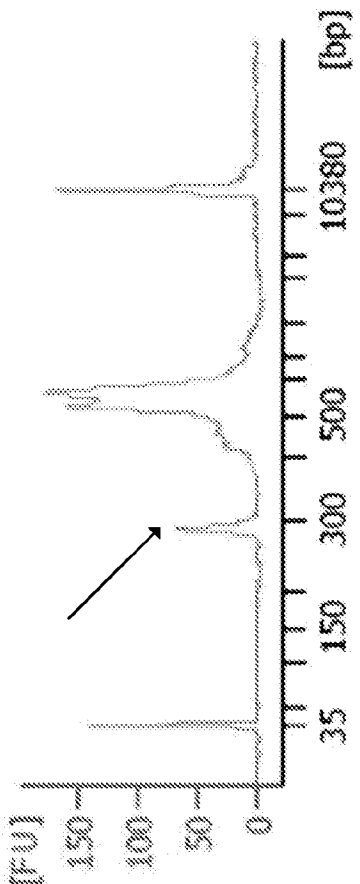
FIG. 37

79/112

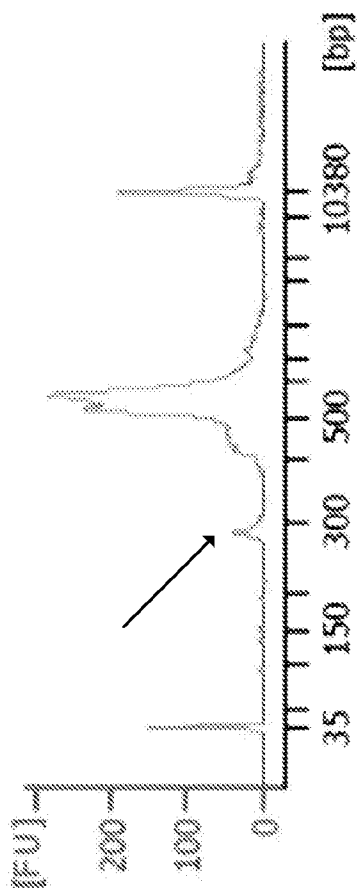
**FIG. 38A** FC1 - 1:100 Dilution, 100% Hot



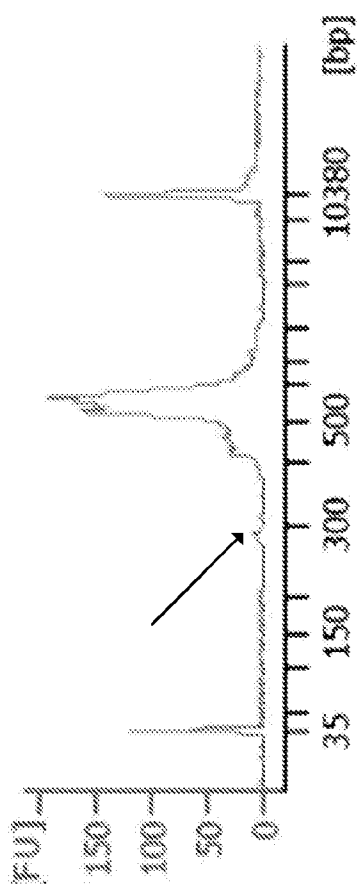
**FIG. 38B** FC1 - 1:800 Dilution, 100% Hot



**FIG. 38C** FC1 - 1:100 Dilution, 10% Hot



**FIG. 38D** FC1 - 1:100 Dilution, 1% Hot

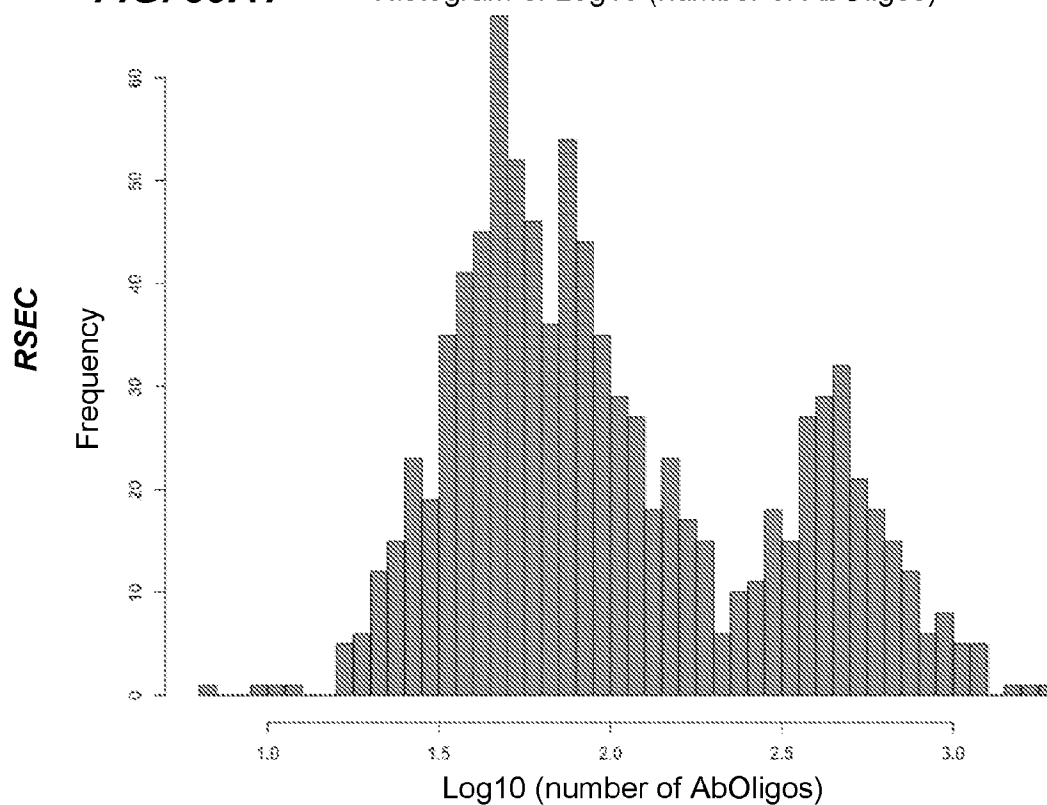


**80/112**

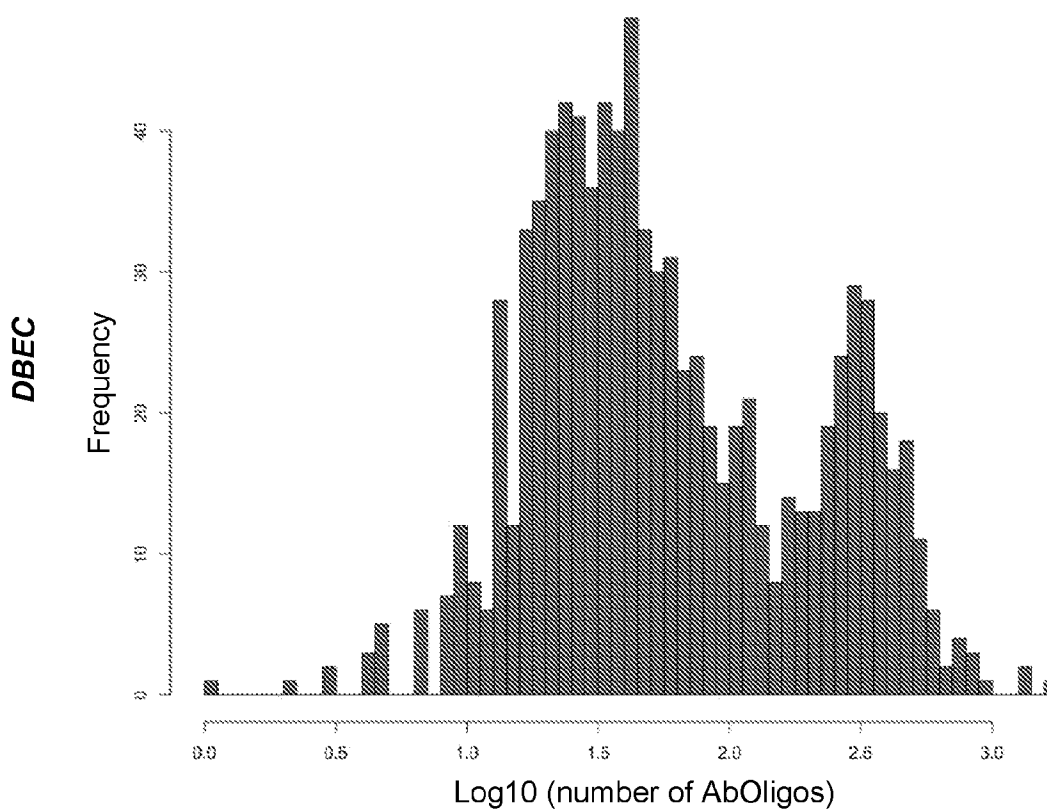
FC2 (1:100 10% hot Ab)

**FIG. 39A1**

Histogram of Log10 (number of AbOligos)

**FIG. 39B1**

Histogram of Log10 (number of AbOligos)

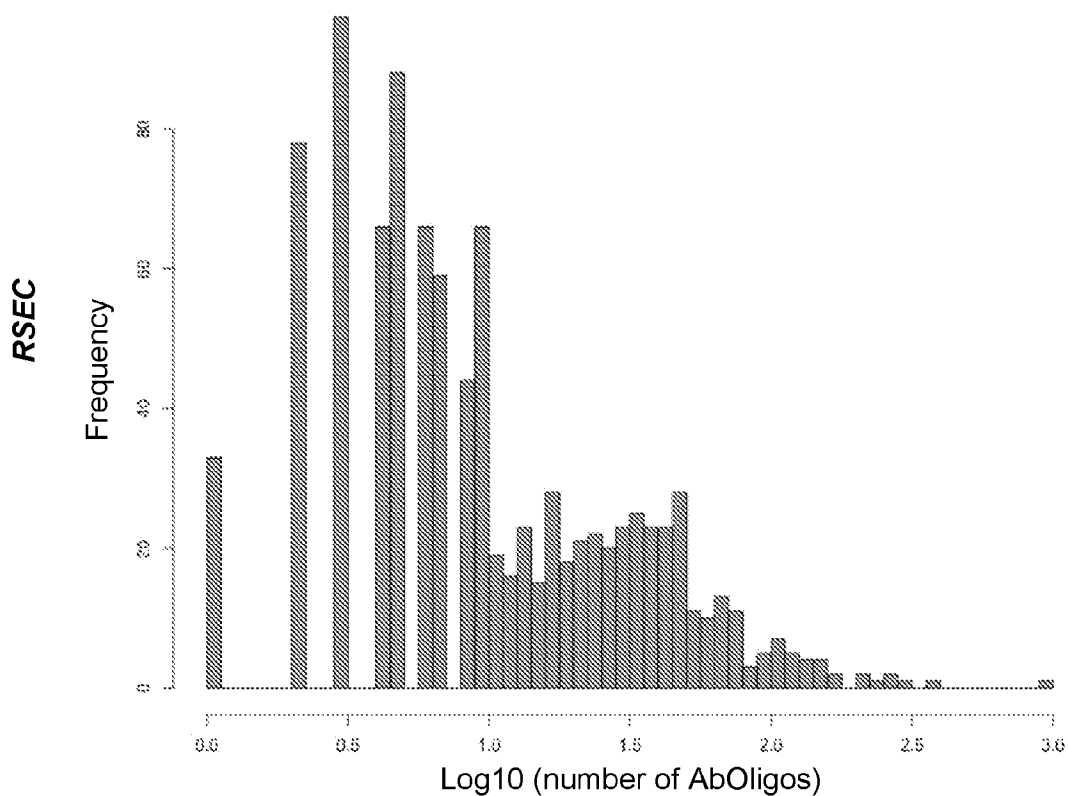


81/112

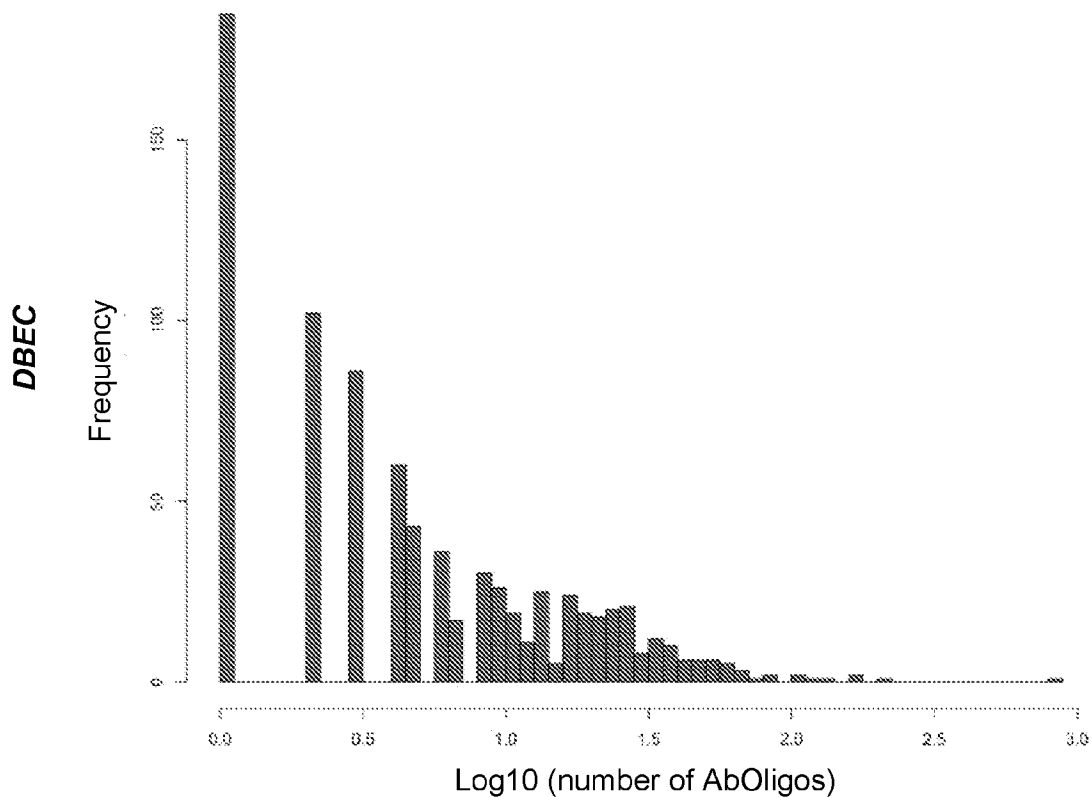
FC3 (1:100 1% hot Ab)

**FIG. 39A2**

Histogram of Log10 (number of AbOligos)

**FIG. 39B2**

Histogram of Log10 (number of AbOligos)

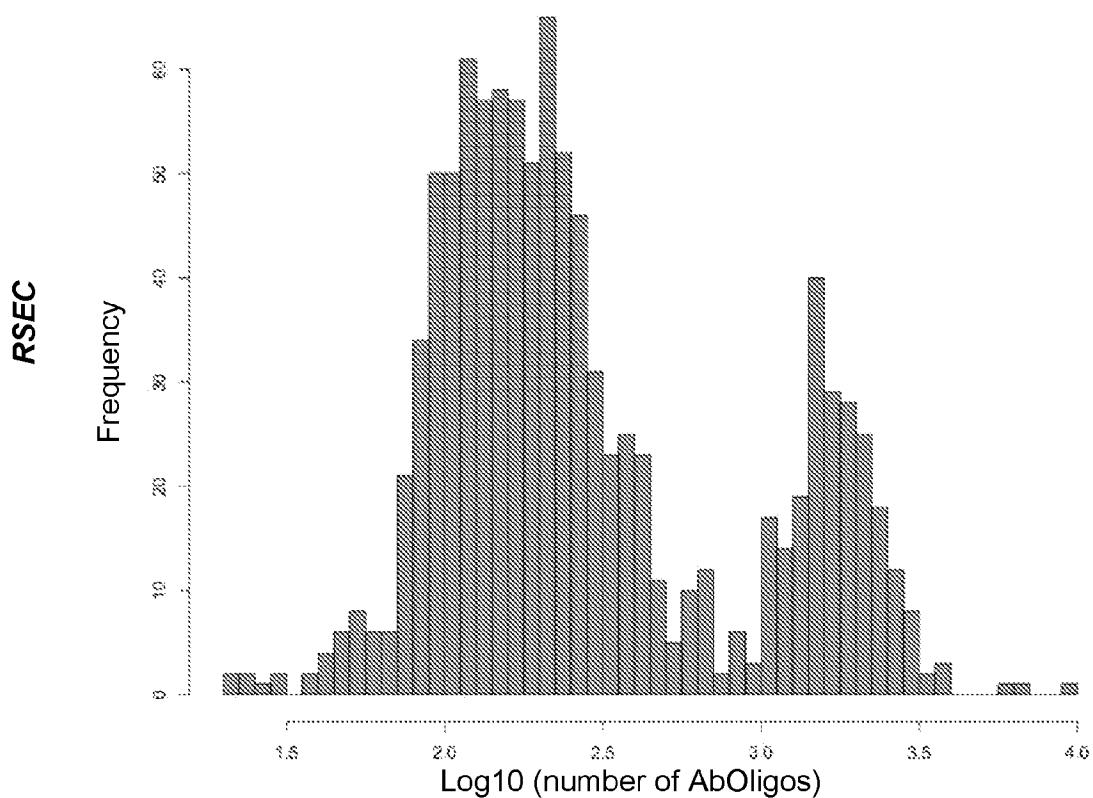


82/112

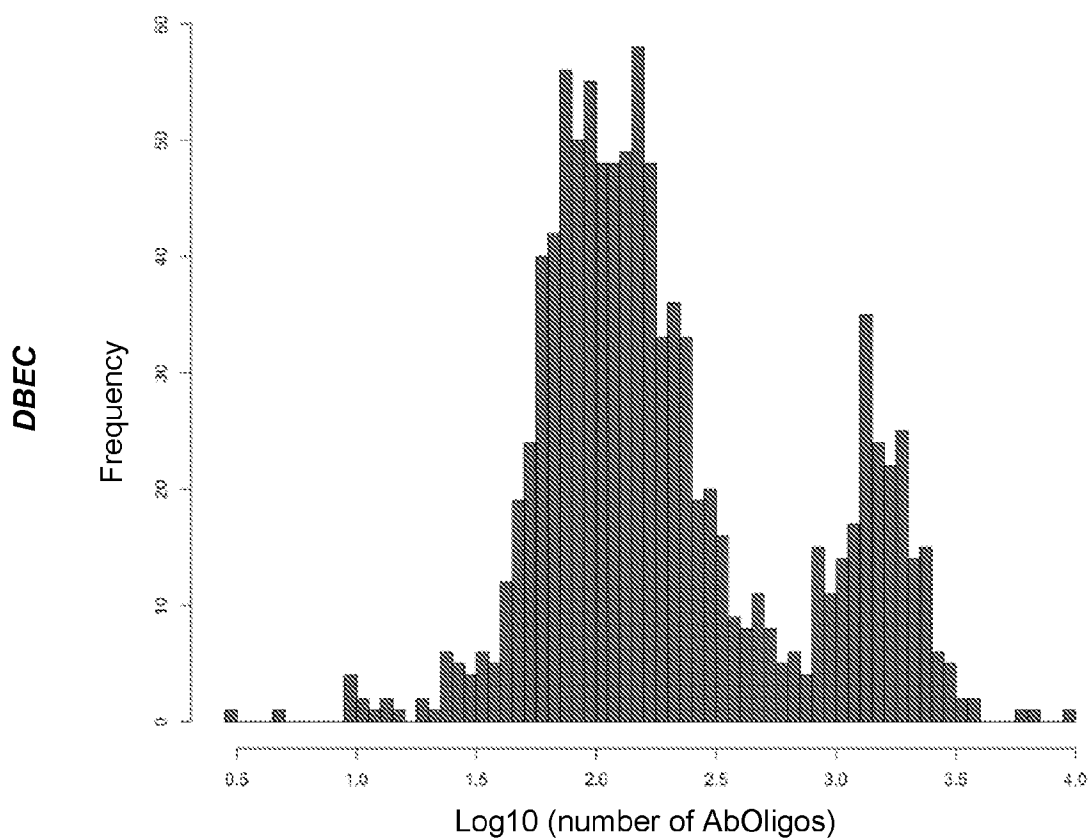
FC4 (1:800 100% hot Ab)

**FIG. 39A3**

Histogram of Log10 (number of AbOligos)

**FIG. 39B3**

Histogram of Log10 (number of AbOligos)



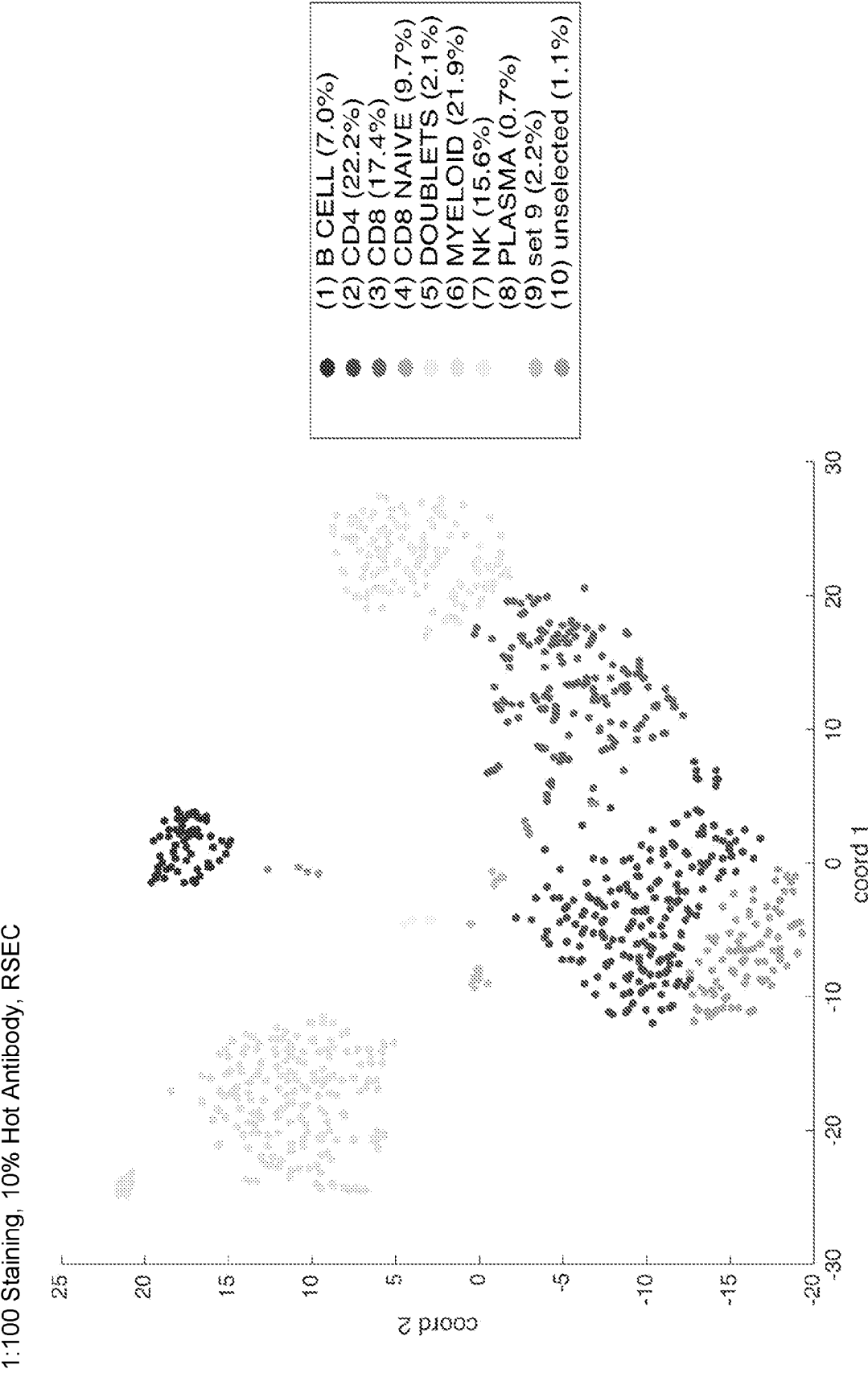
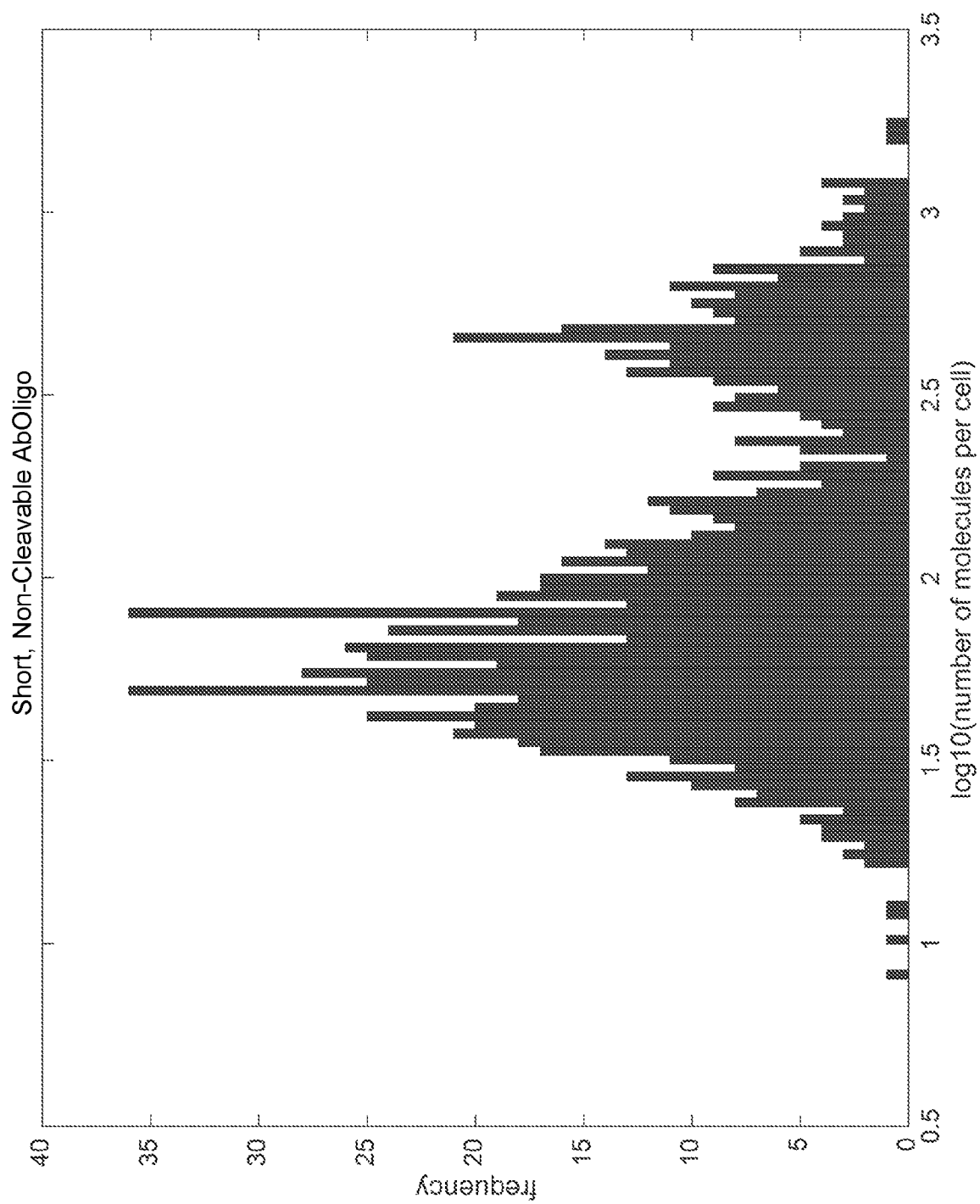


FIG. 40A

84/112

**FIG. 40B**



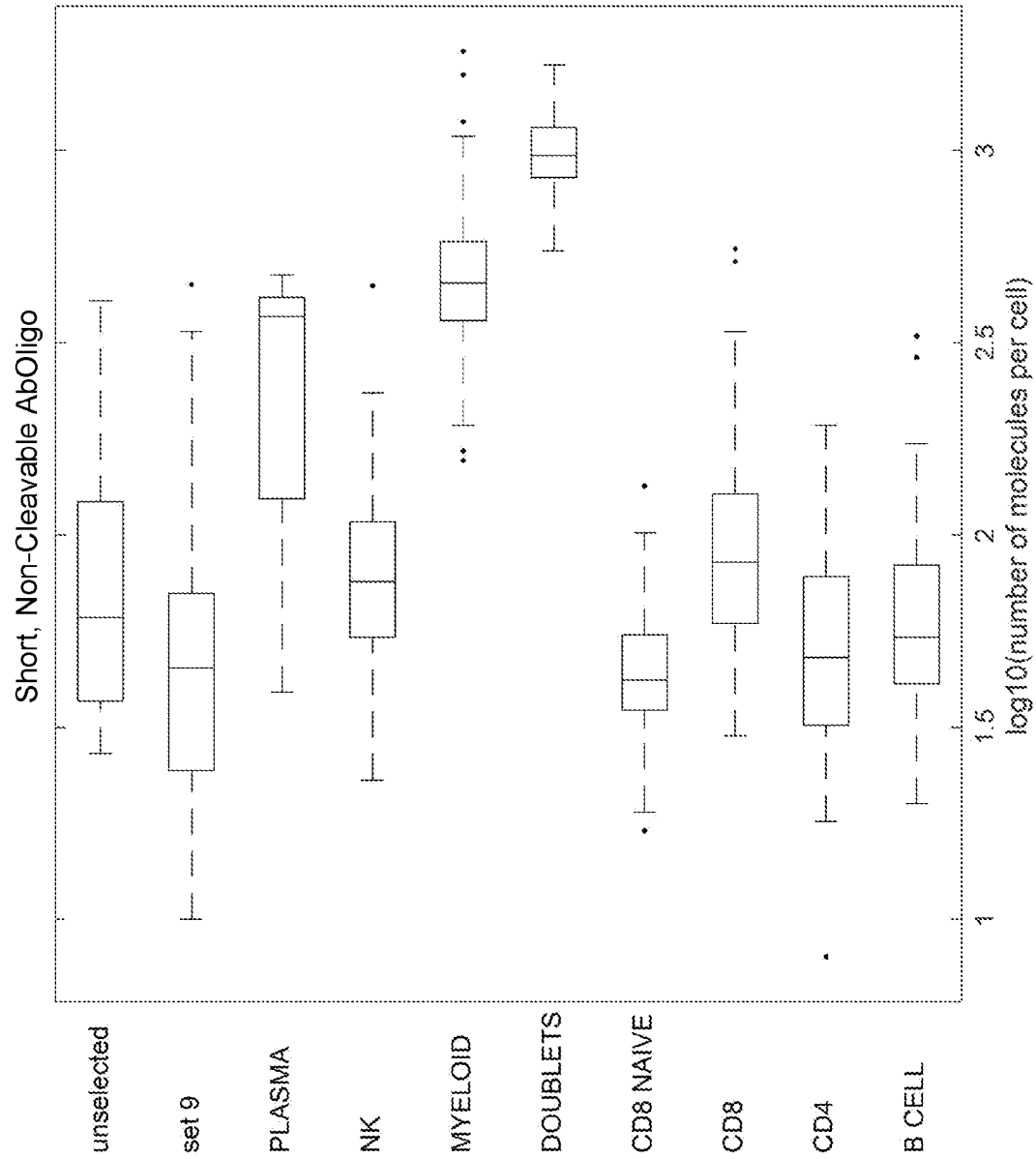
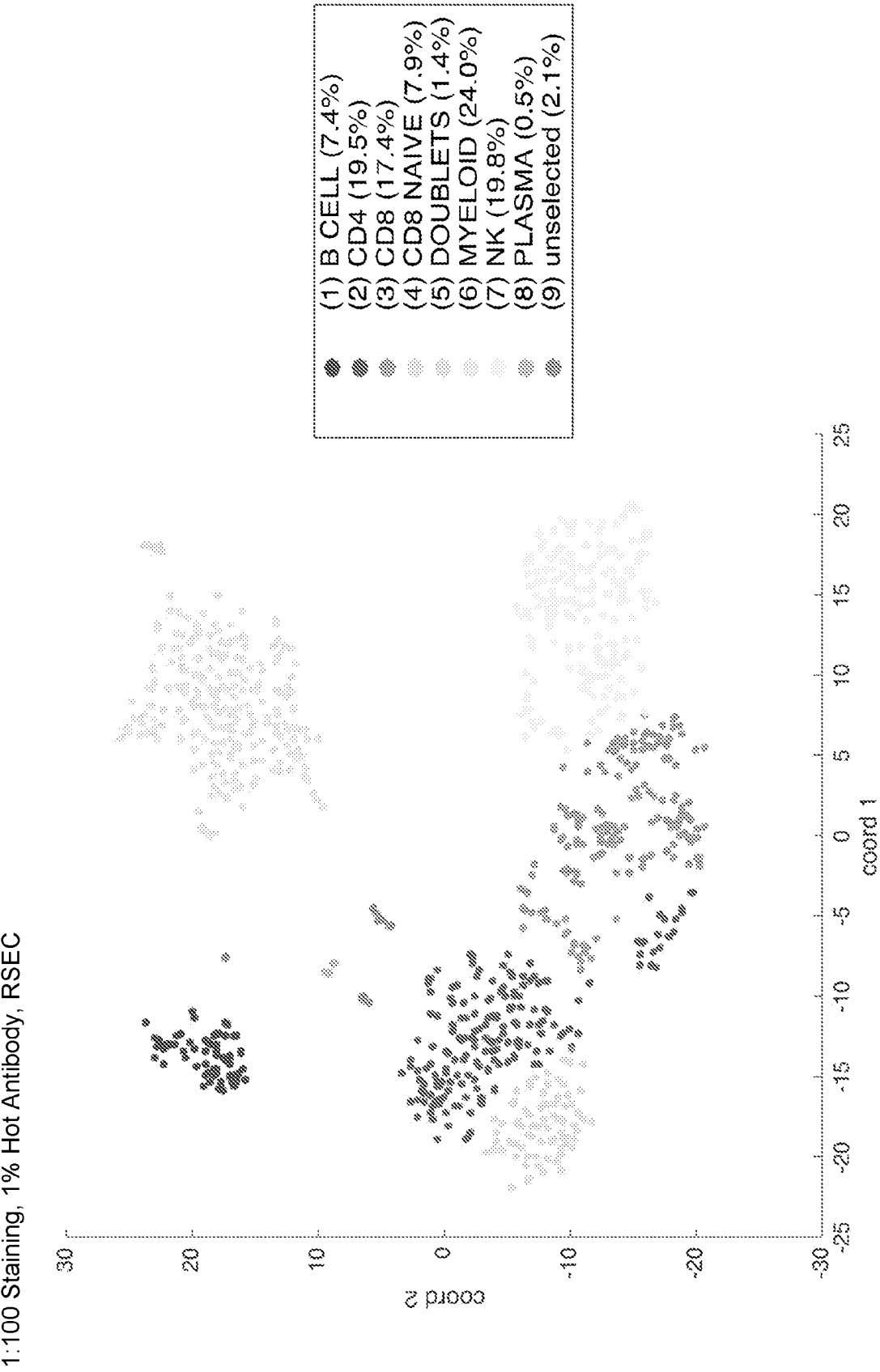
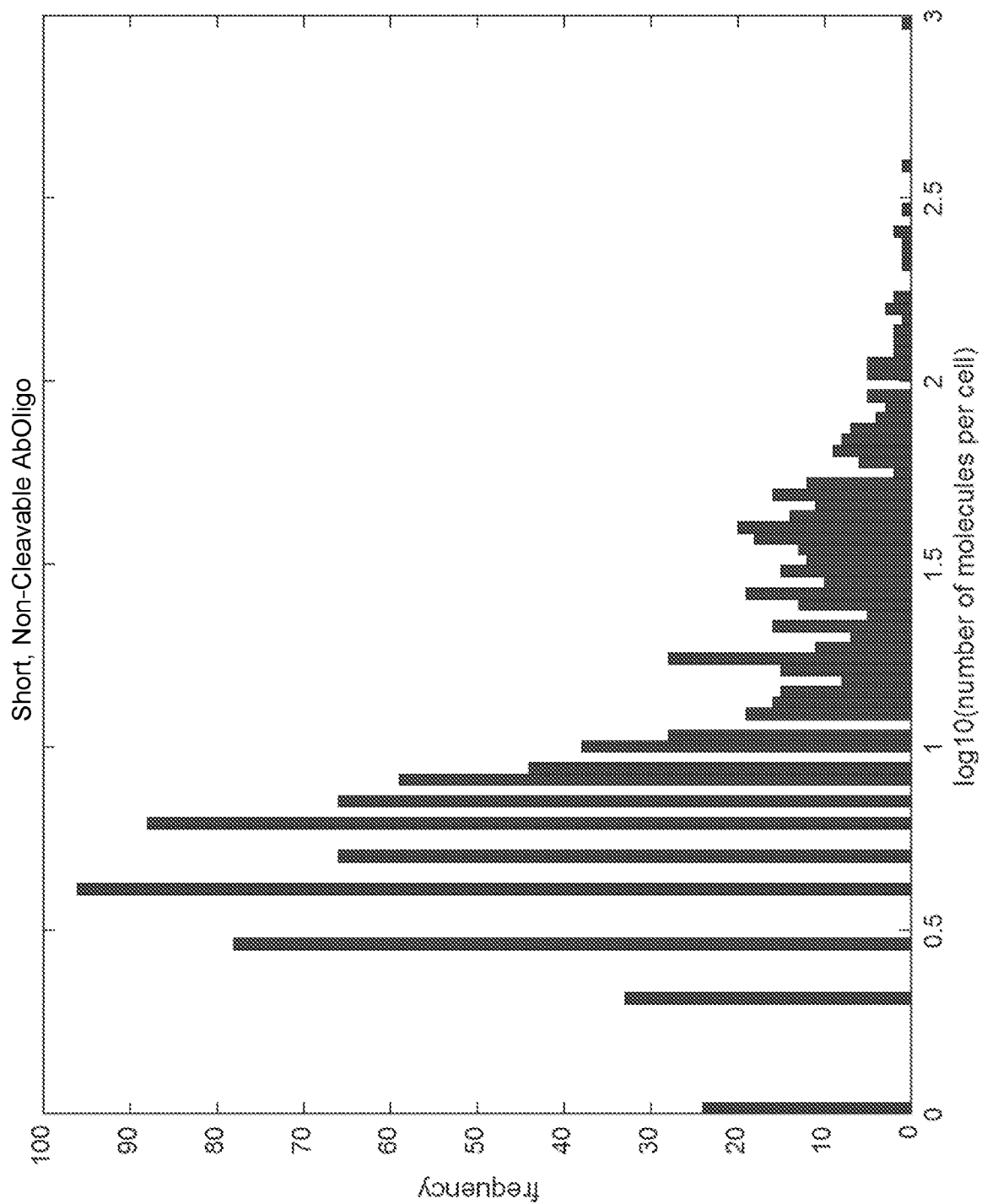


FIG. 40C

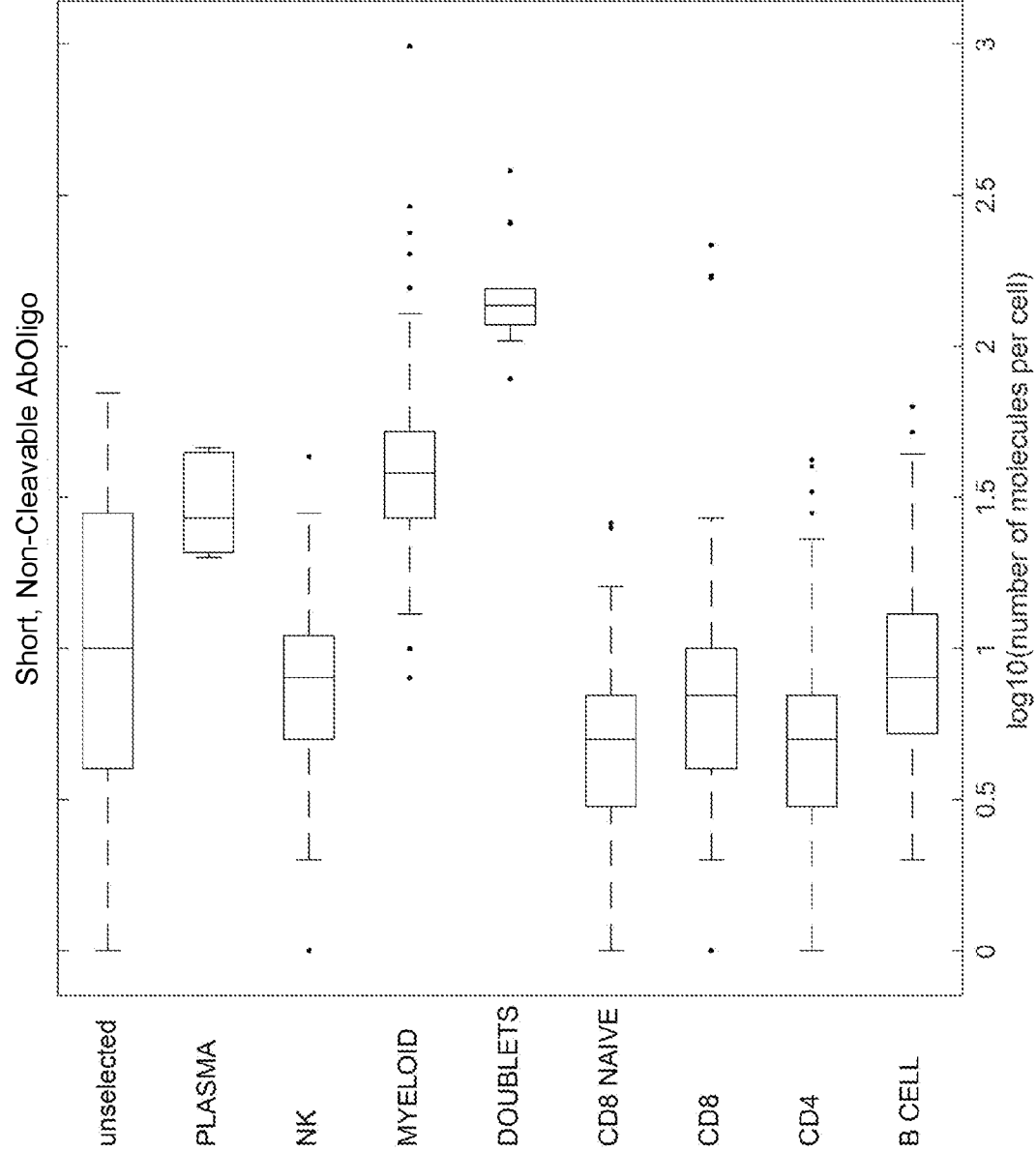


**FIG. 41A**

87/112



**FIG. 41B**



**FIG. 41C**

89/112

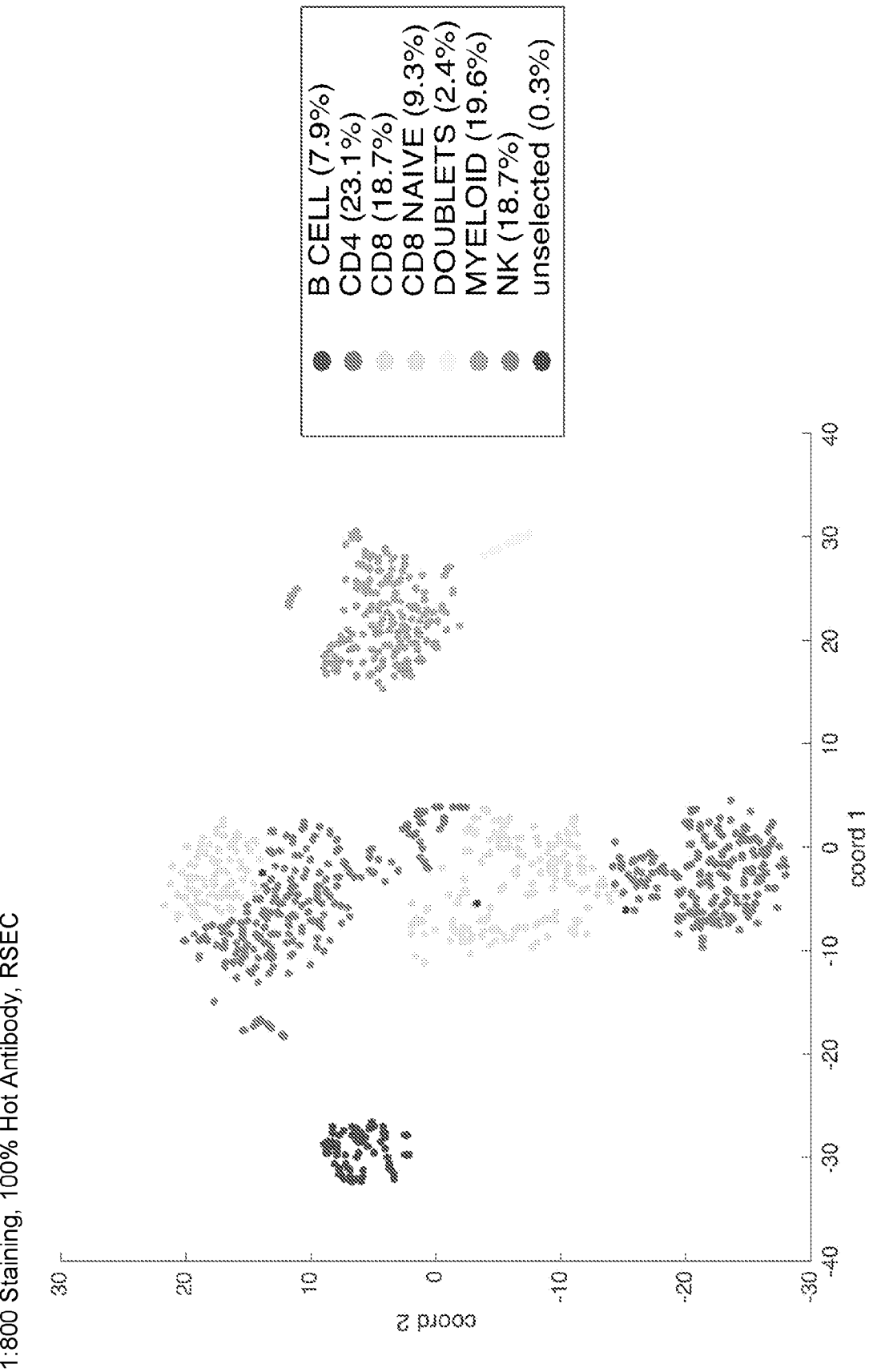
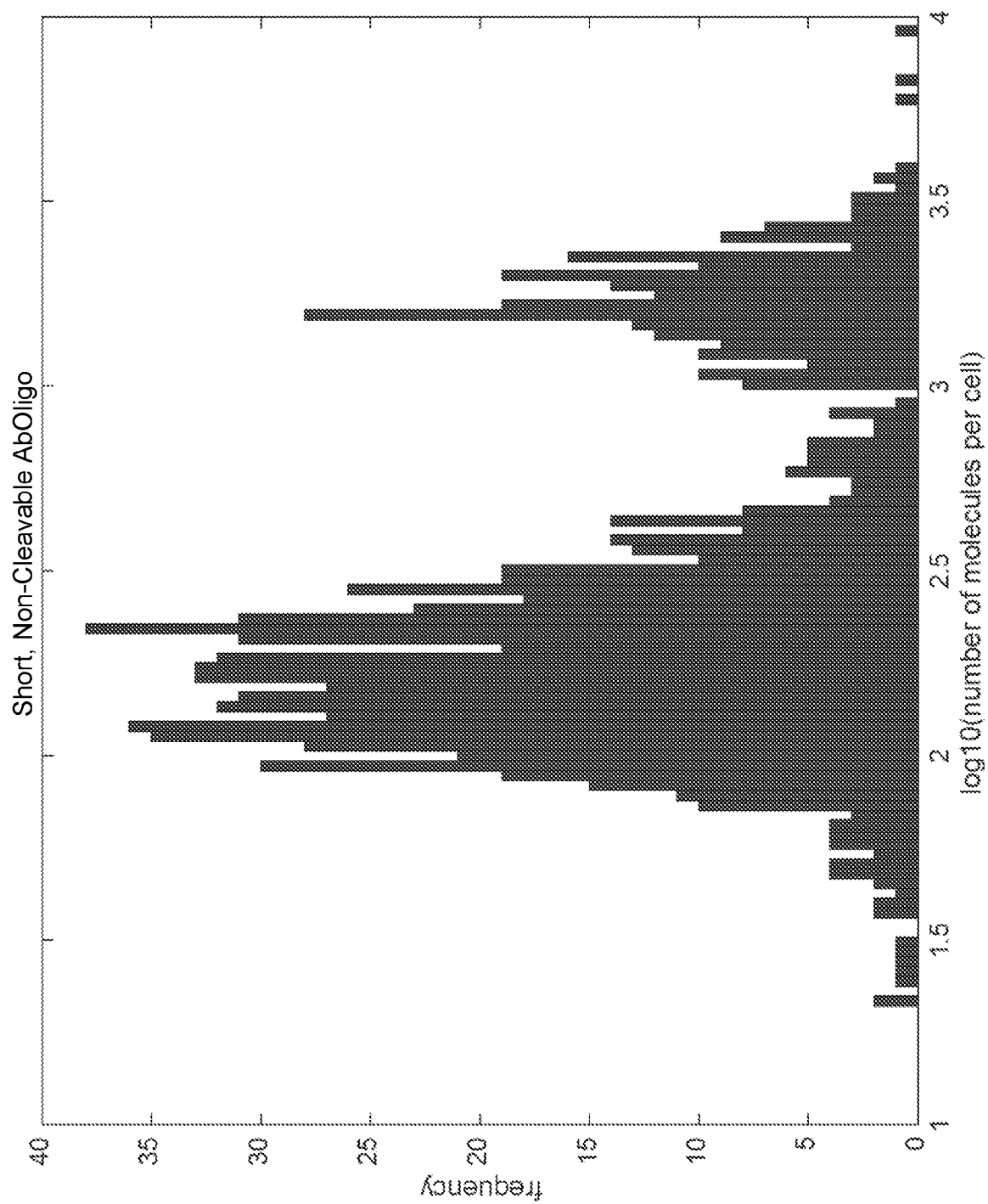
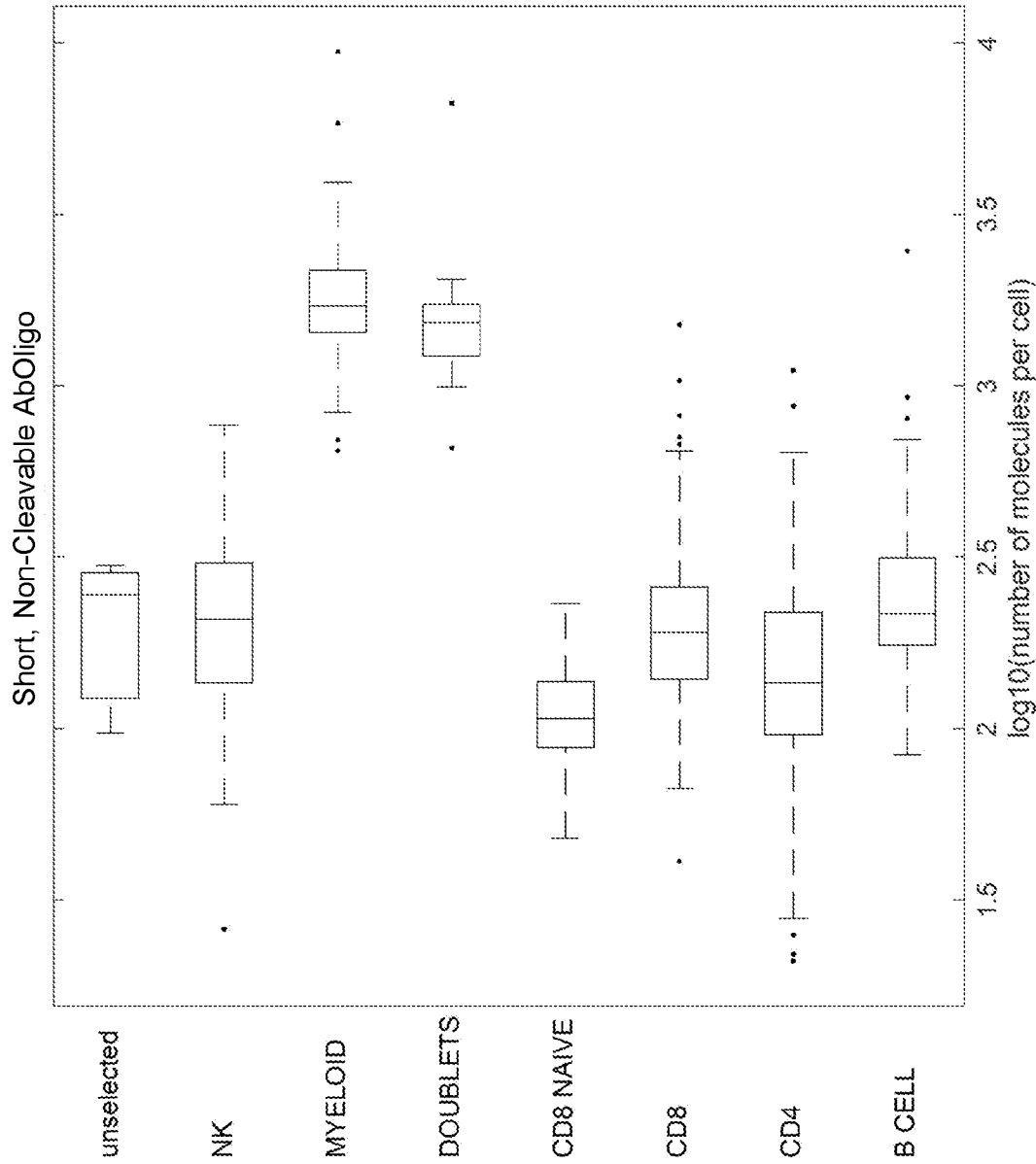


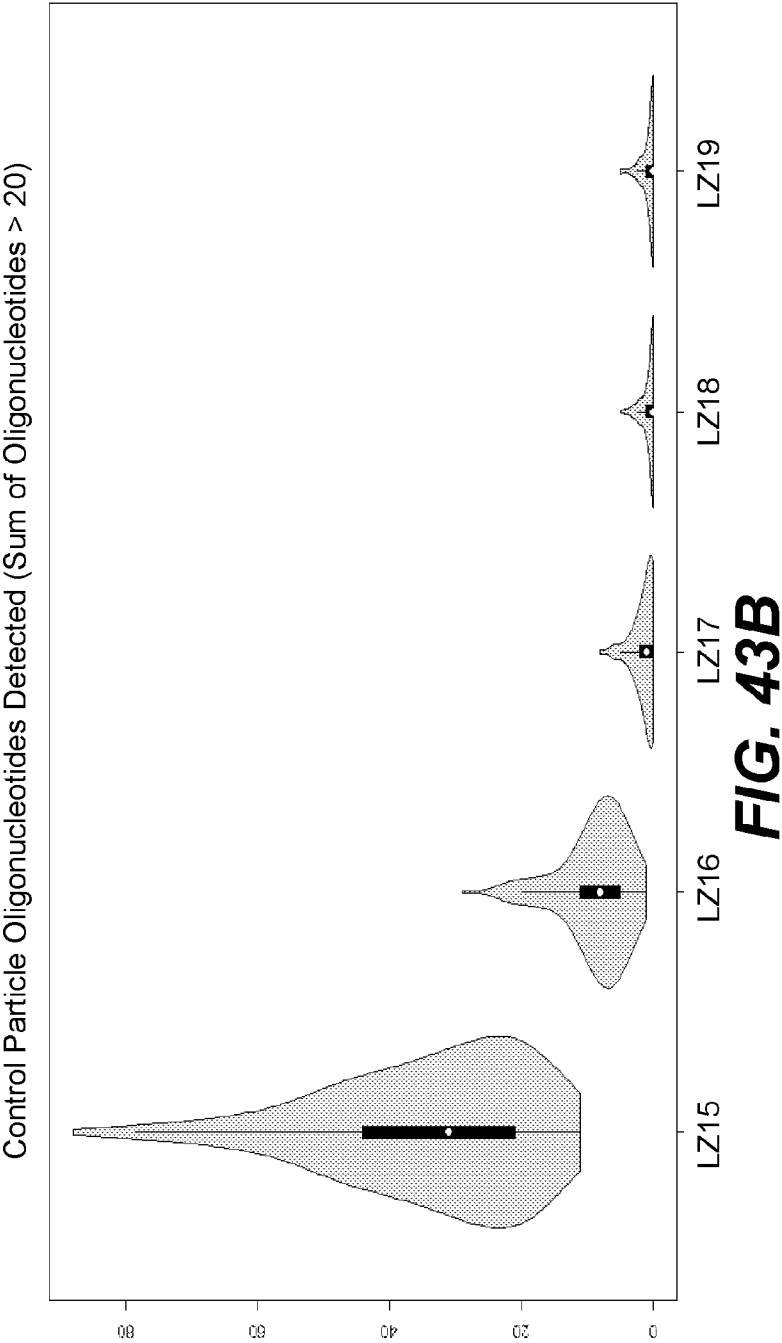
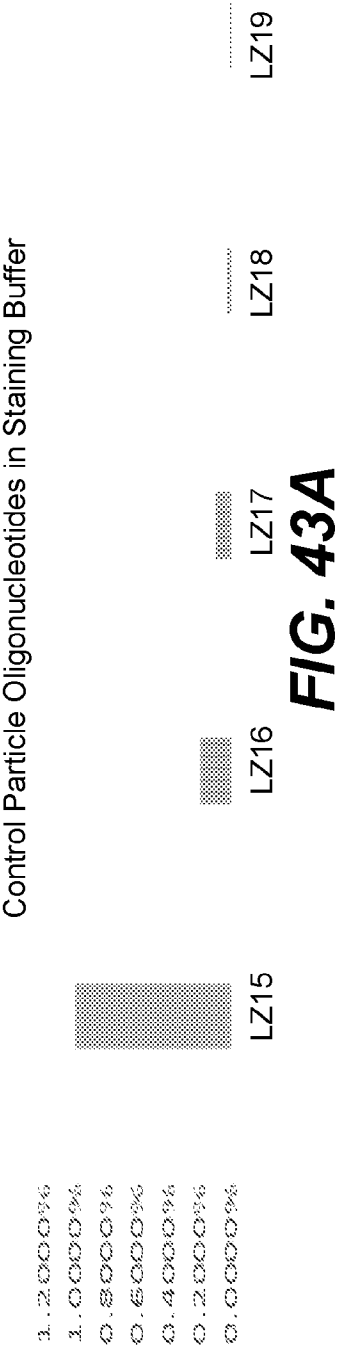
FIG. 42A

90/112

**FIG. 42B**

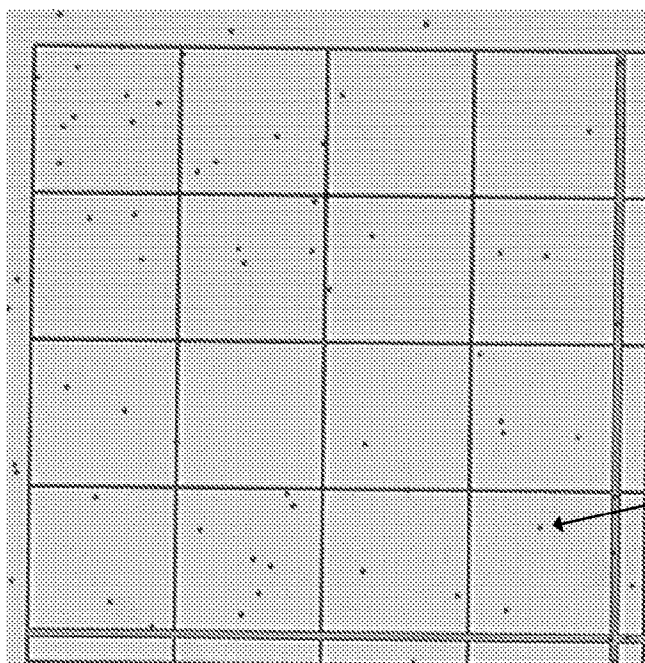


**FIG. 42C**

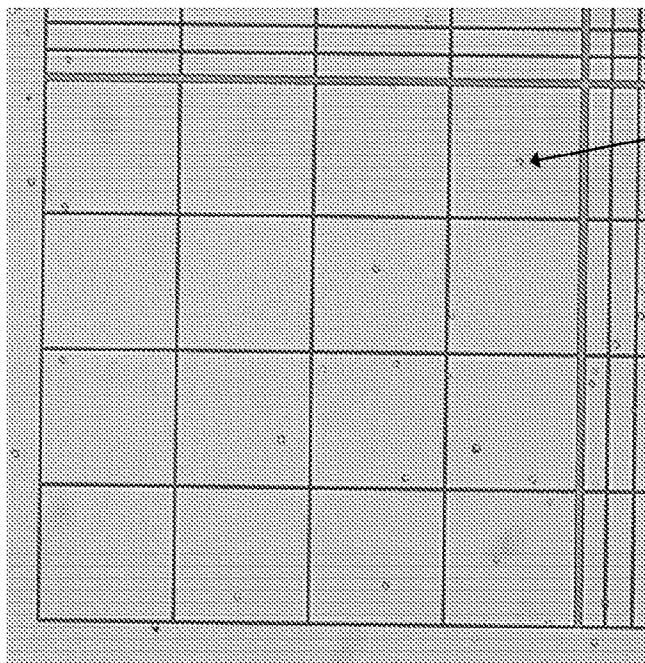




93/112

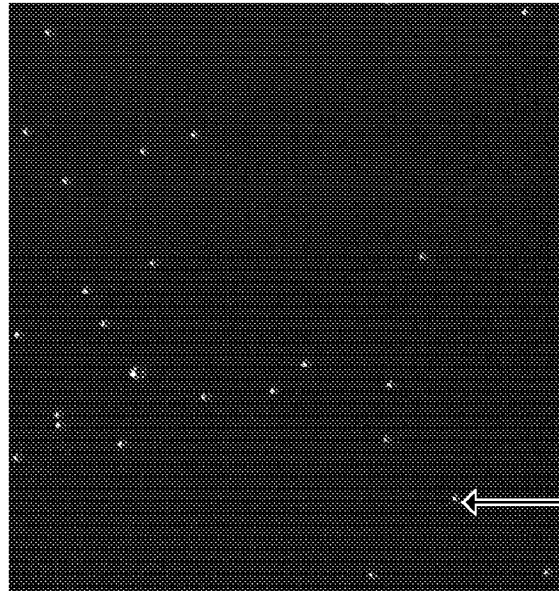


Control particle  
**FIG. 44B**



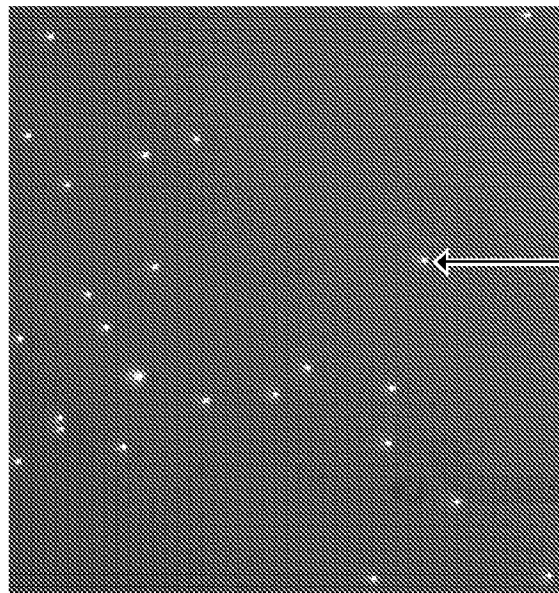
Cell  
**FIG. 44A**

94/112



Control particle

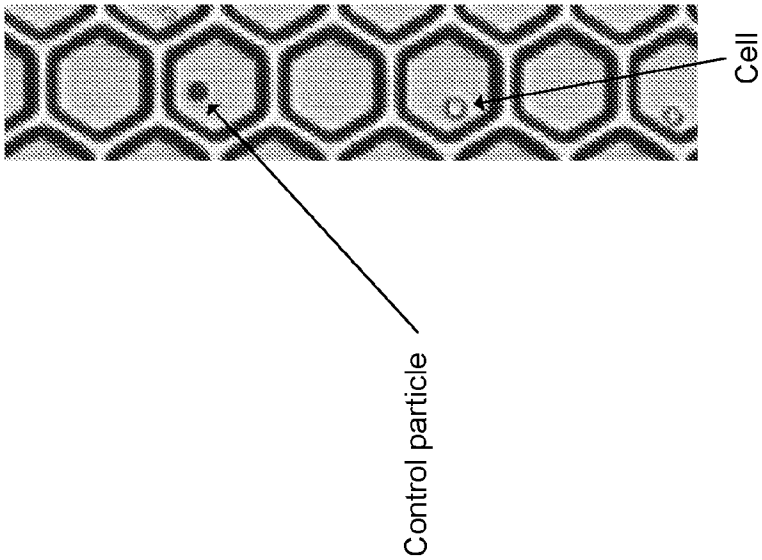
**FIG. 45B**



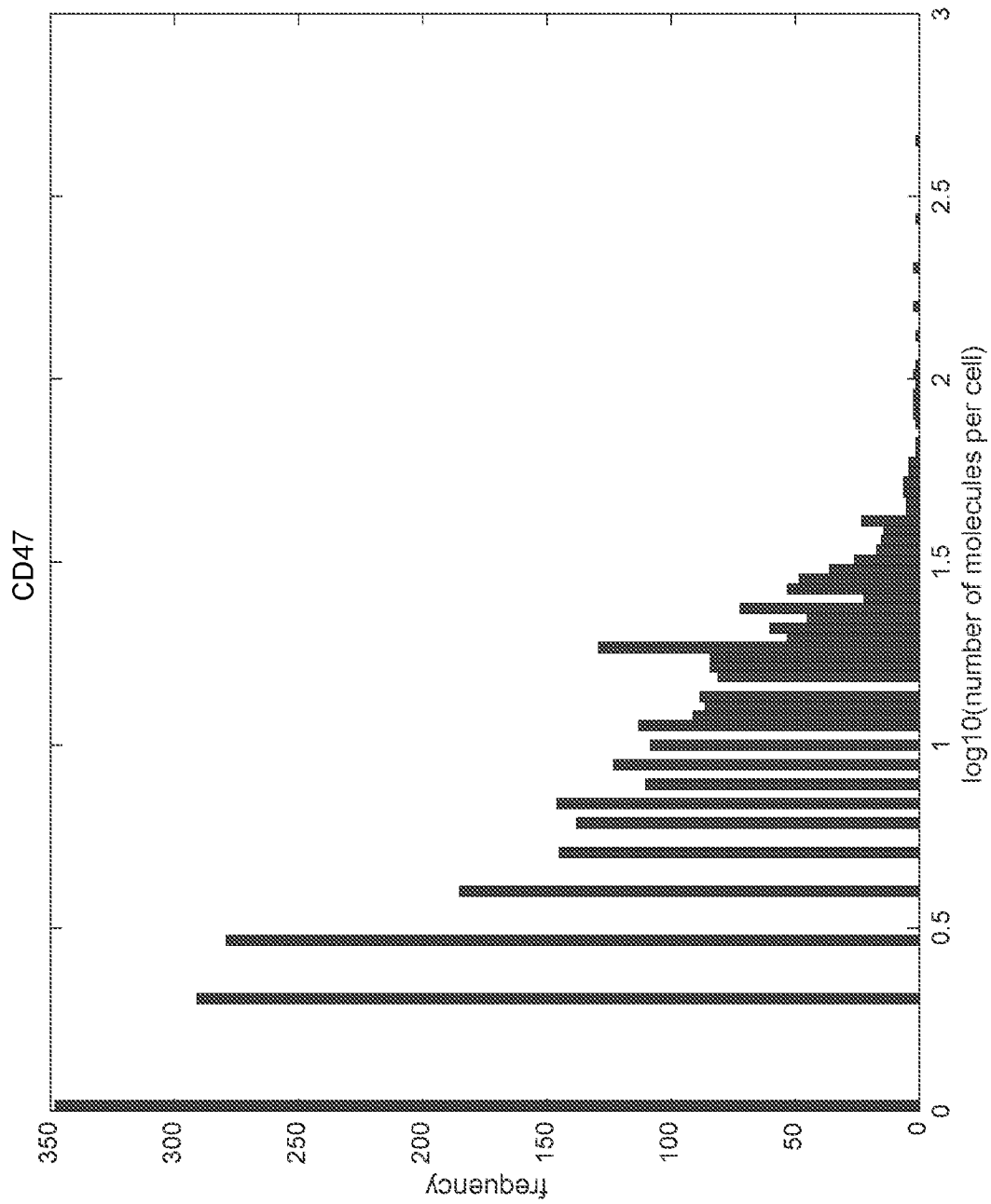
Control particle

**FIG. 45A**

95/112

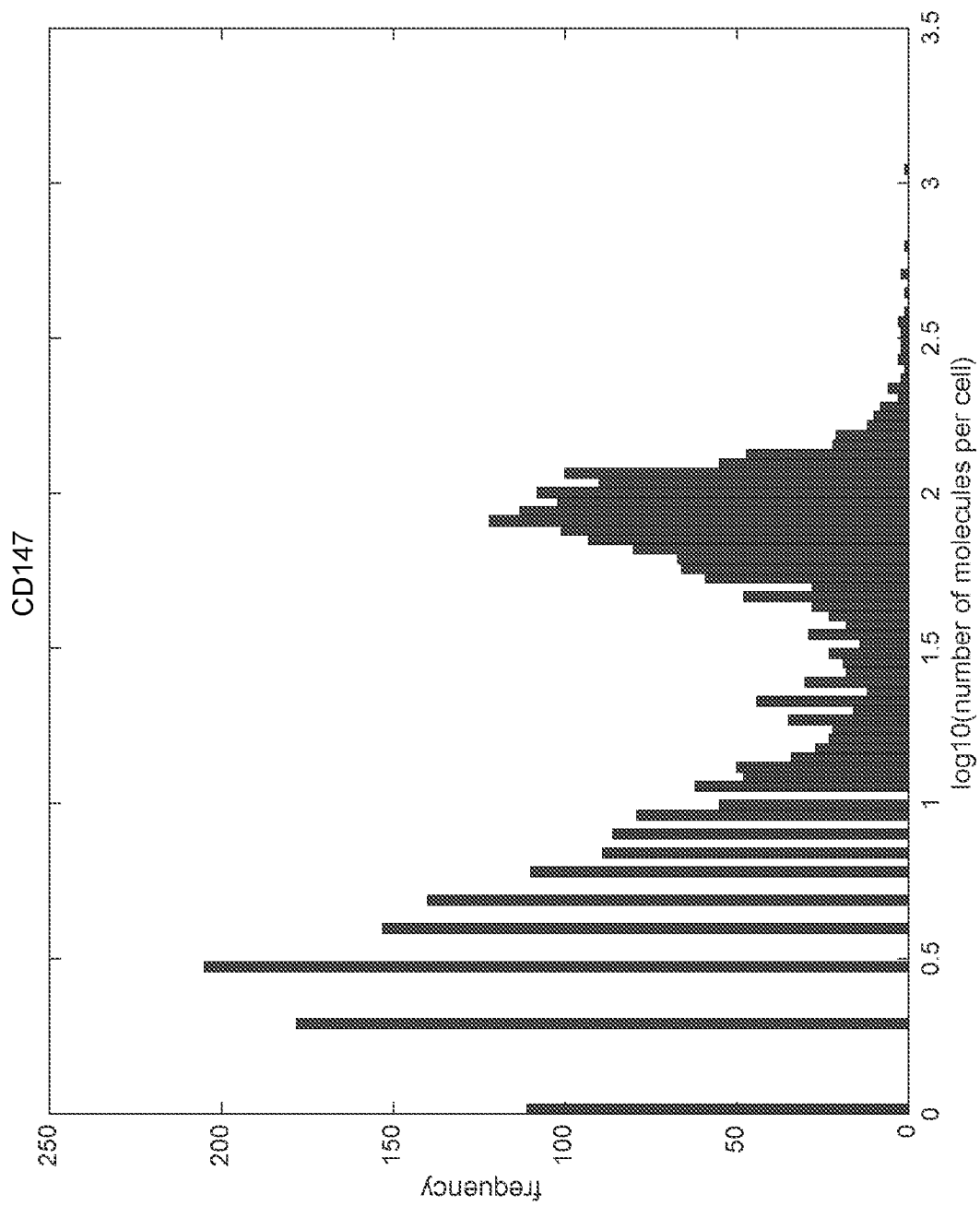


**FIG. 46**

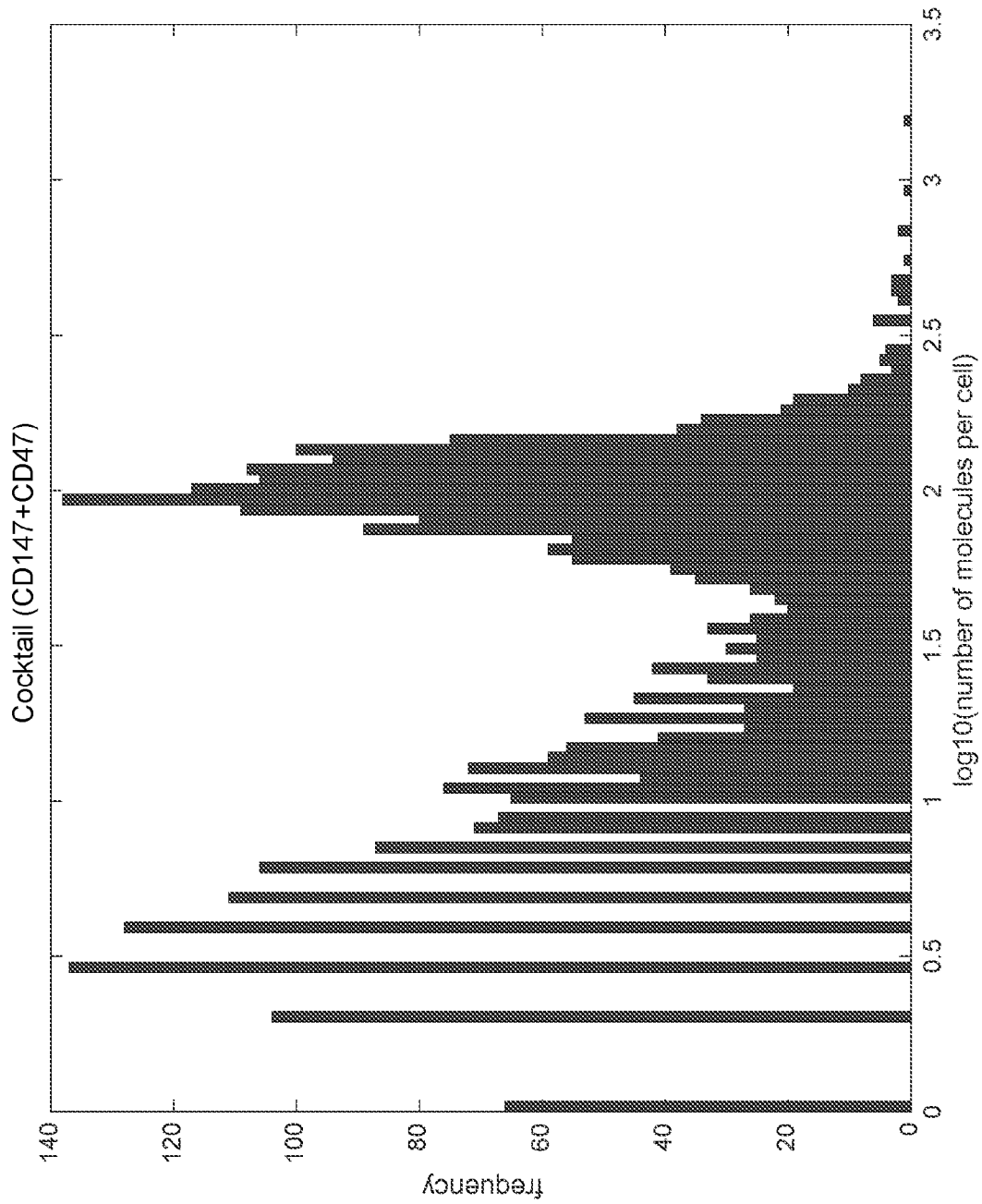


**FIG. 47A**

97/112



**FIG. 47B**



**FIG. 47C**

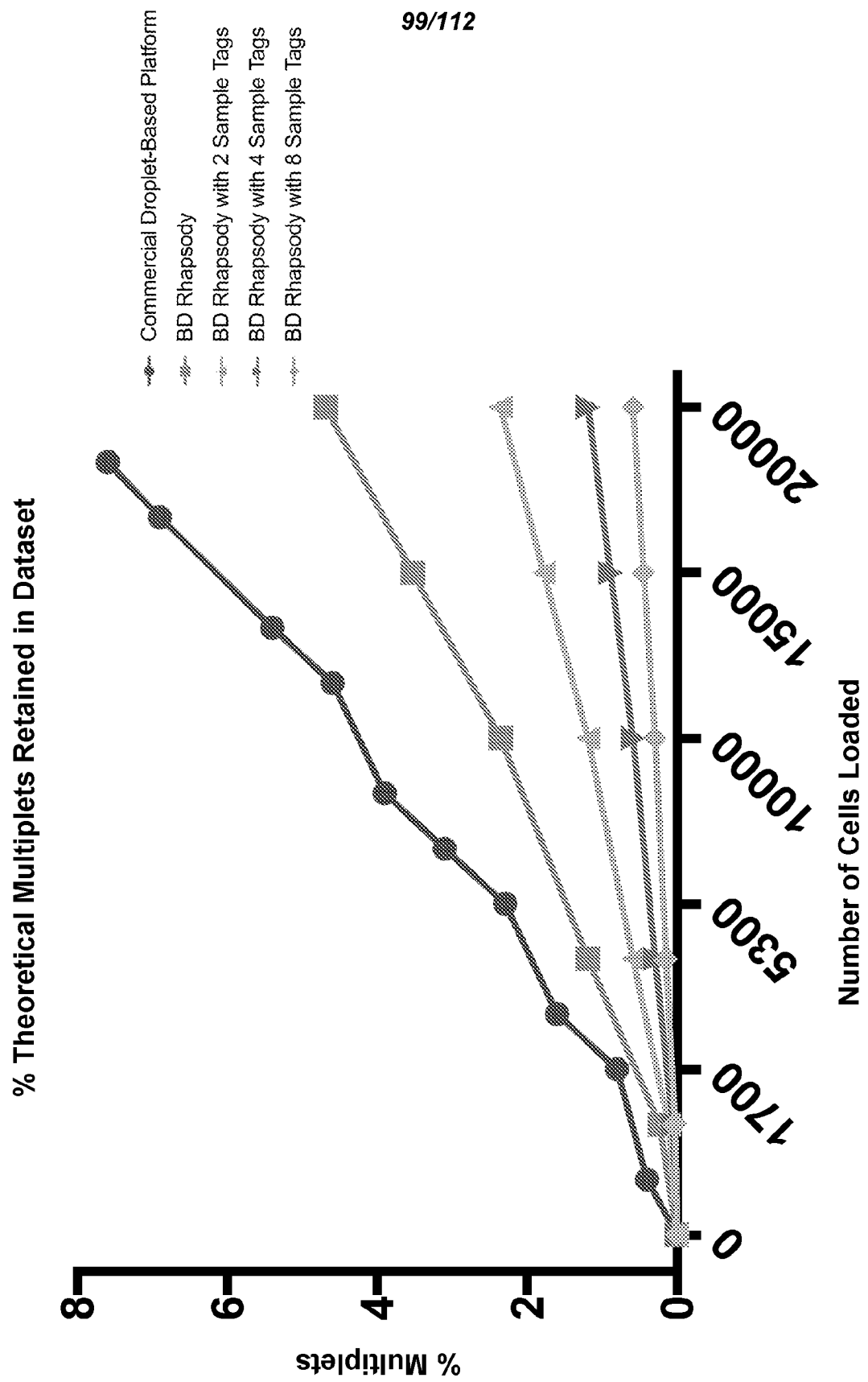
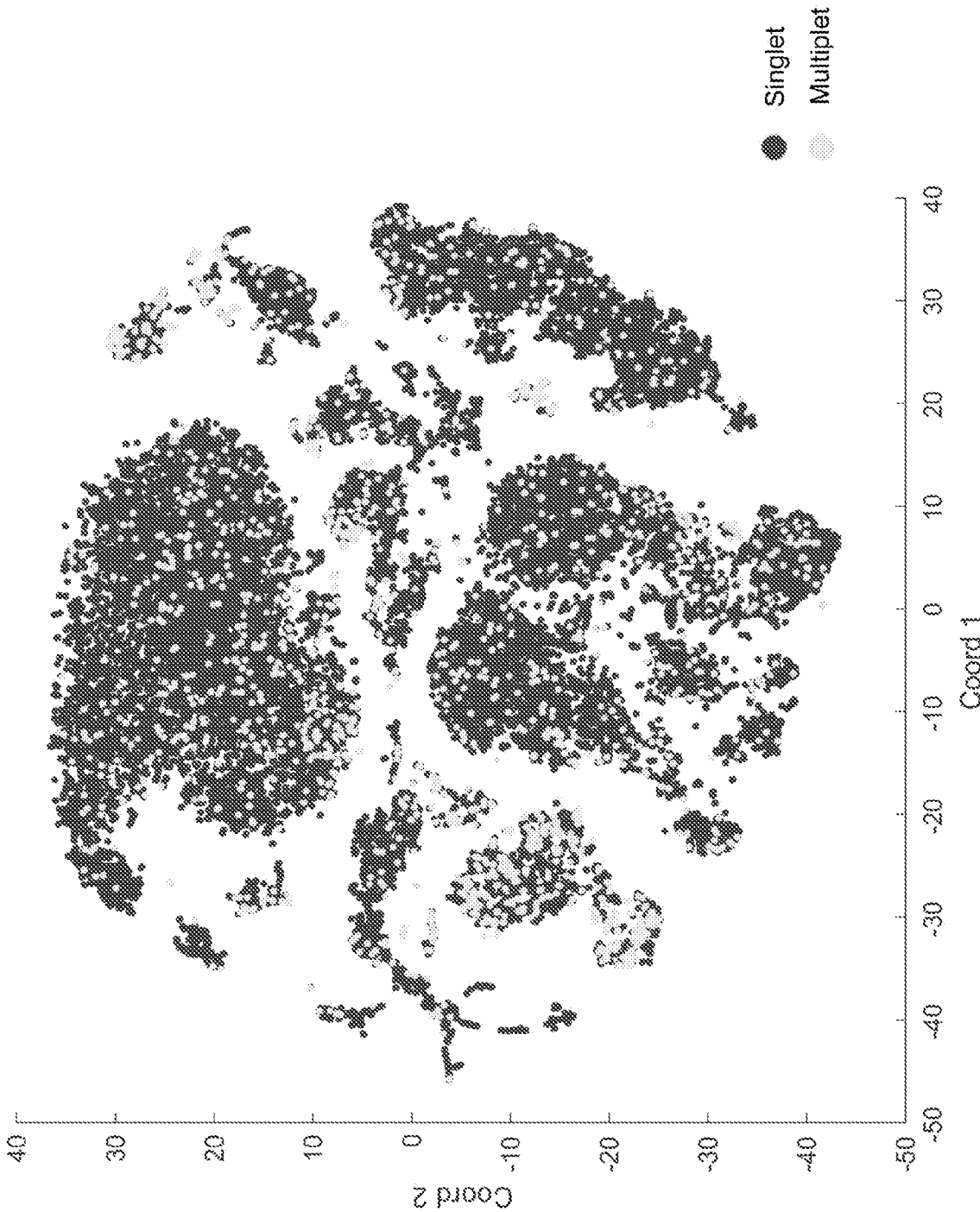


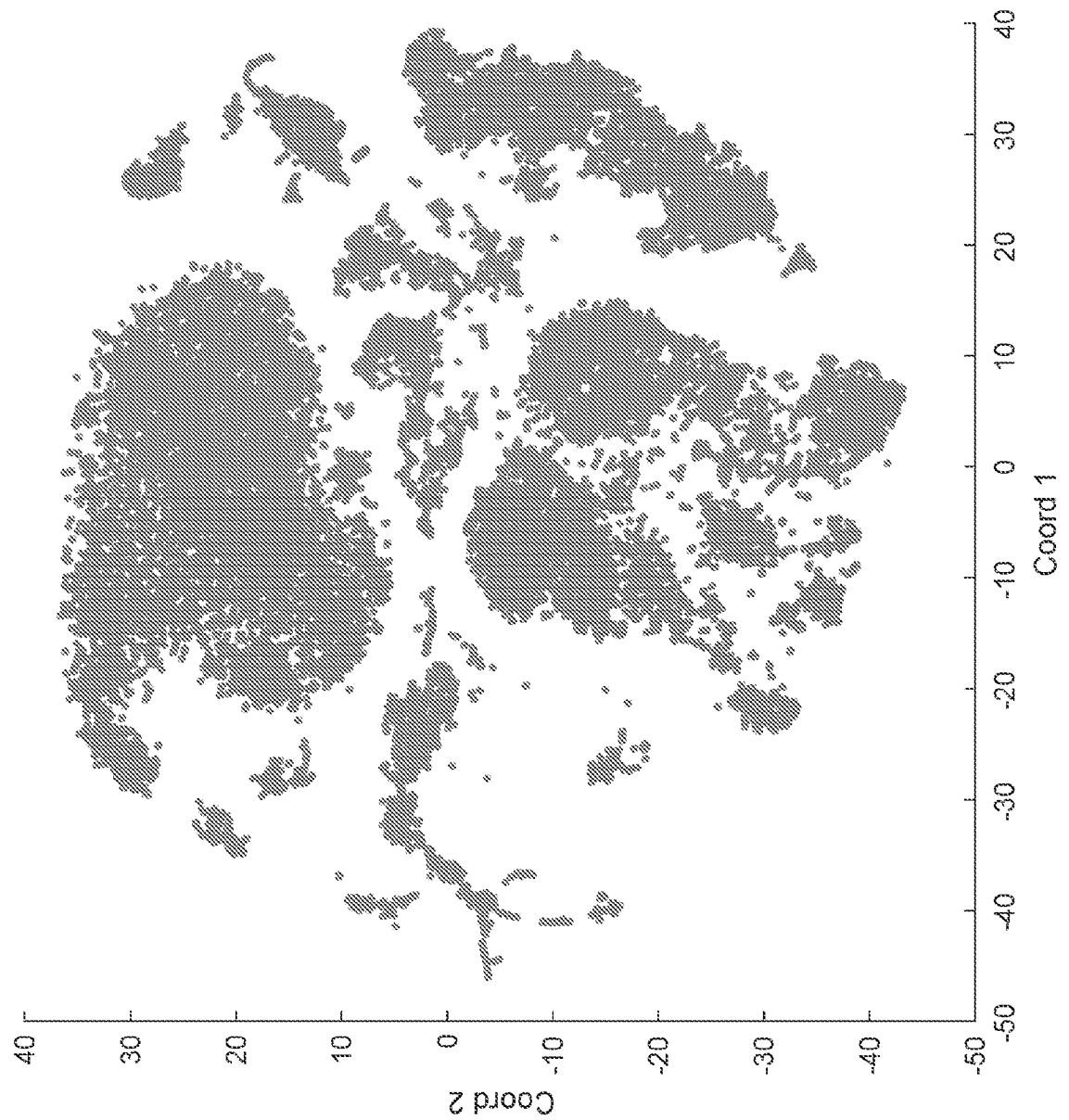
FIG. 48



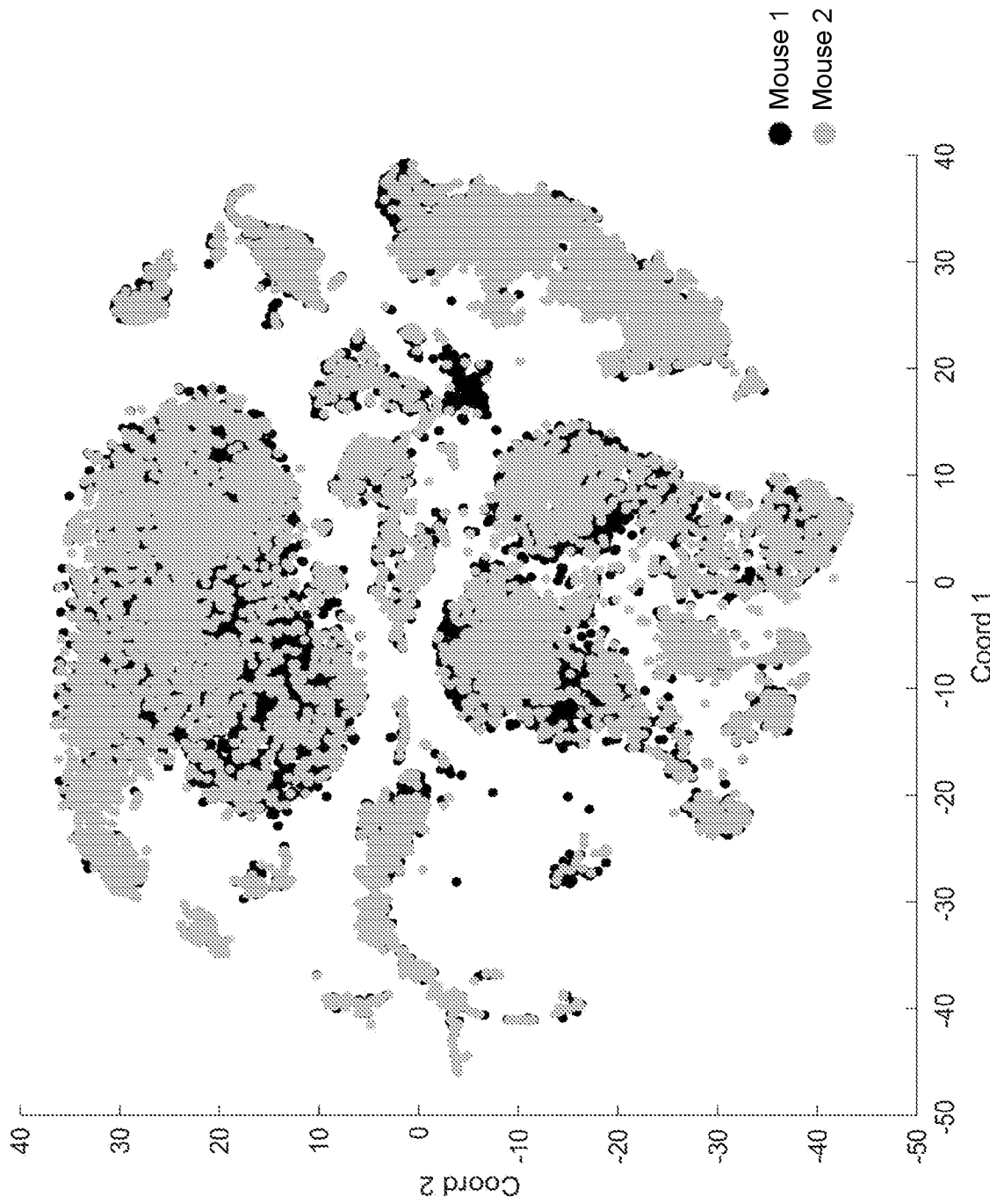
**FIG. 49A**



101/112



**FIG. 49B**



**FIG. 50A**

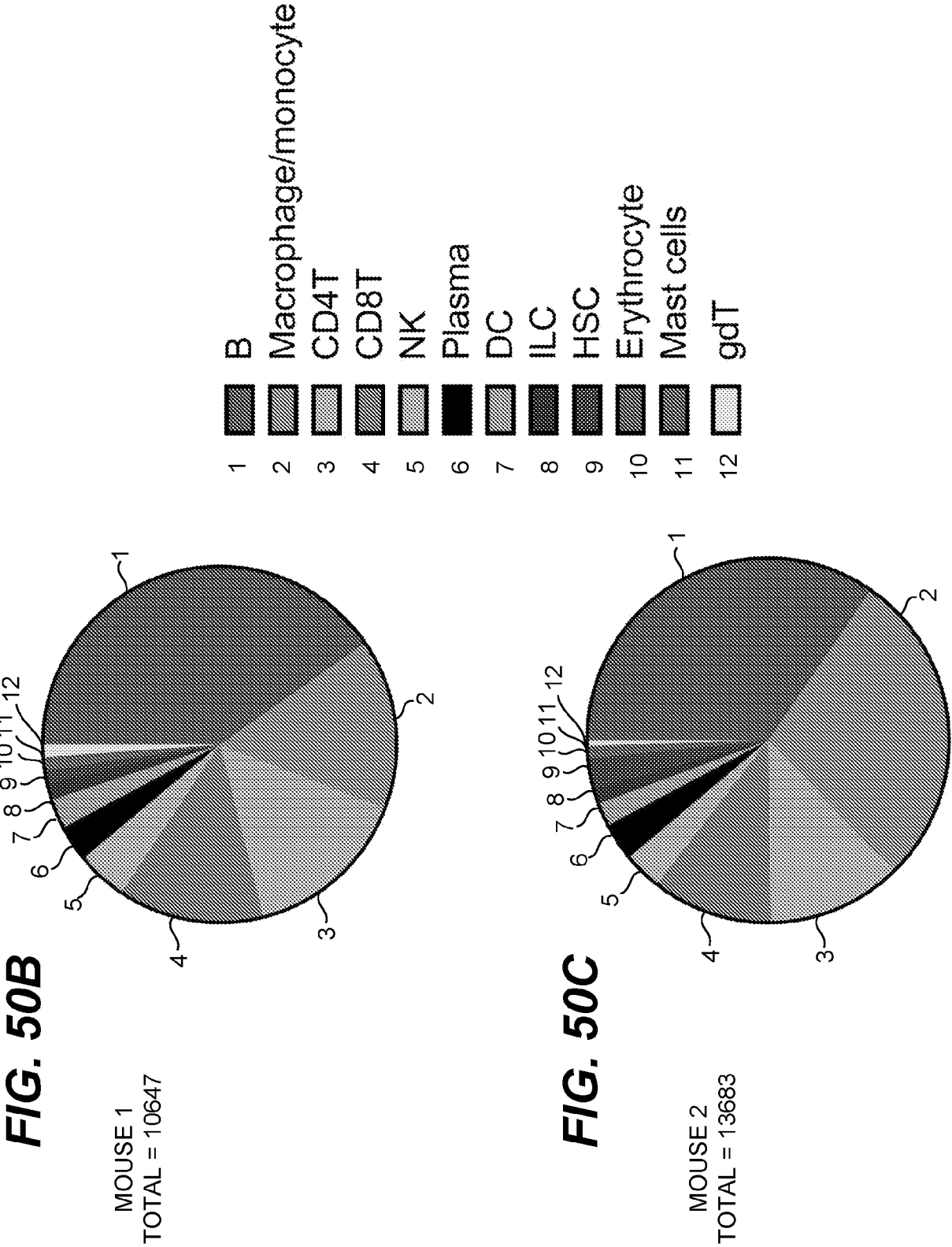


FIG. 51A

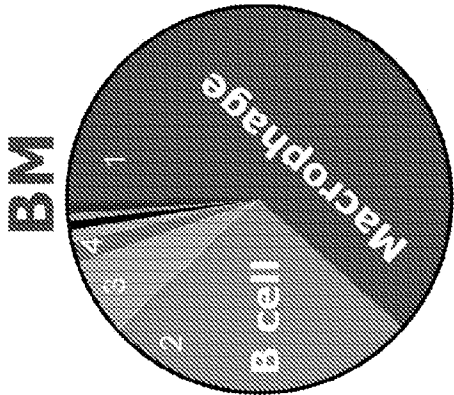


FIG. 51B

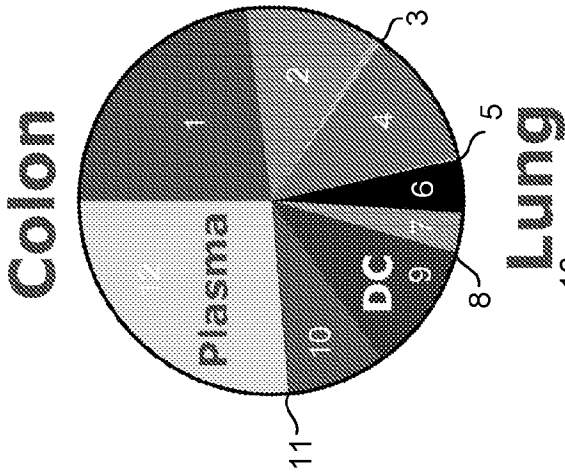
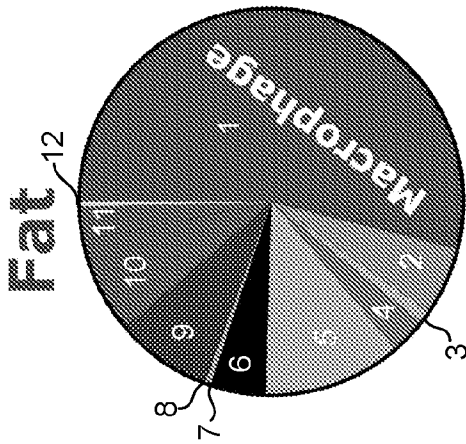
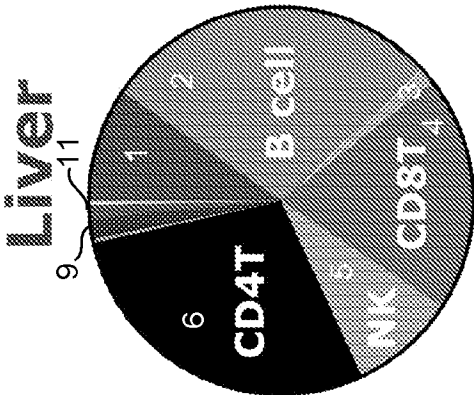


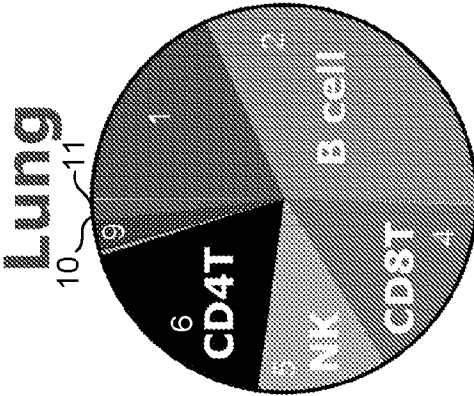
FIG. 51C



**Liver**



**Lung**



**Spleen**

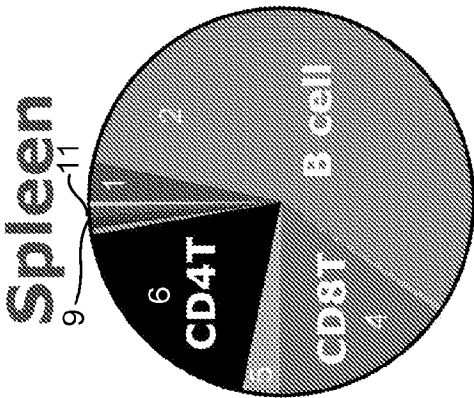
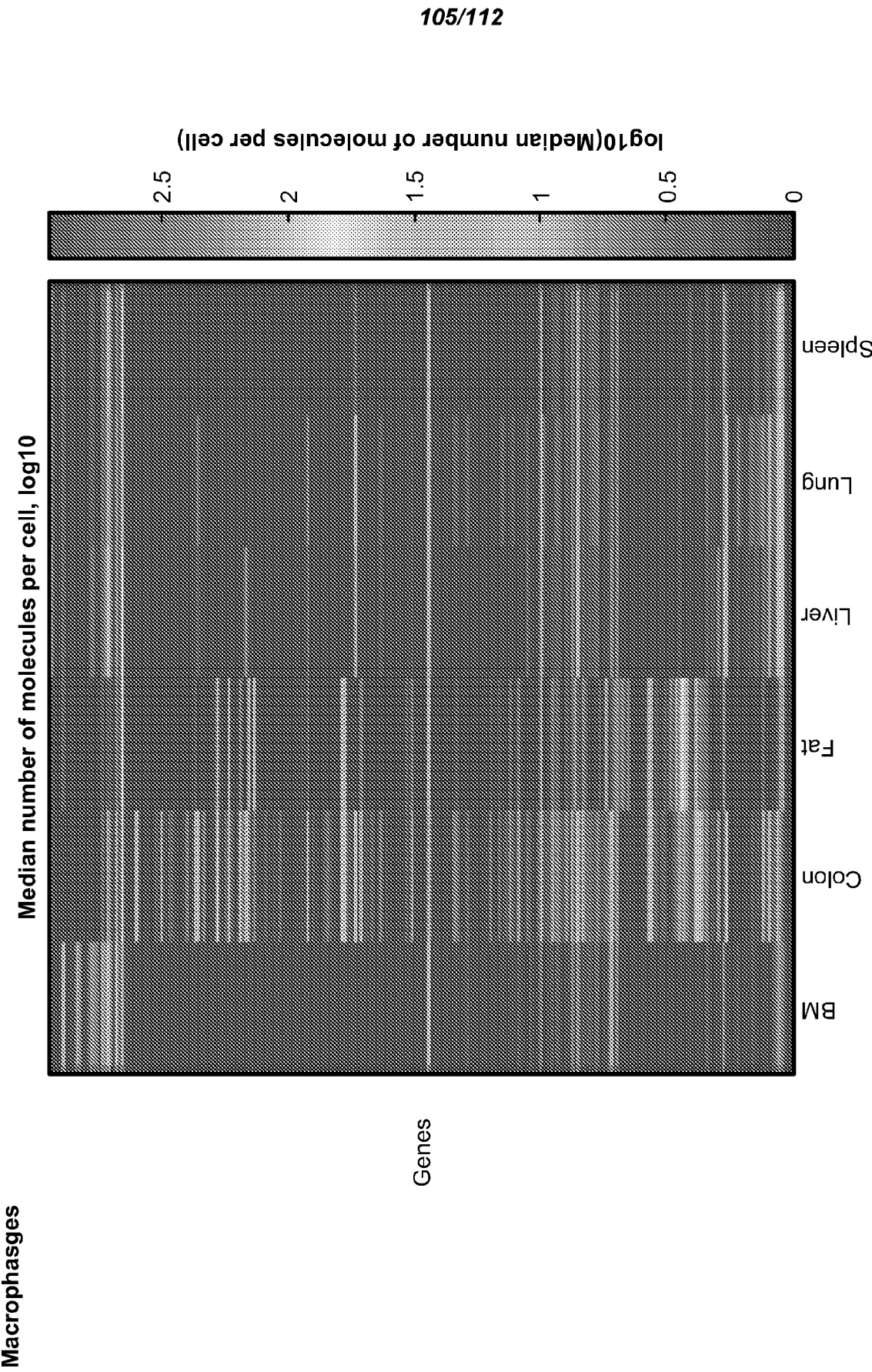


FIG. 51D

FIG. 51E

FIG. 51F



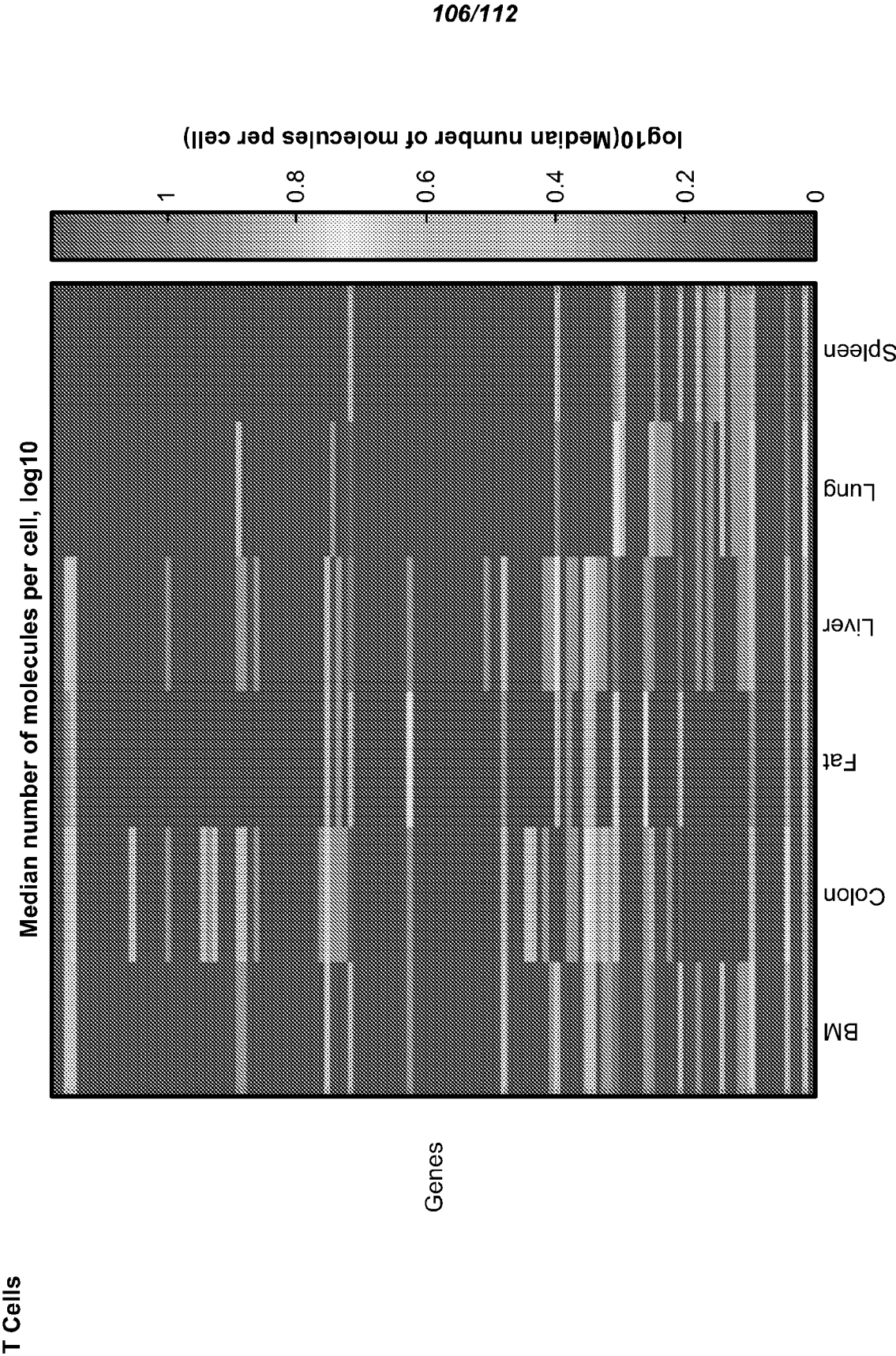
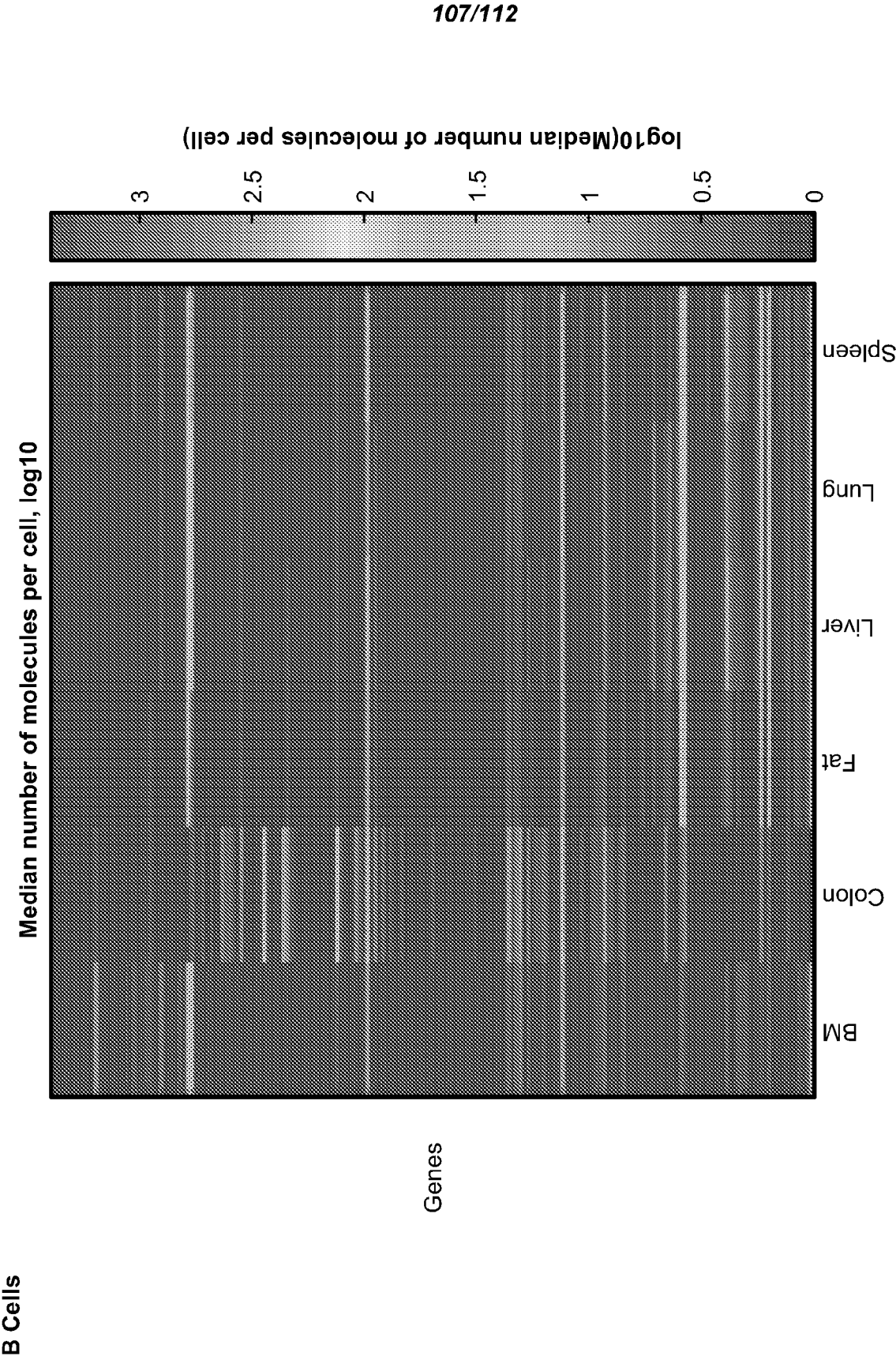
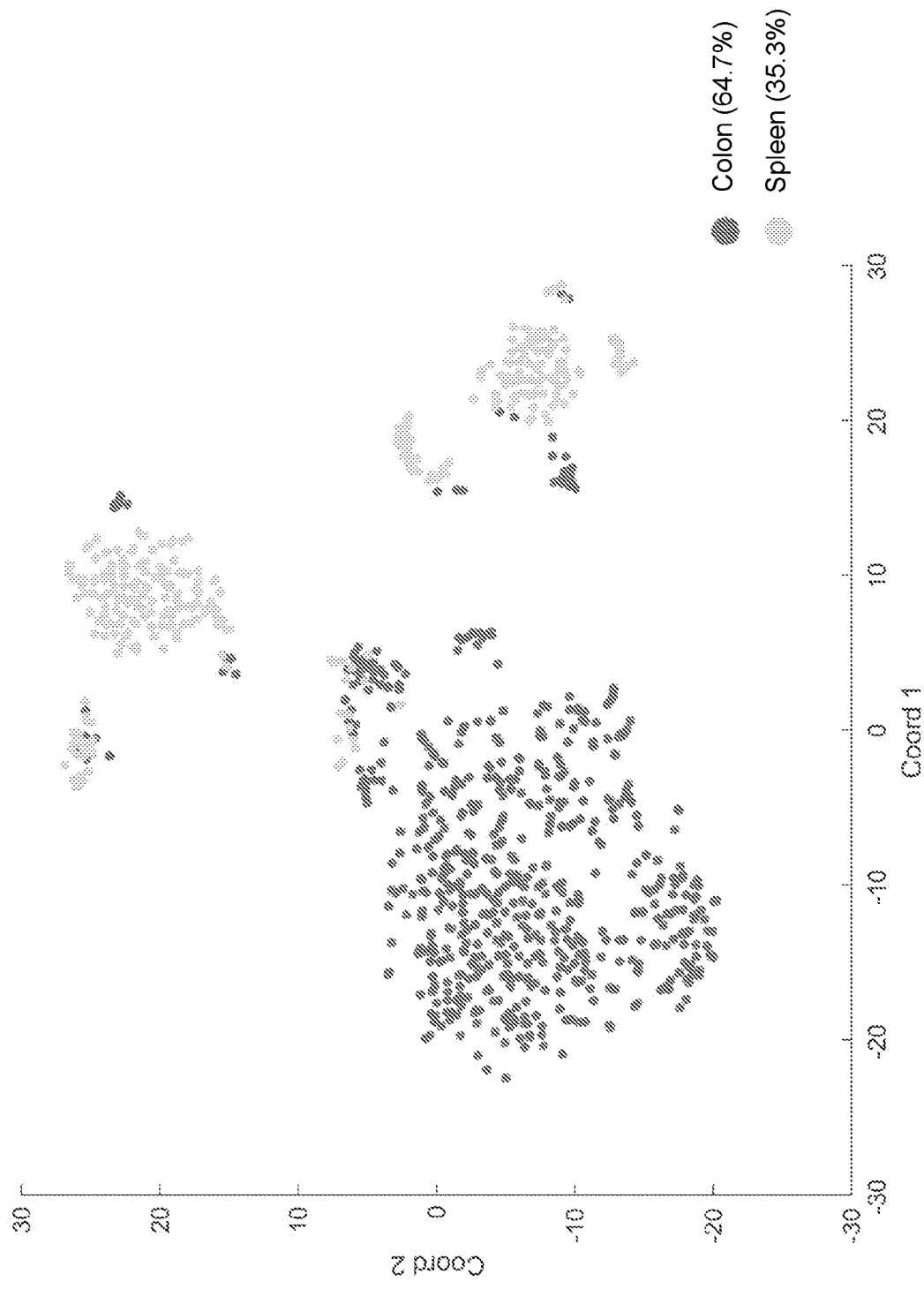


FIG. 52B

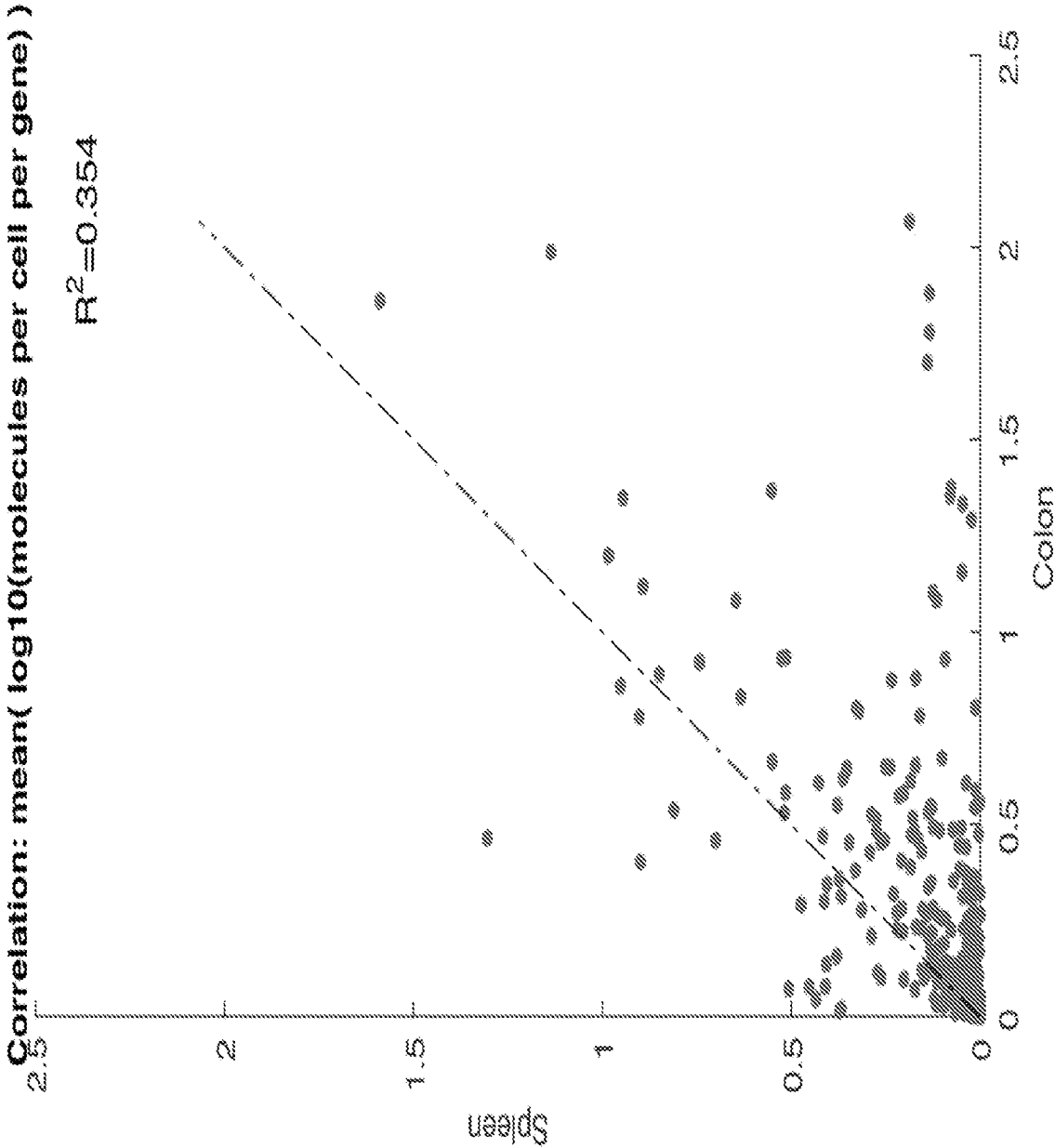


108/112



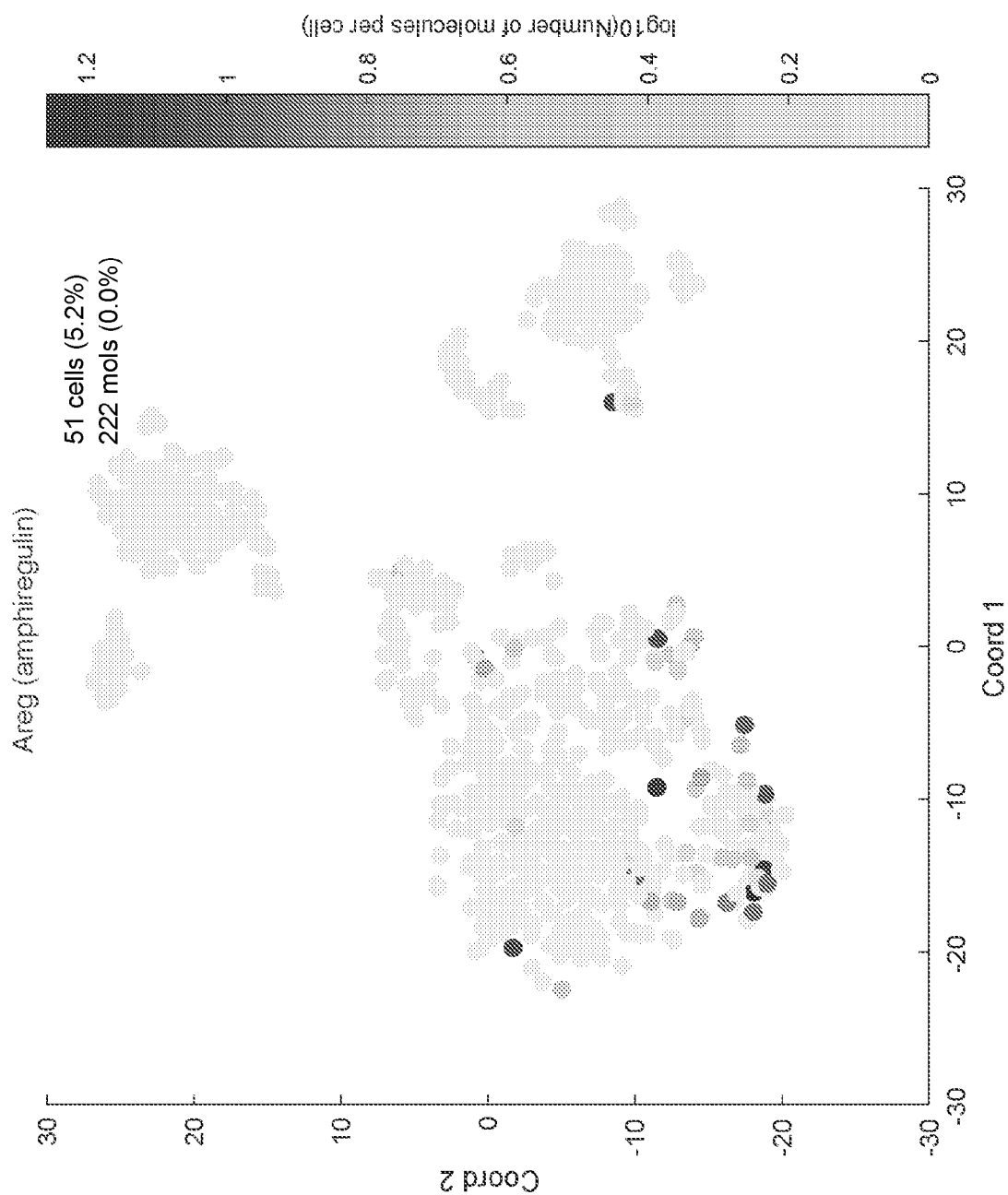
**FIG. 53A**





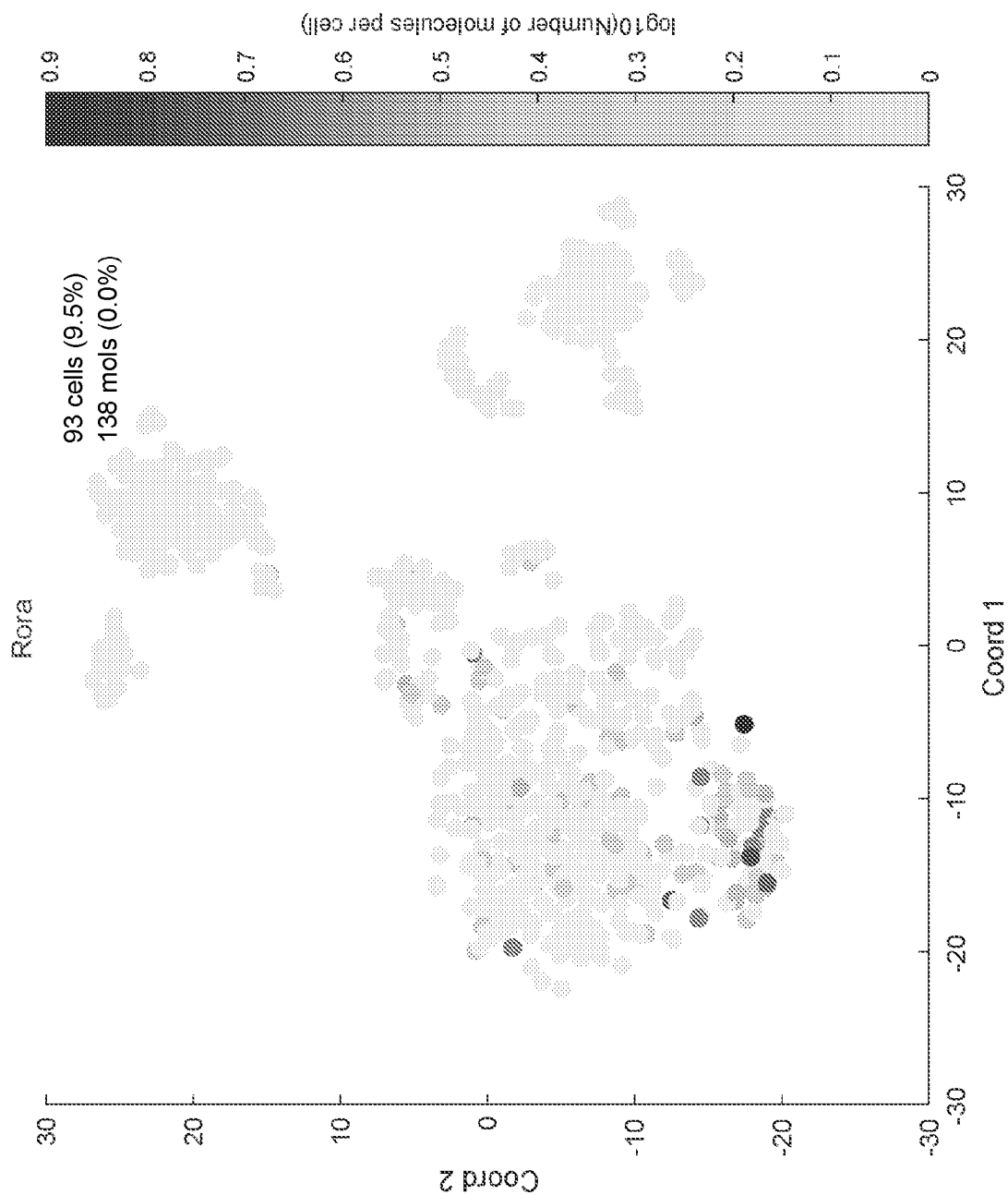
**FIG. 53B**

110/112



**FIG. 53C**

111/112



**FIG. 53D**

112/112



**FIG. 54**

Sequence\_Listing\_BDCRI\_033W0.TXT  
SEQUENCE LISTING

<110> Cellular Research, Inc.  
Chang, Christina  
Fan, Christina  
Shum, Eleen  
Martin, Jody  
Bansal, Nidhanjali  
Ghadiali, James  
Lazaruk, Katherine  
Lam, Gretchen

<120> SAMPLE INDEXING FOR SINGLE CELLS

<130> BDCRI.033W0

<150> 62/515,285  
<151> 2017-06-05

<150> 62/532,905  
<151> 2017-07-14

<150> 62/532,949  
<151> 2017-07-14

<150> 62/532,971  
<151> 2017-07-14

<150> 62/554,425  
<151> 2017-09-05

<150> 62/578,957  
<151> 2017-10-30

<150> 62/645,703  
<151> 2018-03-20

<160> 8

<170> PatentIn version 3.5

<210> 1  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> note="Description of Artificial Sequence: Synthetic  
Oligonucleotide"

<400> 1  
aaaaaaaaaa aaaaaaaaaa

# Sequence\_Listing\_BDCRI\_033W0.TXT

<210> 2  
 <211> 20  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> note="Description of Artificial Sequence: Synthetic  
 Oligonucleotide"

<400> 2  
 tttttttttt tttttttttt 20

<210> 3  
 <211> 95  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> note="Description of Artificial Sequence: Synthetic  
 Oligonucleotide"

<220>  
 <221> 5AmMC6  
 <222> (1)..(1)  
 <223> 5' Amino Modifier C6

<400> 3  
 gttgtcaaga tgctaccgtt cagagtacgt ggagttgggtg gcccgacccc gagcgctacg 60  
 agccccccgg aaaaaaaaaa aaaaaaaaaa aaaaaa 95

<210> 4  
 <211> 200  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> note="Description of Artificial Sequence: Synthetic  
 Oligonucleotide"

<220>  
 <221> 5AmMC6  
 <222> (1)..(1)  
 <223> 5' Amino Modifier C6

<400> 4

Sequence\_Listing\_BDCRI\_033W0.TXT

|  |     |
|--|-----|
| gttgtcaaga tgctaccggt cagagctact gtccgaagtt accgtgtatc taccacgggt  | 60  |
| ggtttttcga atccggaaaa gatagtaata agtgtttttag ttggaataag tcgcaacttt | 120 |
| tggagacgggt tacctctcaa tttttctgat ccgtaggccc cccgatctcg gcctcaaaaa | 180 |
| aaaaaaaaaa aaaaaaaaaa  | 200 |

<210> 5  
 <211> 25  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> note="Description of Artificial Sequence: Synthetic  
 Oligonucleotide"

|                             |    |
|-----------------------------|----|
| <400> 5                     |    |
| gttgtcaaga tgctaccggt cagag | 25 |

<210> 6  
 <211> 95  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> note="Description of Artificial Sequence: Synthetic  
 Oligonucleotide"

|  |    |
|--|----|
| <400> 6  |    |
| gttgtcaaga tgctaccggt cagagcccca tgtctagtag ctattgggtcc cctatcctca | 60 |
| gattcgttta aaaaaaaaaa aaaaaaaaaa aaaaa                             | 95 |

<210> 7  
 <211> 26  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> note="Description of Artificial Sequence: Synthetic  
 Oligonucleotide"

|                              |    |
|------------------------------|----|
| <400> 7                      |    |
| tttttttttt tttttttttt tttttt | 26 |

<210> 8  
 <211> 20

Sequence\_Listing\_BDCRI\_033W0.TXT

<212> DNA

<213> Artificial Sequence

<220>

<223> note="Description of Artificial Sequence: Synthetic  
Oligonucleotide"

<400> 8

acacgacgct cttccgatct

20