(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 August 2023 (10.08.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/150763 A1

(51) International Patent Classification:
C12Q 1/6806 (2018.01)

(21) International Application Number:
PCT/US2023/062070

(22) International Filing Date:
06 February 2023 (06.02.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/307,559 07 February 2022 (07.02.2022) US

(71) Applicant: **BECTON, DICKINSON AND COMPANY** [US/US]; 1 Becton Drive, Franklin Lakes, New Jersey 07417-1880 (US).

(72) Inventors: **SONG, Hye-Won**; 2350 Qume Drive, San Jose, California 95131 (US). **PROSEN, Dennis Edward**; 2350 Qume Drive, San Jose, California 95131 (US).

(74) Agent: **LIU, Jing**; Sheppard, Mullin, Richter & Hampton LLP, 12275 El Camino Real, Suite 100, San Diego, California 92130 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, 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, ME, 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).

(54) Title: PREPARING NUCLEIC ACID FOR FURTHER ANALYSIS OF THEIR SEQUENCE

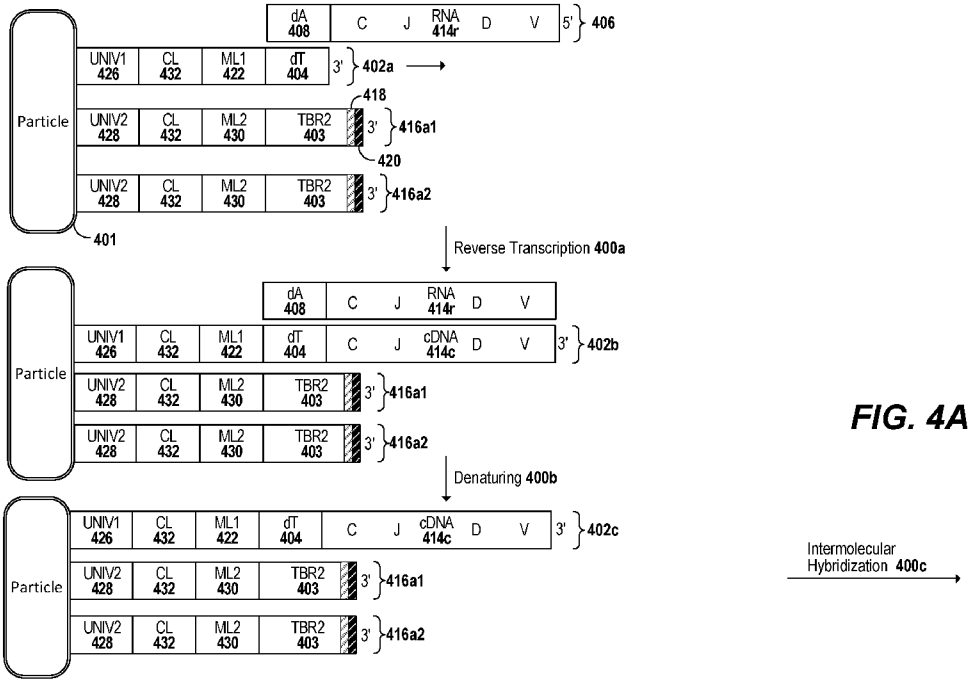


FIG. 4A

(57) Abstract: Disclosed herein include systems, methods, compositions, and kits for full-length whole transcriptome analysis (WTA). Some embodiments comprise 5'-based, 3'-based, and internal-based gene expression profiling. Immune repertoire profiling methods are also provided in some embodiments.

[Continued on next page]

WO 2023/150763 A1

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

PREPARING NUCLEIC ACID FOR FURTHER ANALYSIS OF THEIR SEQUENCE

RELATED APPLICATIONS

[0001] This application claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Patent Application Ser. No. 63/307,559, filed February 7, 2022, the content of this related application is incorporated herein by reference in its entirety for all purposes.

BACKGROUND

Field

[0002] The present disclosure relates generally to the field of molecular biology, and for particular to multiomics analyses using molecular barcoding.

Description of the Related Art

[0003] Methods and techniques of molecular barcoding are useful for single cell transcriptomics analysis, including deciphering gene expression profiles to determine the states of cells using, for example, reverse transcription, polymerase chain reaction (PCR) amplification, and next generation sequencing (NGS). Molecular barcoding is also useful for single cell proteomics analysis. There is a need for methods and techniques for molecular barcoding of nucleic acid targets on one or both the 5' ends and the 3' ends. There is a need for compositions, systems and methods that can quantitatively analyze gene expression of cells efficiently. There is a need for compositions, systems and methods that can obtain the sequence of internal regions of nucleic acid targets. There is a need for compositions, systems and methods that can obtain the full length sequence of nucleic acid targets.

SUMMARY

[0004] Disclosed herein include methods for labeling nucleic acid targets in a sample. In some embodiments, the method comprises: contacting copies of a nucleic acid target with a first plurality of oligonucleotide barcodes, wherein each oligonucleotide barcode of the first plurality of oligonucleotide barcodes comprises a first universal sequence, a first molecular label, and a first target-binding region capable of hybridizing to the nucleic acid target. In some embodiments, the method comprises: extending the first plurality of oligonucleotide barcodes hybridized to the copies of the nucleic acid target to generate a plurality of barcoded nucleic acid molecules each comprising the first universal sequence, the first molecular label, and a sequence complementary to at least a portion of the nucleic acid target. In some embodiments, the method comprises: contacting the barcoded nucleic acid molecules with a second plurality of oligonucleotide barcodes for hybridization. In some embodiments, each oligonucleotide barcode

of the second plurality of oligonucleotide barcodes comprises a second universal sequence, a cleavage domain, and a blocking group. In some embodiments, the blocking group is capable of preventing extension of the oligonucleotide barcode, wherein the cleavage domain is positioned 5' of the blocking group, and wherein a cleaving enzyme is capable of cleaving the oligonucleotide barcode at a point within or adjacent to the cleavage domain when the cleavage domain is hybridized to a barcoded nucleic acid molecule. In some embodiments, the method comprises: contacting the second plurality of oligonucleotide barcodes hybridized to the barcoded nucleic acid molecules with the cleaving enzyme, thereby removing the blocking group from said oligonucleotide barcodes. In some embodiments, the method comprises: extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes hybridized to the barcoded nucleic acid molecules to generate a plurality of extended barcoded nucleic acid molecules.

[0005] In some embodiments, each extended barcoded nucleic acid molecule of the plurality of extended barcoded nucleic acid molecules comprise a sequence of at least a portion of the nucleic acid target. In some embodiments, the cleaving enzyme is an RNase H enzyme and/or an RNase H2 enzyme, optionally the RNase H2 enzyme is a Pyrococcus abyssi RNase H2 enzyme. In some embodiments, the cleaving enzyme is a hot start cleaving enzyme which is thermostable and has reduced activity at lower temperatures. In some embodiments, the hot start cleaving enzyme is Pyrococcus abyssi RNase H2 comprising (a) a G12A amino acid substitution; (b) a P13T amino acid substitution; (c) a G169A amino acid substitution; or (d) a combination thereof. In some embodiments, the cleaving enzyme is chemically modified. In some embodiments, the cleaving enzyme is a chemically modified hot start cleaving enzyme which is thermostable and has reduced activity at lower temperatures, optionally the cleaving enzyme is reversibly inactivated through interaction with an antibody at lower temperatures. In some embodiments, the cleavage domain comprises one or more ribonucleotides that are capable of being cleaved by a RNase H enzyme. In some embodiments, the cleavage domain comprises one or more of the following moieties: a DNA residue, an abasic residue, a modified nucleoside, or a modified phosphate internucleotide linkage. In some embodiments, the cleavage domain comprises at least one RNA base. In some embodiments, the cleavage domain comprises one or more 2'-modified nucleosides, optionally the one or more modified nucleosides are 2'-fluoronucleosides. In some embodiments, the blocking group is attached to the 3' -terminal nucleotide of the oligonucleotide barcode. In some embodiments, the blocking group is at or near 3' terminal of the oligonucleotide barcode. In some embodiments, the blocking group is a 2', 3'-dideoxynucleotide, a ribonucleotide residue, a 2', 3' SH nucleotide, or a 2'-O-PO_3 nucleotide. In some embodiments, the blocking group comprises a

non-nucleotide modification. In some embodiments, the blocking group further comprises a napthyl-azo compound, a spacer and/or a biotin.

[0006] In some embodiments, extending the extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that has strand displacement activity. In some embodiments, extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that has strand displacement activity is capable of generating extended barcoded nucleic acid molecules comprising a complement of the first molecular label and a complement of the first universal sequence. In some embodiments, extending the extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that does not have strand displacement activity. In some embodiments, the polymerase is selected from the group comprising Phi29 DNA polymerase, E. coli DNA polymerase I, Bsu DNA polymerase, Bst DNA polymerase, Taq DNA polymerase, VENT™ DNA polymerase, DEEPVENT™ DNA polymerase, LongAmp® Taq DNA polymerase, LongAmp® Hot Start Taq DNA polymerase, Crimson LongAmp® Taq DNA polymerase, Crimson Taq DNA polymerase, OneTaq® DNA polymerase, OneTaq® Quick-Load® DNA polymerase, Hemo KlenTaq® DNA polymerase, REDTaq® DNA polymerase, Phusion® DNA polymerase, Phusion® High-Fidelity DNA polymerase, Platinum Pfx DNA polymerase, AccuPrime Pfx DNA polymerase, Klenow fragment, Pwo DNA polymerase, Pfu DNA polymerase, T4 DNA polymerase, T7 DNA polymerase, derivatives thereof, or any combination thereof. In some embodiments, extending the 3' ends of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes using a mesophilic DNA polymerase, a thermophilic DNA polymerase, a psychrophilic DNA polymerase, or any combination thereof. In some embodiments, extending the 3' ends of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes using a DNA polymerase lacking at least one of 5' to 3' exonuclease activity and 3' to 5' exonuclease activity, and optionally the DNA polymerase comprises a Klenow Fragment. In some embodiments, extending the first plurality of oligonucleotide barcodes comprises extending the first plurality of oligonucleotide barcodes using a reverse transcriptase. In some embodiments, the reverse transcriptase is capable of terminal transferase activity. In some embodiments, the reverse transcriptase with strand displacement activity is a PrimeScript reverse transcriptase, M-MuLV reverse transcriptase, SmartScribe reverse transcriptase, Maxima H Minus Reverse Transcriptase, and/or Superscript II reverse transcriptase. In some embodiments, the reverse transcriptase comprises a

viral reverse transcriptase, optionally the viral reverse transcriptase is a murine leukemia virus (MLV) reverse transcriptase or a Moloney murine leukemia virus (MMLV) reverse transcriptase.

[0007] In some embodiments, the each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second molecular label, wherein at least 10 of the second plurality of oligonucleotide barcodes comprise different second molecular label sequences, optionally each second molecular label comprises at least 6 nucleotides, further optionally the second molecular label sequence is a random sequence. In some embodiments, the second plurality of oligonucleotide barcodes are hybridized to the barcoded nucleic acid molecules via hybridization between the second molecular label and the sequence complementary to at least a portion of the nucleic acid target. In some embodiments, the each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second target-binding region. In some embodiments, the first target-binding region and/or the second target-binding region comprises a poly(dA) region, a poly(dT) region, a random sequence, a gene-specific sequence, or any combination thereof. In some embodiments, oligonucleotide barcode of the second plurality of oligonucleotide barcodes are hybridized to the barcoded nucleic acid molecules via hybridization between the second target-binding region and the sequence complementary to at least a portion of the nucleic acid target. In some embodiments, at least 10 of the second plurality of oligonucleotide barcodes comprise different second target-binding regions, optionally at least two of the target-binding regions are capable of binding the complements of different nucleic acid targets, further optionally at least two of the target-binding regions are capable of hybridizing different regions of the complement of the same nucleic acid target. In some embodiments, two or more oligonucleotide barcodes of the second plurality of oligonucleotide barcodes are capable of hybridizing to different regions of the complement of the same nucleic acid target to generate two or more extended barcoded nucleic acid molecules. In some embodiments, said two or more extended barcoded nucleic acid molecules are capable of being generated by two or more oligonucleotide barcodes of the second plurality of oligonucleotide barcodes hybridizing to different regions of the complement of the same nucleic acid target. In some embodiments, said the two or more extended barcoded nucleic acid molecules collectively comprise the sequence of the entire at least about 50% of the nucleic acid target.

[0008] In some embodiments, the method comprises: denaturing the plurality of barcoded nucleic acid molecules. In some embodiments, the method comprises: denaturing the plurality of extended barcoded nucleic acid molecules. In some embodiments, the method comprises: determining the copy number of the nucleic acid target in the sample based on: the

number of first molecular labels with distinct sequences associated with the plurality of barcoded nucleic acid molecules, or products thereof. In some embodiments, the method comprises: determining the copy number of the nucleic acid target in the sample based on: the number of first molecular labels with distinct sequences, second molecular labels with distinct sequences, or a combination thereof, associated with the plurality of extended barcoded nucleic acid molecules, or products thereof. In some embodiments, determining the copy number of the nucleic acid target comprises determining the copy number of each of a plurality of nucleic acid targets in the sample based on: the number of first molecular labels with distinct sequences associated with barcoded nucleic acid molecules of the plurality of barcoded nucleic acid molecules, or products thereof, comprising a sequence of the each of the plurality of nucleic acid targets; and/or the number of first molecular labels with distinct sequences, second molecular labels with distinct sequences, or a combination thereof, associated with extended barcoded nucleic acid molecules of the plurality of extended barcoded nucleic acid molecules comprising a sequence of the each of the plurality of nucleic acid targets. In some embodiments, the sequence of the each of the plurality of nucleic acid targets comprises a subsequence of the each of the plurality of nucleic acid targets. In some embodiments, the sequence of the nucleic acid target in the plurality of barcoded nucleic acid molecules comprises a subsequence of the nucleic acid target. In some embodiments, the nucleic acid target comprises mRNA. In some embodiments, the sample comprises a single cell, optionally an immune cell, and further optionally a B cell or a T cell. In some embodiments, the sample comprises a plurality of cells, a plurality of single cells, a tissue, a tumor sample, or any combination thereof. In some embodiments, single cell comprises a circulating tumor cell.

[0009] In some embodiments, the first universal sequence of each oligonucleotide barcode of the first plurality of oligonucleotide barcodes is 5' of the first molecular label and the first target-binding region; and/or the second universal sequence of each oligonucleotide barcode of the second plurality of oligonucleotide barcodes is 5' of the second molecular label and/or the second target-binding region.

[0010] In some embodiments, the method comprises: amplifying the plurality of barcoded nucleic acid molecules using an amplification primer and a primer comprising the first universal sequence, or a portion thereof, thereby generating a first plurality of single-labeled nucleic acid molecules comprising the sequence of the nucleic acid target, or a portion thereof, wherein determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of first molecular labels with distinct sequences associated with the first plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, the method comprises:

amplifying the plurality of extended barcoded nucleic acid molecules using an amplification primer and a primer comprising the second universal sequence, or a portion thereof, thereby generating a second plurality of single-labeled nucleic acid molecules comprising the sequence of the nucleic acid target, or a portion thereof, wherein determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of second molecular labels with distinct sequences associated with the second plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, the amplification primer comprises a fourth universal sequence. In some embodiments, the amplification primer is a target-specific primer. In some embodiments, the target-specific primer specifically hybridizes to an immune receptor, a constant region of an immune receptor, a variable region of an immune receptor, a diversity region of an immune receptor, and/or the junction of a variable region and diversity region of an immune receptor. In some embodiments, the immune receptor is a T cell receptor (TCR) and/or a B cell receptor (BCR) receptor, and optionally the TCR comprises TCR alpha chain, TCR beta chain, TCR gamma chain, TCR delta chain, or any combination thereof; and the BCR receptor comprises BCR heavy chain and/or BCR light chain.

[0011] In some embodiments, the method comprises: hybridizing random primers to the plurality of barcoded nucleic acid molecules and extending the random primers to generate a first plurality of extension products, wherein the random primers comprise a third universal sequence, or a complement thereof; and amplifying the first plurality of extension products using a primer capable of hybridizing to the third universal sequence, or a complement thereof, and a primer capable of hybridizing to the first universal sequence, or a complement thereof, thereby generating a third plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of first molecular labels with distinct sequences associated with the third plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, the method comprises: hybridizing random primers to the plurality of extended barcoded nucleic acid molecules and extending the random primers to generate a second plurality of extension products, wherein the random primers comprise a third universal sequence, or a complement thereof; and amplifying the second plurality of extension products using a primer capable of hybridizing to the third universal sequence, or a complement thereof, and a primer capable of hybridizing to the second universal sequence, or a complement thereof, thereby generating a fourth plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, determining the copy number of the nucleic acid target in the sample comprises:

determining the copy number of the nucleic acid target in the sample based on the number of second molecular labels with distinct sequences associated with the fourth plurality of single-labeled nucleic acid molecules, or products thereof.

[0012] In some embodiments, the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence are the same. In some embodiments, the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence are different. In some embodiments, the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence comprise the binding sites of sequencing primers and/or sequencing adaptors, complementary sequences thereof, and/or portions thereof. In some embodiments, the sequencing adaptors comprise a P5 sequence, a P7 sequence, complementary sequences thereof, and/or portions thereof. In some embodiments, the sequencing primers comprise a Read 1 sequencing primer, a Read 2 sequencing primer, complementary sequences thereof, and/or portions thereof.

[0013] In some embodiments, the method comprises: obtaining sequence information of the plurality of barcoded nucleic acid molecules, or products thereof. In some embodiments, obtaining the sequence information comprises attaching sequencing adaptors to the plurality of extended barcoded nucleic acid molecules, or products thereof. In some embodiments, the method comprises: obtaining sequence information of the plurality of extended barcoded nucleic acid molecules, or products thereof. In some embodiments, obtaining the sequence information comprises attaching sequencing adaptors to the extended barcoded nucleic acid molecules, barcoded nucleic acid molecules, products thereof, or any combination thereof. In some embodiments, the method comprises: obtaining sequence information of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof, In some embodiments, obtaining the sequence information comprises attaching sequencing adaptors to one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof. In some embodiments, obtaining sequence information of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof, comprises: obtaining sequencing data comprising a plurality of sequencing reads of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof, wherein each of the plurality of sequencing reads comprise (1) a cell label sequence, (2) a molecular label sequence, and/or (3) a subsequence of the nucleic acid target.

[0014] In some embodiments, the method comprises: for each unique cell label sequence, which indicates a single cell of the sample: aligning each of the plurality of

sequencing reads of the nucleic acid target to generate an aligned sequence of the nucleic acid target. In some embodiments, the aligned sequence of the nucleic acid target comprises at least 50% of the cDNA sequence of the nucleic acid target, at least 70% of the cDNA sequence of the nucleic acid target, at least 90% of the cDNA sequence of the nucleic acid target, or the full length of the cDNA sequence of the nucleic acid target. In some embodiments, the nucleic acid target is an immune receptor, optionally the immune receptor comprises BCR light chain, BCR heavy chain, TCR alpha chain, TCR beta chain, TCR gamma chain, TCR delta chain, or any combination thereof. In some embodiments, the aligned sequence of the nucleic acid target comprises the complementarity determining region 1 (CDR1), the complementarity determining region 2 (CDR2), the complementarity determining region 3 (CDR3), the variable region, the full length of the variable region, or a combination thereof. In some embodiments, the aligned sequence of the nucleic acid target comprises the variable region, the diversity region, the junction of a variable region diversity region and/or the constant region, or any combination thereof. In some embodiments, obtaining the sequence information comprises obtaining the sequence information of the BCR light chain and the BCR heavy chain of a single cell, and optionally the sequence information of the BCR light chain and the BCR heavy chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the CDR3, or any combination thereof, of the BCR light chain and/or the BCR heavy chain. In some embodiments, the method comprises: pairing the BCR light chain and the BCR heavy chain of the single cell based on the obtained sequence information. In some embodiments, the sample comprises a plurality of single cells, the method comprising pairing the BCR light chain and the BCR heavy chain of at least 50% of said single cells based on the obtained sequence information. In some embodiments, obtaining the sequence information comprises obtaining the sequence information of the TCR alpha chain and the TCR beta chain of a single cell, and optionally the sequence information of the TCR alpha chain and the TCR beta chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the CDR3, or any combination thereof, of the TCR alpha chain and/or the TCR beta chain. In some embodiments, the method comprises: pairing the TCR alpha chain and the TCR beta chain of the single cell based on the obtained sequence information. In some embodiments, the sample comprises a plurality of single cells, the method comprising pairing the TCR alpha chain and the TCR beta chain of at least 50% of said single cells based on the obtained sequence information. In some embodiments, obtaining the sequence information comprises obtaining the sequence information of the TCR gamma chain and the TCR delta chain of a single cell. In some embodiments, the sequence information of the TCR gamma chain and the TCR delta chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the

CDR3, or any combination thereof, of the TCR gamma chain and/or the TCR delta chain. In some embodiments, the method comprises: pairing the TCR gamma chain and the TCR delta chain of the single cell based on the obtained sequence information. In some embodiments, the sample comprises a plurality of single cells, the method comprising pairing the TCR gamma chain and the TCR delta chain of at least 50% of said single cells based on the obtained sequence information.

[0015] In some embodiments, the complement of the molecular label comprises a reverse complementary sequence of the molecular label or a complementary sequence of the molecular label. In some embodiments, the plurality of barcoded nucleic acid molecules comprises barcoded deoxyribonucleic acid (DNA) molecules, barcoded ribonucleic acid (RNA) molecules, or a combination thereof. In some embodiments, the nucleic acid target comprises a nucleic acid molecule, optionally the nucleic acid molecule comprises ribonucleic acid (RNA), messenger RNA (mRNA), microRNA, small interfering RNA (siRNA), RNA degradation product, RNA comprising a poly(A) tail, or any combination thereof, and further optionally the mRNA encodes an immune receptor. In some embodiments, the nucleic acid target comprises a cellular component binding reagent, and/or the nucleic acid molecule is associated with the cellular component binding reagent, optionally the method further comprising dissociating the nucleic acid molecule and the cellular component binding reagent. In some embodiments, at least 10 of the first and/or second pluralities of oligonucleotide barcodes comprise different molecular label sequences. In some embodiments, each molecular label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides. In some embodiments, the first and/or second pluralities of oligonucleotide barcodes are associated with a solid support. In some embodiments, the first and/or second pluralities of oligonucleotide barcodes associated with the same solid support each comprise an identical sample label. In some embodiments, each sample label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides. In some embodiments, the first and/or second pluralities of oligonucleotide barcodes each comprise a cell label. In some embodiments, each cell label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides. In some embodiments, oligonucleotide barcodes of the first and/or second pluralities of oligonucleotide barcodes associated with the same solid support comprise the same cell label. In some embodiments, oligonucleotide barcodes of the first and/or second pluralities of oligonucleotide barcodes associated with different solid supports comprise different cell labels. In some embodiments, the method comprises: extending the oligonucleotide barcodes in the presence of one or more of ethylene glycol, polyethylene glycol, 1,2-propanediol, dimethyl

sulfoxide (DMSO), glycerol, formamide, 7-deaza-GTP, acetamide, tetramethylammonium chloride salt, betaine, or any combination thereof.

[0016] In some embodiments, the solid support comprises a synthetic particle, a planar surface, or a combination thereof. In some embodiments, the sample comprises a single cell, the method comprising associating a synthetic particle comprising the first and second pluralities of oligonucleotide barcodes with the single cell in the sample. In some embodiments, the method comprises: lysing the single cell after associating the synthetic particle with the single cell, optionally lysing the single cell comprises heating the sample, contacting the sample with a detergent, changing the pH of the sample, or any combination thereof. In some embodiments, the synthetic particle and the single cell are in the same partition, and optionally the partition is a well or a droplet. In some embodiments, at least one oligonucleotide barcode of the first and/or second pluralities of oligonucleotide barcodes is immobilized or partially immobilized on the synthetic particle, or at least one oligonucleotide barcode of the first and/or second pluralities of oligonucleotide barcodes is enclosed or partially enclosed in the synthetic particle. In some embodiments, the synthetic particle is disruptable, optionally a disruptable hydrogel particle. In some embodiments, the synthetic particle comprises a bead, optionally the bead is 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. In some embodiments, the synthetic particle comprises 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. In some embodiments, each oligonucleotide barcode of the first and/or second pluralities of oligonucleotide barcodes comprises a linker functional group. In some embodiments, the synthetic particle comprises a solid support functional group. In some embodiments, the support functional group and the linker functional group are associated with each other, and optionally the linker functional group and the support functional group are individually selected from the group consisting of C6, biotin, streptavidin, primary amine(s), aldehyde(s), ketone(s), and any combination thereof.

[0017] There are provided, in some embodiments, solid supports. In some embodiments, the solid support associated with one or both of a first and second pluralities of oligonucleotide barcodes disclosed herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] FIG. 1 illustrates a non-limiting exemplary barcode.

[0019] FIG. 2 shows a non-limiting exemplary workflow of barcoding and digital counting.

[0020] FIG. 3 is a schematic illustration showing a non-limiting exemplary process for generating an indexed library of targets barcoded at the 3'-ends from a plurality of targets.

[0021] FIGS. 4A-4H show schematic illustrations of non-limiting exemplary workflows of determining the sequences of a nucleic acid target (e.g., the V(D)J region of an immune receptor) and for full-length whole transcriptome analysis (WTA) using barcoding methods and compositions provided herein.

[0022] FIGS. 5A-5E show schematic illustrations of non-limiting exemplary workflows of determining the sequences of a nucleic acid target (e.g., the V(D)J region of an immune receptor) and for full-length whole transcriptome analysis (WTA) using barcoding methods and compositions provided herein.

DETAILED DESCRIPTION

[0023] 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.

[0024] 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.

[0025] 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™ assay (Cellular Research, Inc. (Palo Alto, CA)) and Rhapsody™ assay (Becton, Dickinson and Company (Franklin Lakes, NJ)), can correct for bias induced by PCR and library preparation steps by using molecular labels (MLs) to label mRNAs during reverse transcription (RT).

[0026] The Precise™ assay can utilize a non-depleting pool of stochastic barcodes with large number, for example 6561 to 65536, unique molecular label sequences 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 label sequences, and the numbers of mRNA molecules.

[0027] Disclosed herein include methods for labeling nucleic acid targets in a sample. In some embodiments, the method comprises: contacting copies of a nucleic acid target with a first plurality of oligonucleotide barcodes, wherein each oligonucleotide barcode of the first plurality of oligonucleotide barcodes comprises a first universal sequence, a first molecular label, and a first target-binding region capable of hybridizing to the nucleic acid target. In some embodiments, the method comprises: extending the first plurality of oligonucleotide barcodes hybridized to the copies of the nucleic acid target to generate a plurality of barcoded nucleic acid molecules each comprising the first universal sequence, the first molecular label, and a sequence complementary to at least a portion of the nucleic acid target. In some embodiments, the method comprises: contacting the barcoded nucleic acid molecules with a second plurality of oligonucleotide barcodes for hybridization. In some embodiments, each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second universal sequence, a cleavage domain, and a blocking group. In some embodiments, the blocking group is capable of preventing extension of the oligonucleotide barcode, wherein the cleavage domain is positioned 5' of the blocking group, and wherein a cleaving enzyme is capable of cleaving the oligonucleotide barcode at a point within or adjacent to the cleavage domain when the cleavage domain is hybridized to a barcoded nucleic acid molecule. In some embodiments, the method

comprises: contacting the second plurality of oligonucleotide barcodes hybridized to the barcoded nucleic acid molecules with the cleaving enzyme, thereby removing the blocking group from said oligonucleotide barcodes. In some embodiments, the method comprises: extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes hybridized to the barcoded nucleic acid molecules to generate a plurality of extended barcoded nucleic acid molecules. There are provided, in some embodiments, solid supports. In some embodiments, the solid support associated with one or both of a first and second pluralities of oligonucleotide barcodes disclosed herein.

Definitions

[0028] 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.

[0029] 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 adaptors can be linear. The adaptors can be pre-adenylated adaptors. 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 adaptor 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' adaptors can comprise identical and/or universal nucleic acid sequences and the 3' adaptors 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 adaptors (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.

[0030] 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).

[0031] 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, a "complementary" sequence can refer to a "complement" or a "reverse complement" of a sequence. It is understood from the disclosure that if a molecule can hybridize to another molecule it may be complementary, or partially complementary, to the molecule that is hybridizing.

[0032] 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.

[0033] 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.

[0034] 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).

[0035] 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.

[0036] 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.

[0037] 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.

[0038] 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_2 component parts.

[0039] 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.

[0040] 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.

[0041] 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_2), 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 (Tm=+3 to +10 °C), stability towards 3'-exonucleolytic degradation and good solubility properties.

[0042] 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—CH3) 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).

[0043] 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.

[0044] 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.

[0045] 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."

[0046] 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.

[0047] 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.

[0048] 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."

[0049] 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."

[0050] 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).

[0051] 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

[0052] Barcoding, such as stochastic barcoding, has been described in, for example, US 2015/0299784, WO 2015/031691, 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.

[0053] 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.

[0054] 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.

[0055] 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).

[0056] 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

[0057] 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

[0058] 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.

[0059] 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.

[0060] 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.

[0061] 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

[0062] 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.

[0063] 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.

[0064] 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

[0065] 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.

[0066] 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

[0067] 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).

[0068] 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). In some embodiments, the unique molecular label sequence is partially or entirely encompassed by a particle (e.g., a hydrogel bead).

[0069] 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

[0070] 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).

[0071] 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).

[0072] For barcoding (e.g., 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.

[0073] 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

[0074] 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.

[0075] 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, quatramer, 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.

[0076] 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

[0077] 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

[0078] 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.

[0079] 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.

[0080] The antibody fragment can be, for example, a portion of an antibody such as F(ab')2, 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

[0081] 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

[0082] 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

[0083] 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.

[0084] 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.

[0085] 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.

[0086] In some embodiments, the particle can be disruptable (e.g., dissolvable, 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.

[0087] 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.

[0088] 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(phthaladehyde) (PPA), poly(hexyl viologen) (PHV), poly(L-lysine) (PLL), poly(L-arginine) (PARG), poly(lactic-co-glycolic acid) (PLGA).

[0089] 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.

[0090] 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.

[0091] 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.

[0092] 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 can cause melting of a bead such that the bead wall disintegrates. In some embodiments, the heat can increase the internal pressure of the inner components of the bead such that the bead ruptures or explodes. In some embodiments, the heat can 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.

[0093] 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 Fe_3O_4 nanoparticles into polyelectrolyte containing beads triggers rupture in the presence of an oscillating magnetic field stimulus.

[0094] 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.

[0095] 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_2 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.

[0096] 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 barcodes. 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.

[0097] 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.

[0098] 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.

[0099] 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.

[0100] 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.

[0101] 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 (Fe_3O_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.

[0102] 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.

[0103] 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, be about, 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, 10^9, a number or a range between any two of these values, or more. 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.

[0104] 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, be about, 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, 10^9, or a number or a range between any two of these values. 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.

[0105] The size of the beads can vary. For example, the diameter of the bead can range from 0.1 micrometer to 50 micrometers. 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, or 50 micrometers, or a number or a range between any two of these values.

[0106] 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, be about, be at least, or be at

most, 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). 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, be about, be at least, or be at most, 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.

[0107] 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.

[0108] 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.

[0109] 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.

[0110] 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.

[0111] 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 lacks 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.

[0112] 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.

[0113] 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 able to be pumped throughout the circulatory system.

Substrates and Microwell Array

[0114] 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

[0115] 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.

[0116] 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

[0117] 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.

[0118] 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

[0119] 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.

[0120] 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.

[0121] 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 or 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 or 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 0.5 M. 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, 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 or at most about 1, 5, 10, 15, 20, 25, or 30 mM. 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 or at most about 1, 5, 10, 15, or 20 mM. 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.

[0122] 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

[0123] 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.

[0124] 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.

[0125] 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.

[0126] 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

[0127] 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.

[0128] 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.

[0129] 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

[0130] 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 molecular label).

[0131] 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.

[0132] 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 Qβ replicase (Qβ) 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.

[0133] 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 labeled amplicon). 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.

[0134] 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.

[0135] 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.

[0136] 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.

[0137] 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.

[0138] 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.

[0139] 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.

[0140] 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.

[0141] 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.

[0142] 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.

[0143] 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 Qβ replicase (Qβ), 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).

[0144] 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.

[0145] In some embodiments, the method comprises repeatedly amplifying the labeled nucleic acid to produce multiple amplicons. The methods disclosed herein can comprise conducting about, or at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, or a number or a range between any two of these values, amplification reactions.

[0146] 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.

[0147] 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.

[0148] 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.

[0149] 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.

[0150] 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, 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.

[0151] 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 sequence, a cell label sequence, 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.

[0152] In some embodiments, the cell label sequence can include 3 to 20 nucleotides. In some embodiments, the molecular label sequence 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.

[0153] 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.

[0154] 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.

[0155] 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 1st PCR primer pool and a 1st 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 1st 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.

[0156] 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 2nd 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 2nd 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 2nd 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.

[0157] 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 NGS.

Full Length Single Cell RNA Sequencing

[0158] High-throughput single-cell RNA-sequencing has transformed the understanding of complex and heterogenous biological samples. However, most methods enable only 3' analysis of the mRNA transcript information, which may limit analysis of splice variants, alternative transcription start sites and highly variable loci due to rearrangement such as the VDJ junction of T cell and B cell receptors and antibodies. For both T and B cells,

currently available C priming-based approaches can read into V(D)J but misses upstream V region. Thus, currently available methods can limit the ability to get full length nucleic acid target (e.g., V(D)J-containing transcript) information. A particular problem in the art is that we need to know the VDJ sequence as a longer read because there are many VDJ due to the numerous recombination events possible. There is a need for methods of both counting of sequences (e.g., V(D)J-containing transcripts) and identification of said sequences (in particular full-length sequence identification).

[0159] There are provided, in some embodiments, methods to obtain full-length V(D)J information (e.g., by Illumina sequencing on the Rhapsody system). T and B cell receptors contain V segments, D segments (for TCR beta and BCR heavy chain only), J segments as well as a constant region at the 3' prime end of the mRNA. CDR3, which is made of the V(D)J junction, contains the bulk of the repertoire diversity and is short enough to be sequenced on the Illumina short read platform. However, full-length V segment and D segment and J segment information is also useful and cannot be easily obtained without long read sequencing technologies, as Illumina short read capability limits ability to get full-length V(D)J information. The methods provided herein can enable a user to obtain both CDR3 information as well as full-length V segment, full-length D segment sequence and/or full-length J segment sequence from a single library and sequencing run compatible with Illumina sequencers. Thus, some embodiments of the methods provided herein yield full-length immune receptor mRNA sequences.

[0160] Disclosed herein include methods for generating a whole transcriptome analysis (WTA) library from cDNA product (e.g., cDNA product of the BD Rhapsody Single Cell Analysis System) for sequencing on compatible sequencers (e.g., Illumina® sequencers). In some embodiments, targeted RNA analysis generally yields higher sensitivity for low expressing targets, while WTA RNA analysis provides a larger breadth of interrogated genes. Successful amplification of desired targets in some embodiments, may need a thorough understanding of each RNA's 3' end and its use of polyadenylation sites in model system of interest. The WTA method, for example using with BD Rhapsody, allows obtaining of several levels of information, including 1) identify interesting RNA targets in their model system to identify a panel of genes to design; 2) characterize the 3' ends of the transcriptome in their model system as an input to a new generation of PCR panel design for BD Rhapsody, and 3) allows users to do discovery biology by assaying a wider breath of genes compared to the standard BD Rhapsody targeted approach.

[0161] There are provided, in some embodiments, systems, methods, compositions, and kits for 3'-based, internal-based, and/or 5'-based whole transcriptome analysis (WTA).

Disclosed herein include methods and compositions for 5', 3', and internal WTA library generation. The disclosed methods and compositions can be employed for full length single cell RNA sequencing using, for example, the Rhapsody system. Disclosed methods and compositions enable sequencing the full length whole transcriptome analysis (WTA) of mRNAs in single cells using BD Rhapsody System. Currently available WTA assays sequence the 3' end of mRNA by capturing the poly A tail at the 3' end of mRNAs and adding a unique molecular index (UMI) and cell label (CL). Additionally, currently available methods can profile the 5' end of mRNA by adding a sequence to the 5' end of the cDNA that bridges back to the bead, and adding UMI and CL at the 5' end of the transcript. However, sequencers can only profile small fragments (usually ~300-600 base pairs); this means any region greater than ~750 bases from the 3' or 5' end cannot be profiled, since no method is available to add UMI and CL into internal region. Currently available high throughput scRNAseq methods offer sequencing either 3' or 5' end of transcript. Some currently available methods enable either a 3' or 5' solution, but not both at the same time. Full length RNAs are not captured by other methods. Low throughput scRNAseq methods currently available can offer full length sequencing but the maximum cell number is much lower and is highly labor intensive process. Due to the limitation of sequencing long sequences on Illumina devices, new methods and compositions are needed. The methods and compositions provided herein can overcome these disadvantages of currently available methods. There are provided, in some embodiments, full length scRNAseq solutions for high throughput experiments. The systems, methods, compositions, and kits provided herein can, in some embodiments, be employed in concert with the methods and compositions described in PCT Patent Application Number PCT/US22/76366, filed on September, 13, 2022, entitled "FULL LENGTH SINGLE CELL RNA SEQUENCING", the content of which is incorporated herein by reference in its entirety.

[0162] There are provided, in some embodiments, compositions and methods for performing full length RNA sequencing (e.g., using the Rhapsody system). Currently available workflows offer 3' or 5' mRNA sequencing analysis at the single cell level but not the full length. A reason said currently available methods cannot provide internal sequence(s) of RNA is that 1) Illumina sequencer can sequence a limited size of library, 2) to obtain single cell information, cell label and UMI (e.g., molecular label) should be at the end of the library. Current capture sequences for RNA/cDNA are through the poly A (3' end of mRNA) or template switching oligonucleotide (5' end of mRNA). Therefore, these approaches are only providing 3' and 5' end sequence(s). To link internal sequence(s) with the cell label, it is important to bring the oligonucleotide near the internal region. Some embodiments of the methods and compositions provided herein use randomer/internal seq of targeted panel

sequence(s) on the beads to add the cell label information linked to internal mRNA sequence(s). In some embodiments, approaches described herein utilize polymerase(s) that have strand displacement activity to generate multiple copies of internal sequence and/or polymerase(s) without strand displacement activity to generate a few area of internal sequence. These two approaches can be utilized depending on the purpose of full-length sequencing. If a user desires full coverage of cDNA sequence, the first approach can be used. If a user desires to sequence internal area but does not want full coverage due to the high cost of sequencing, the latter method will be suitable. If a user desires internal sequence for all mRNAs, a randomer can be used. If a user desires specific targeted gene set for internal sequence (i.e. mutation hot spots), targeted gene specific sequence can be added. FIGS. 4A-4H show schematic illustrations of non-limiting exemplary workflows of determining the sequences of a nucleic acid target (e.g., the V(D)J region of an immune receptor) and for full-length whole transcriptome analysis (WTA) using barcoding methods and compositions provided herein. The work flow can comprise one or more of the following steps: Reverse Transcription **400a**, Denaturing **400b**, Intermolecular Hybridization **400c**, Contacting with Cleavage Agent **400d**, Contacting with Polymerase without Strand Displacement Activity **400e**, Contacting with Polymerase with Strand Displacement Activity **400f**, Denaturing and Primer Hybridization **400g**, Denaturing and Primer Hybridization **400h**, Primer Extension, Library Generation, and Sequencing **400i**, and Primer Extension, Library Generation, and Sequencing **400j**. A nucleic acid target **406** (e.g., mRNA) can comprise a coding sequence (e.g., **414r**) and a polyA tail (e.g., **408**). The first plurality of oligonucleotide barcodes **402a** can comprise a first universal sequence **426**, a first molecular label **422**, and a sequence complementary to at least a portion of the nucleic acid target (e.g., a first target-binding region **404**). Each oligonucleotide barcode of the second plurality of oligonucleotide barcodes **416** (e.g., **416a1**, **416a2**) can comprises a second universal sequence **428**, a cleavage domain **418**, and a blocking group **420**, a second molecular label **430**, a second target binding region (**403**). The the cleavage domain can be positioned 5' of the blocking group. Oligonucleotide barcodes can comprise a cell label **432**. Oligonucleotide barcodes can be attached to a solid support **401**. The workflow can comprise generation of one or more of the following products: **402b**, **402c**, **414c**, **402c**, **416c**, **416d**, and **416e**. Products can include a reverse complement (**rc**) of the elements described herein. The workflow can comprise contacting products with Random Primer **446** and this can comprise a third universal sequence **448**.

[0163] In some embodiments, solid supports (e.g., Rhapsody beads) provided herein have oligo dT capture sequence and/or template switching oligo capture sequences on the beads to capture through 3' or 5' end of mRNA/cDNA. Adding randomer/gene specific sequence(s)

for capture oligonucleotides in the solid supports (e.g., Rhapsody beads) has not been employed. However, as described herein, having these oligos in the solid supports (e.g., Rhapsody beads) can allow the generation of internal sequence containing library to make full length RNAseq possible. By adding the blocked primer, a user can control the time of usage of the primer and that can reduce noise generation during library preparation.

[0164] FIGS. 5A-5E show schematic illustrations of non-limiting exemplary workflows of determining the sequences of a nucleic acid target (e.g., the V(D)J region of an immune receptor) and for full-length whole transcriptome analysis (WTA) using barcoding methods and compositions provided herein.

[0165] FIG. 5A depicts a non-limiting exemplary solid support (e.g., bead). Full length RNAseq can comprise the depicted bead (e.g., rhapsody bead) and associated oligonucleotides. The workflow can comprise mRNA capture and reverse transcription, followed by mRNA denaturation or use of Rnase H to remove RNA (upper oligonucleotide). A randomer (UMI, molecular label) or gene specific sequence can have a rhAmp end to prevent from this sequence to bind RNAs and reverse transcription (lower oligonucleotide). rhAmp PCR can comprise blocked inactive primers annealing to a target. Said primers can have an overhang and an internal RNA base 5' of a blocking moiety. RNase H2 can recognize hybridized internal RNA base and cleavage can occur on the 5' side of the RNA base after primer hybridization to the target DNA. DNA polymerase can now extend newly unblocked primers.

[0166] FIG. 5B depicts a non-limiting exemplary embodiment of first and second workflows provided herein. In some embodiments, the workflow is a first workflow comprising full length RNAseq without strand displacement activity. The first workflow can comprise hybridization of Randomers on the cDNA. The first workflow can comprise Rnase H2 treatment to release the blocker and polymerase reaction. In some embodiments of the first workflow, due to no strand displacement activity, polymerization will stop at the end of other randomer binding region. This first workflow can generate DNA pieces of different regions of cDNA. This first workflow can be ideal when medium level of full-length coverage is desired by a user. In some embodiments, the workflow is a second workflow comprising full length RNAseq with strand displacement activity. The second workflow can comprise hybridization of Randomers on the cDNA. The second workflow can comprise Rnase H2 treatment to release the blocker and Polymerase reaction. In some embodiments, due to strand displacement activity, polymerization will keep going and generate multiple different size DNA in the second workflow. FIG. 5C depicts a non-limiting exemplary embodiment of second workflow (Full length RNAseq (with strand displacement activity)) provided herein. Due to strand displacement activity in the second

workflow, polymerization can keep going and generate multiple different size DNA. This can be ideal if full coverage of full-length sequence is desired by a user.

[0167] FIG. 5D depicts non-limiting exemplary embodiments of full length RNAseq without PTA. In some embodiments, the polymerization will generate different size products. FIG. 5E depicts non-limiting exemplary embodiments of RPE-PCR (employing a random with a R2 sequence). The UMI from T1 oligos can be used for randomer and still can be used to filter PCR amplicons. In some embodiments, the UMI from T1 oligos can be used to count molecule number. In some embodiments, UMI from T1 oligos is not used to count molecule number (e.g., there will be more than 1 UMI/mRNA transcript). In some embodiments, to have R2 adaptor, RPE is repeated. In some embodiments, randomers from other beads can bind to the cDNA. In some embodiments, instead of employing a Randomer, sequence specific primers can be used to generate full coverage of certain genes of interest (e.g., cancer related mutation detection in gene candidates) at the single cell level.

[0168] Disclosed herein include methods for labeling nucleic acid targets in a sample. In some embodiments, the method comprises: contacting copies of a nucleic acid target with a first plurality of oligonucleotide barcodes, wherein each oligonucleotide barcode of the first plurality of oligonucleotide barcodes comprises a first universal sequence, a first molecular label, and a first target-binding region capable of hybridizing to the nucleic acid target. In some embodiments, the method comprises: extending the first plurality of oligonucleotide barcodes hybridized to the copies of the nucleic acid target to generate a plurality of barcoded nucleic acid molecules each comprising the first universal sequence, the first molecular label, and a sequence complementary to at least a portion of the nucleic acid target. In some embodiments, the method comprises: contacting the barcoded nucleic acid molecules with a second plurality of oligonucleotide barcodes for hybridization. In some embodiments, each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second universal sequence, a cleavage domain, and a blocking group. In some embodiments, the blocking group is capable of preventing extension of the oligonucleotide barcode, wherein the cleavage domain is positioned 5' of the blocking group, and wherein a cleaving enzyme is capable of cleaving the oligonucleotide barcode at a point within or adjacent to the cleavage domain when the cleavage domain is hybridized to a barcoded nucleic acid molecule. In some embodiments, the method comprises: contacting the second plurality of oligonucleotide barcodes hybridized to the barcoded nucleic acid molecules with the cleaving enzyme, thereby removing the blocking group from said oligonucleotide barcodes. In some embodiments, the method comprises: extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide

barcodes hybridized to the barcoded nucleic acid molecules to generate a plurality of extended barcoded nucleic acid molecules.

[0169] In some embodiments, each extended barcoded nucleic acid molecule of the plurality of extended barcoded nucleic acid molecules comprise a sequence of at least a portion of the nucleic acid target. In some embodiments, the cleaving enzyme is an RNase H enzyme and/or an RNase H2 enzyme, optionally the RNase H2 enzyme is a Pyrococcus abyssi RNase H2 enzyme. In some embodiments, the cleaving enzyme is a hot start cleaving enzyme which is thermostable and has reduced activity at lower temperatures. In some embodiments, the hot start cleaving enzyme is Pyrococcus abyssi RNase H2 comprising (a) a G12A amino acid substitution; (b) a P13T amino acid substitution; (c) a G169A amino acid substitution; or (d) a combination thereof. In some embodiments, the cleaving enzyme is chemically modified. In some embodiments, the cleaving enzyme is a chemically modified hot start cleaving enzyme which is thermostable and has reduced activity at lower temperatures, optionally the cleaving enzyme is reversibly inactivated through interaction with an antibody at lower temperatures. In some embodiments, the cleavage domain comprises one or more ribonucleotides that are capable of being cleaved by a RNase H enzyme. In some embodiments, the cleavage domain comprises one or more of the following moieties: a DNA residue, an abasic residue, a modified nucleoside, or a modified phosphate internucleotide linkage. In some embodiments, the cleavage domain comprises at least one RNA base. In some embodiments, the cleavage domain comprises one or more 2'-modified nucleosides, optionally the one or more modified nucleosides are 2'- fluoronucleosides. In some embodiments, the blocking group is attached to the 3' -terminal nucleotide of the oligonucleotide barcode. In some embodiments, the blocking group is at or near 3' terminal of the oligonucleotide barcode. In some embodiments, the blocking group is a 2', 3'-dideoxynucleotide, a ribonucleotide residue, a 2', 3' SH nucleotide, or a 2'-O-PO3 nucleotide. In some embodiments, the blocking group comprises a non-nucleotide modification. In some embodiments, the blocking group further comprises a napthyl-azo compound, a spacer and/or a biotin.

[0170] In some embodiments, extending the extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that has strand displacement activity. In some embodiments, extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that has strand displacement activity is capable of generating extended barcoded nucleic acid molecules comprising a complement of the first molecular label and a complement of the first universal sequence. In some embodiments, extending the extending the 3' ends of

oligonucleotide barcodes of the second plurality of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that does not have strand displacement activity. In some embodiments, the polymerase is selected from the group comprising Phi29 DNA polymerase, E.coli DNA polymerase I, Bsu DNA polymerase, Bst DNA polymerase, Taq DNA polymerase, VENT™ DNA polymerase, DEEPVENT™ DNA polymerase, LongAmp® Taq DNA polymerase, LongAmp® Hot Start Taq DNA polymerase, Crimson LongAmp® Taq DNA polymerase, Crimson Taq DNA polymerase, OneTaq® DNA polymerase, OneTaq® Quick-Load® DNA polymerase, Hemo KlenTaq® DNA polymerase, REDTaq® DNA polymerase, Phusion® DNA polymerase, Phusion® High-Fidelity DNA polymerase, Platinum Pfx DNA polymerase, AccuPrime Pfx DNA polymerase, Klenow fragment, Pwo DNA polymerase, Pfu DNA polymerase, T4 DNA polymerase, T7 DNA polymerase, derivatives thereof, or any combination thereof. In some embodiments, extending the 3' ends of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes using a mesophilic DNA polymerase, a thermophilic DNA polymerase, a psychrophilic DNA polymerase, or any combination thereof. In some embodiments, extending the 3' ends of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes using a DNA polymerase lacking at least one of 5' to 3' exonuclease activity and 3' to 5' exonuclease activity, and optionally the DNA polymerase comprises a Klenow Fragment. In some embodiments, extending the first plurality of oligonucleotide barcodes comprises extending the first plurality of oligonucleotide barcodes using a reverse transcriptase. In some embodiments, the reverse transcriptase is capable of terminal transferase activity. In some embodiments, the reverse transcriptase with strand displacement activity is a PrimeScript reverse transcriptase, M-MuLV reverse transcriptase, SmartScribe reverse transcriptase, Maxima H Minus Reverse Transcriptase, and/or Superscript II reverse transcriptase. In some embodiments, the reverse transcriptase comprises a viral reverse transcriptase, optionally the viral reverse transcriptase is a murine leukemia virus (MLV) reverse transcriptase or a Moloney murine leukemia virus (MMLV) reverse transcriptase.

[0171] In some embodiments, the each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second molecular label, wherein at least 10 of the second plurality of oligonucleotide barcodes comprise different second molecular label sequences, optionally each second molecular label comprises at least 6 nucleotides, further optionally the second molecular label sequence is a random sequence. In some embodiments, the second plurality of oligonucleotide barcodes are hybridized to the barcoded nucleic acid molecules via hybridization between the second molecular label and the sequence

complementary to at least a portion of the nucleic acid target. In some embodiments, the each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second target-binding region. In some embodiments, the first target-binding region and/or the second target-binding region comprises a poly(dA) region, a poly(dT) region, a random sequence, a gene-specific sequence, or any combination thereof. In some embodiments, oligonucleotide barcode of the second plurality of oligonucleotide barcodes are hybridized to the barcoded nucleic acid molecules via hybridization between the second target-binding region and the sequence complementary to at least a portion of the nucleic acid target. In some embodiments, at least 10 of the second plurality of oligonucleotide barcodes comprise different second target-binding regions, optionally at least two of the target-binding regions are capable of binding the complements of different nucleic acid targets, further optionally at least two of the target-binding regions are capable of hybridizing different regions of the complement of the same nucleic acid target. In some embodiments, two or more oligonucleotide barcodes of the second plurality of oligonucleotide barcodes are capable of hybridizing to different regions of the complement of the same nucleic acid target to generate two or more extended barcoded nucleic acid molecules. In some embodiments, said two or more extended barcoded nucleic acid molecules are capable of being generated by two or more oligonucleotide barcodes of the second plurality of oligonucleotide barcodes hybridizing to different regions of the complement of the same nucleic acid target. In some embodiments, said the two or more extended barcoded nucleic acid molecules collectively comprise the sequence of the entire at least about 50% of the nucleic acid target.

[0172] In some embodiments, the method comprises: denaturing the plurality of barcoded nucleic acid molecules. In some embodiments, the method comprises: denaturing the plurality of extended barcoded nucleic acid molecules. In some embodiments, the method comprises: determining the copy number of the nucleic acid target in the sample based on: the number of first molecular labels with distinct sequences associated with the plurality of barcoded nucleic acid molecules, or products thereof. In some embodiments, the method comprises: determining the copy number of the nucleic acid target in the sample based on: the number of first molecular labels with distinct sequences, second molecular labels with distinct sequences, or a combination thereof, associated with the plurality of extended barcoded nucleic acid molecules, or products thereof. In some embodiments, determining the copy number of the nucleic acid target comprises determining the copy number of each of a plurality of nucleic acid targets in the sample based on: the number of first molecular labels with distinct sequences associated with barcoded nucleic acid molecules of the plurality of barcoded nucleic acid molecules, or products thereof, comprising a sequence of the each of the plurality of nucleic acid

targets; and/or the number of first molecular labels with distinct sequences, second molecular labels with distinct sequences, or a combination thereof, associated with extended barcoded nucleic acid molecules of the plurality of extended barcoded nucleic acid molecules comprising a sequence of the each of the plurality of nucleic acid targets. In some embodiments, the sequence of the each of the plurality of nucleic acid targets comprises a subsequence of the each of the plurality of nucleic acid targets. In some embodiments, the sequence of the nucleic acid target in the plurality of barcoded nucleic acid molecules comprises a subsequence of the nucleic acid target. In some embodiments, the nucleic acid target comprises mRNA. In some embodiments, the sample comprises a single cell, optionally an immune cell, and further optionally a B cell or a T cell. In some embodiments, the sample comprises a plurality of cells, a plurality of single cells, a tissue, a tumor sample, or any combination thereof. In some embodiments, single cell comprises a circulating tumor cell.

[0173] In some embodiments, the first universal sequence of each oligonucleotide barcode of the first plurality of oligonucleotide barcodes is 5' of the first molecular label and the first target-binding region; and/or the second universal sequence of each oligonucleotide barcode of the second plurality of oligonucleotide barcodes is 5' of the second molecular label and/or the second target-binding region.

[0174] In some embodiments, the method comprises: amplifying the plurality of barcoded nucleic acid molecules using an amplification primer and a primer comprising the first universal sequence, or a portion thereof, thereby generating a first plurality of single-labeled nucleic acid molecules comprising the sequence of the nucleic acid target, or a portion thereof, wherein determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of first molecular labels with distinct sequences associated with the first plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, the method comprises: amplifying the plurality of extended barcoded nucleic acid molecules using an amplification primer and a primer comprising the second universal sequence, or a portion thereof, thereby generating a second plurality of single-labeled nucleic acid molecules comprising the sequence of the nucleic acid target, or a portion thereof, wherein determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of second molecular labels with distinct sequences associated with the second plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, the amplification primer comprises a fourth universal sequence. In some embodiments, the amplification primer is a target-specific primer. In some embodiments, the target-specific primer specifically hybridizes to an immune receptor, a

constant region of an immune receptor, a variable region of an immune receptor, a diversity region of an immune receptor, and/or the junction of a variable region and diversity region of an immune receptor. In some embodiments, the immune receptor is a TCR and/or a BCR receptor, and optionally the TCR comprises TCR alpha chain, TCR beta chain, TCR gamma chain, TCR delta chain, or any combination thereof; and the BCR receptor comprises BCR heavy chain and/or BCR light chain.

[0175] In some embodiments, the method comprises: hybridizing random primers to the plurality of barcoded nucleic acid molecules and extending the random primers to generate a first plurality of extension products, wherein the random primers comprise a third universal sequence, or a complement thereof; and amplifying the first plurality of extension products using a primer capable of hybridizing to the third universal sequence, or a complement thereof, and a primer capable of hybridizing to the first universal sequence, or a complement thereof, thereby generating a third plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of first molecular labels with distinct sequences associated with the third plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, the method comprises: hybridizing random primers to the plurality of extended barcoded nucleic acid molecules and extending the random primers to generate a second plurality of extension products, wherein the random primers comprise a third universal sequence, or a complement thereof; and amplifying the second plurality of extension products using a primer capable of hybridizing to the third universal sequence, or a complement thereof, and a primer capable of hybridizing to the second universal sequence, or a complement thereof, thereby generating a fourth plurality of single-labeled nucleic acid molecules, or products thereof. In some embodiments, determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of second molecular labels with distinct sequences associated with the fourth plurality of single-labeled nucleic acid molecules, or products thereof.

[0176] In some embodiments, the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence are the same. In some embodiments, the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence are different. In some embodiments, the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence comprise the binding sites of sequencing primers and/or sequencing adaptors, complementary sequences thereof, and/or portions thereof. In some embodiments, the

sequencing adaptors comprise a P5 sequence, a P7 sequence, complementary sequences thereof, and/or portions thereof. In some embodiments, the sequencing primers comprise a Read 1 sequencing primer, a Read 2 sequencing primer, complementary sequences thereof, and/or portions thereof.

[0177] In some embodiments, the method comprises: obtaining sequence information of the plurality of barcoded nucleic acid molecules, or products thereof. In some embodiments, obtaining the sequence information comprises attaching sequencing adaptors to the plurality of extended barcoded nucleic acid molecules, or products thereof. In some embodiments, the method comprises: obtaining sequence information of the plurality of extended barcoded nucleic acid molecules, or products thereof. In some embodiments, obtaining the sequence information comprises attaching sequencing adaptors to the extended barcoded nucleic acid molecules, barcoded nucleic acid molecules, products thereof, or any combination thereof. In some embodiments, the method comprises: obtaining sequence information of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof, In some embodiments, obtaining the sequence information comprises attaching sequencing adaptors to one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof. In some embodiments, obtaining sequence information of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof, comprises: obtaining sequencing data comprising a plurality of sequencing reads of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof, wherein each of the plurality of sequencing reads comprise (1) a cell label sequence, (2) a molecular label sequence, and/or (3) a subsequence of the nucleic acid target.

[0178] In some embodiments, the method comprises: for each unique cell label sequence, which indicates a single cell of the sample: aligning each of the plurality of sequencing reads of the nucleic acid target to generate an aligned sequence of the nucleic acid target. In some embodiments, the aligned sequence of the nucleic acid target comprises at least 50% of the cDNA sequence of the nucleic acid target, at least 70% of the cDNA sequence of the nucleic acid target, at least 90% of the cDNA sequence of the nucleic acid target, or the full length of the cDNA sequence of the nucleic acid target. In some embodiments, the nucleic acid target is an immune receptor, optionally the immune receptor comprises BCR light chain, BCR heavy chain, TCR alpha chain, TCR beta chain, TCR gamma chain, TCR delta chain, or any combination thereof. In some embodiments, the aligned sequence of the nucleic acid target comprises the complementarity determining region 1 (CDR1), the complementarity determining region 2 (CDR2), the complementarity determining region 3 (CDR3), the variable region, the

full length of the variable region, or a combination thereof. In some embodiments, the aligned sequence of the nucleic acid target comprises the variable region, the diversity region, the junction of a variable region diversity region and/or the constant region, or any combination thereof. In some embodiments, obtaining the sequence information comprises obtaining the sequence information of the BCR light chain and the BCR heavy chain of a single cell, and optionally the sequence information of the BCR light chain and the BCR heavy chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the CDR3, or any combination thereof, of the BCR light chain and/or the BCR heavy chain. In some embodiments, the method comprises: pairing the BCR light chain and the BCR heavy chain of the single cell based on the obtained sequence information. In some embodiments, the sample comprises a plurality of single cells, the method comprising pairing the BCR light chain and the BCR heavy chain of at least 50% of said single cells based on the obtained sequence information. In some embodiments, obtaining the sequence information comprises obtaining the sequence information of the TCR alpha chain and the TCR beta chain of a single cell, and optionally the sequence information of the TCR alpha chain and the TCR beta chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the CDR3, or any combination thereof, of the TCR alpha chain and/or the TCR beta chain. In some embodiments, the method comprises: pairing the TCR alpha chain and the TCR beta chain of the single cell based on the obtained sequence information. In some embodiments, the sample comprises a plurality of single cells, the method comprising pairing the TCR alpha chain and the TCR beta chain of at least 50% of said single cells based on the obtained sequence information. In some embodiments, obtaining the sequence information comprises obtaining the sequence information of the TCR gamma chain and the TCR delta chain of a single cell. In some embodiments, the sequence information of the TCR gamma chain and the TCR delta chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the CDR3, or any combination thereof, of the TCR gamma chain and/or the TCR delta chain. In some embodiments, the method comprises: pairing the TCR gamma chain and the TCR delta chain of the single cell based on the obtained sequence information. In some embodiments, the sample comprises a plurality of single cells, the method comprising pairing the TCR gamma chain and the TCR delta chain of at least 50% of said single cells based on the obtained sequence information.

[0179] In some embodiments, the complement of the molecular label comprises a reverse complementary sequence of the molecular label or a complementary sequence of the molecular label. In some embodiments, the plurality of barcoded nucleic acid molecules comprises barcoded deoxyribonucleic acid (DNA) molecules, barcoded ribonucleic acid (RNA)

molecules, or a combination thereof. In some embodiments, the nucleic acid target comprises a nucleic acid molecule, optionally the nucleic acid molecule comprises ribonucleic acid (RNA), messenger RNA (mRNA), microRNA, small interfering RNA (siRNA), RNA degradation product, RNA comprising a poly(A) tail, or any combination thereof, and further optionally the mRNA encodes an immune receptor. In some embodiments, the nucleic acid target comprises a cellular component binding reagent, and/or the nucleic acid molecule is associated with the cellular component binding reagent, optionally the method further comprising dissociating the nucleic acid molecule and the cellular component binding reagent. In some embodiments, at least 10 of the first and/or second pluralities of oligonucleotide barcodes comprise different molecular label sequences. In some embodiments, each molecular label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides. In some embodiments, the first and/or second pluralities of oligonucleotide barcodes are associated with a solid support. In some embodiments, the first and/or second pluralities of oligonucleotide barcodes associated with the same solid support each comprise an identical sample label. In some embodiments, each sample label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides. In some embodiments, the first and/or second pluralities of oligonucleotide barcodes each comprise a cell label. In some embodiments, each cell label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides. In some embodiments, oligonucleotide barcodes of the first and/or second pluralities of oligonucleotide barcodes associated with the same solid support comprise the same cell label. In some embodiments, oligonucleotide barcodes of the first and/or second pluralities of oligonucleotide barcodes associated with different solid supports comprise different cell labels. In some embodiments, the method comprises: extending the oligonucleotide barcodes in the presence of one or more of ethylene glycol, polyethylene glycol, 1,2-propanediol, dimethyl sulfoxide (DMSO), glycerol, formamide, 7-deaza-GTP, acetamide, tetramethylammonium chloride salt, betaine, or any combination thereof.

[0180] In some embodiments, the solid support comprises a synthetic particle, a planar surface, or a combination thereof. In some embodiments, the sample comprises a single cell, the method comprising associating a synthetic particle comprising the first and second pluralities of oligonucleotide barcodes with the single cell in the sample. In some embodiments, the method comprises: lysing the single cell after associating the synthetic particle with the single cell, optionally lysing the single cell comprises heating the sample, contacting the sample with a detergent, changing the pH of the sample, or any combination thereof. In some embodiments, the synthetic particle and the single cell are in the same partition, and optionally the partition is a well or a droplet. In some embodiments, at least one oligonucleotide barcode of

the first and/or second pluralities of oligonucleotide barcodes is immobilized or partially immobilized on the synthetic particle, or at least one oligonucleotide barcode of the first and/or second pluralities of oligonucleotide barcodes is enclosed or partially enclosed in the synthetic particle. In some embodiments, the synthetic particle is disruptable, optionally a disruptable hydrogel particle. In some embodiments, the synthetic particle comprises a bead, optionally the bead is 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. In some embodiments, the synthetic particle comprises 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. In some embodiments, each oligonucleotide barcode of the first and/or second pluralities of oligonucleotide barcodes comprises a linker functional group. In some embodiments, the synthetic particle comprises a solid support functional group. In some embodiments, the support functional group and the linker functional group are associated with each other, and optionally the linker functional group and the support functional group are individually selected from the group consisting of C6, biotin, streptavidin, primary amine(s), aldehyde(s), ketone(s), and any combination thereof.

[0181] There are provided, in some embodiments, solid supports. In some embodiments, the solid support associated with one or both of a first and second pluralities of oligonucleotide barcodes disclosed herein.

[0182] The methods and systems described herein can be used with methods and systems using antibodies associated with (e.g., attached to or conjugated with) oligonucleotides (also referred to herein as AbOs or AbOligos). Some embodiments of using AbOs to determine protein expression profiles in single cells and tracking sample origins have been described in US2018/0088112, and US2018/0346970; the content of each is incorporated by reference herein in its entirety. In some embodiments, the method disclosed herein allows V(D)J profiling of T cells and B cells, 3' targeted, 5' targeted, 3' whole transcriptome amplification (WTA), 5' WTA, protein expression profiling with AbO, and/or sample multiplexing on a single experiment. Methods for determining the sequences of a nucleic acid target (e.g., the V(D)J region of an immune receptor) using 5' barcoding and/or 3' barcoding are described in US2020/0109437; the content of which is incorporated herein by reference in its entirety. Systems, methods, compositions, and kits for molecular barcoding on the 5'-end of a nucleic acid target have been described in, for example, US2019/0338278, the content of which is incorporated herein by

reference in its entirety. The systems, methods, compositions, and kits for internal-based gene expression profiling provided herein can, in some embodiments, be employed in concert with the methods to obtain full-length V(D)J information (e.g., by Illumina sequencing on the Rhapsody system) using a combined 5' barcoding and random priming approach described in U.S. Patent Application Number 17/091,639, filed on November 6, 2020, entitled "USING RANDOM PRIMING TO OBTAIN FULL-LENGTH V(D)J INFORMATION FOR IMMUNE REPERTOIRE SEQUENCING"; the content of which is incorporated herein by reference in its entirety. The systems, methods, compositions, and kits for internal-based gene expression profiling provided herein can, in some embodiments, be employed in concert with random priming and extension (RPE)-based whole transcriptome analysis methods and compositions have been described in U.S. Patent Application No. 16/677,012; the content of which is incorporated herein by reference in its entirety. The systems, methods, compositions, and kits for internal-based gene expression profiling provided herein can, in some embodiments, be employed in concert with the blocker oligonucleotides described in US20210238661A1, the content of which is incorporated herein by reference in its entirety. Some embodiments of the compositions and methods disclosed herein comprise a first plurality of oligonucleotide barcodes and a second plurality of oligonucleotide barcodes, which have described in US20210371909A1 and content of which is incorporated herein by reference in its entirety. The systems, methods, compositions, and kits provided herein can, in some embodiments, be employed in concert with template switch oligonucleotides comprising a blocking sequence, which, in some embodiments, can reduce the generation of undesirable extension products during library preparation, and which have described in PCT. Patent Application Number PCT/US22/75577, filed on August 29, 2022, entitled "TEMPLATE SWITCH OLIGONUCLEOTIDE (TSO) FOR MRNA 5' ANALYSIS", and content of which is incorporated herein by reference in its entirety.

[0183] In some embodiments, the extension products and/or the amplification products disclosed herein may be used for sequencing. Any suitable sequencing method known in the art can be used, preferably high-throughput approaches. For example, cyclic array sequencing using platforms such as Roche 454, Illumina Solexa, ABI-SOLiD, ION Torrent, Complete Genomics, Pacific Bioscience, Helicos, or the Polonator platform, may also be utilized. Sequencing may comprise MiSeq sequencing and/or HiSeq sequencing.

[0184] Disclosed herein includes systems, methods, compositions, and kits for attachment of barcodes (e.g., stochastic barcodes) with molecular labels (or molecular indices) to the 5'-ends of nucleic acid targets being barcoded or labeled (e.g., deoxyribonucleic acid molecules, and ribonucleic acid molecules). The 5'-based and internal-based transcript counting

methods disclosed herein can complement, or supplement, for example, 3'-based transcript counting methods (e.g., Rhapsody™ assay (Becton, Dickinson and Company, Franklin Lakes, NJ), Chromium™ Single Cell 3' Solution (10X Genomics, San Francisco, CA)). The barcoded nucleic acid targets can be used for sequence identification, transcript counting, alternative splicing analysis, mutation screening, and/or full length sequencing in a high throughput manner. Transcript counting on the 5'-end (5' relative to the target nucleic acid targets being labeled) can reveal alternative splicing isoforms and variants (including, but not limited to, splice variants, single nucleotide polymorphisms (SNPs), insertions, deletions, substitutions.) on, or closer to, the 5'-ends of nucleic acid molecules. In some embodiments, the method can involve intramolecular hybridization. Methods for determining the sequences of a nucleic acid target (e.g., the V(D)J region of an immune receptor) using 5' barcoding and/or 3' barcoding are described in US2020/0109437; the content of which is incorporated herein by reference in its entirety. Systems, methods, compositions, and kits for molecular barcoding on the 5'-end of a nucleic acid target have been described in US2019/0338278, the content of which is incorporated herein by reference in its entirety.

[0185] The methods of the disclosure can be used for identifying VDJ regions of BCR, TCR, and antibodies. VDJ recombination, also known as somatic recombination, is a mechanism of genetic recombination in the early stages of immunoglobulin (Ig) (e.g., BCR) and TCR production of the immune system. VDJ recombination can nearly randomly combine Variable (V), Diverse (D) and Joining (J) gene segments. Because of its randomness in choosing different genes, it is able to diversely encode proteins to match antigens from bacteria, viruses, parasites, dysfunctional cells such as tumor cells and pollens.

[0186] The VDJ region can comprise a large 3 Mb locus comprising variable (V) genes, diversity (D) genes and joining (J) genes. These are the segments that can participate in VDJ recombination. There can be constant genes which may not undergo VDJ recombination. The first event in the VDJ recombination of this locus can be that one of the D genes rearranges to one of the J genes. Following this, one of the V genes can be appended to this DJ rearrangement to form the functional VDJ rearranged gene that then codes for the variable segment of the heavy chain protein. Both of these steps can be catalyzed by recombinase enzymes, which can delete out the intervening DNA.

[0187] This recombination process takes place in a stepwise fashion in progenitor B cells to produce the diversity required for the antibody repertoire. Each B cell may only produce one antibody (e.g., BCR). This specificity can be achieved by allelic exclusion such that functional rearrangement of one allele signals to prevent further recombination of the second allele.

[0188] In some embodiments, the sample comprises an immune cell. An immune cell can include, for example, T cell, B cell, lymphoid stem cell, myeloid progenitor cell, lymphocyte, granulocyte, B-cell progenitor, T cell progenitor, Natural Killer cell, Tc cell, Th cell, plasma cell, memory cell, neutrophil, eosinophil, basophil, mast cell, monocyte, dendritic cell and/or macrophage, or any combination thereof.

[0189] A T cell can be a T cell clone, which can refer to T cells derived from a single T cell or those having identical TCRs. A T cell can be part of a T cell line which can include T cell clones and mixed populations of T cells with different TCRs all of which may recognize the same target (e.g., antigen, tumor, virus). T cells can be obtained from a number of sources, including peripheral blood mononuclear cells, bone marrow, lymph node tissue, spleen tissue, and tumors. T cells can be obtained from a unit of blood collected from a subject, such as using the Ficoll separation. Cells from the circulating blood of an individual can be obtained by apheresis or leukapheresis. The apheresis product can comprise lymphocytes, including T cells, monocytes, granulocytes, B cells, other nucleated white blood cells, red blood cells, and platelets. The cells can be washed and resuspended in media to isolate the cell of interest.

[0190] T cells can be isolated from peripheral blood lymphocytes by lysing the red blood cells and depleting the monocytes, for example, by centrifugation through a PERCOLL™ gradient. A specific subpopulation of T cells, such as CD28+, CD4+, CDC, CD45RA+, and CD45RO+ T cells, can be further isolated by positive or negative selection techniques. For example, T cells can be isolated by incubation with anti-CD3/anti-CD28 (i.e., 3×28)-conjugated beads, such as DYNABEADS® M-450 CD3/CD28 T, or XCYTE DYNABEADS™ for a time period sufficient for positive selection of the desired T cells. Immune cells (e.g., T cells and B cells) can be antigen specific (e.g., specific for a tumor).

[0191] In some embodiments, the cell can be an antigen-presenting cell (APC), such as a B cell, an activated B cell from a lymph node, a lymphoblastoid cell, a resting B-cell, or a neoplastic B cell, e.g. from a lymphoma. An APC can refer to a B-cell or a follicular dendritic cell expressing at least one of the BCRC proteins on its surface.

[0192] The methods of the disclosure can be used to trace the molecular phenotype of single T cells. Different subtypes of T cells can be distinguished by expression of different molecular markers. T cells express a unique TCR from a diverse repertoire of TCRs. In most T cells, the TCR can be composed of a heterodimer of a α and a β chain; each functional chain can be a product of somatic DNA recombination events during T cell development, allowing the expression of over a million different TCRs in a single individual. TCRs can be used to define the identity of individual T cells, allowing for lineage tracing for T cell clonal expansion during an immune response. The immunological methods of the disclosure can be used in a variety of

ways, including but not limited to, identifying unique TCRα and TCRβ chain pairing in single T cells, quantifying TCR and marker expression at the single cell level, identifying TCR diversity in an individual, characterizing the TCR repertoire expressed in different T cell populations, determining functionality of the alpha and beta chain alleles of the TCR, and identifying clonal expansion of T cells during immune response.

[0193] *T-cell receptor chain pairing*

[0194] TCRs are recognition molecules present on the surface of T lymphocytes. The T-cell receptors found on the surface of T-cells can be comprised of two glycoprotein subunits which are referred to as the alpha and beta chains. Both chains can comprise a molecular weight of about 40 kDa and possess a variable and a constant domain. The genes which encode the alpha and beta chains can be organized in libraries of V, D and J regions from which the genes are formed by genetic rearrangement. TCRs can recognize antigen which is presented by an antigen presenting cell as a part of a complex with a specific self-molecule encoded by a histocompatibility gene. The most potent histocompatibility genes are known as the major histocompatibility complex (MHC). The complex which is recognized by T-cell receptors, therefore, consists of and MHC/peptide ligand.

[0195] In some embodiments, the methods, devices, and systems of the disclosure can be used for TCR sequencing and pairing. The methods, devices, and systems of the disclosure can be used for sequencing T-cell receptor alpha and beta chains, pairing alpha and beta chains, and/or determining the functional copy of T-cell receptor alpha chains. A single cell can be contained in a single partition (e.g., well) with a single solid support (e.g., bead). The cell can be lysed. The bead can comprise a stochastic label that can bind to a specific location within an alpha and/or beta chain of a TCR. The TCR alpha and beta molecules associated with solid support can be subjected to the molecular biology methods of the disclosure, including reverse transcription, amplification, and sequencing. TCR alpha and beta chains that comprise the same cellular label can be considered to be from the same single cell, thereby pairing alpha and beta chains of the TCR.

[0196] *Heavy and light chain pairing in antibody repertoires*

[0197] The methods devices and systems of the disclosure can be used for heavy and light chain pairing of BCR receptors and antibodies. The methods of the present disclosure allow for the repertoire of immune receptors and antibodies in an individual organism or population of cells to be determined. The methods of the present disclosure can aid in determining pairs of polypeptide chains that make up immune receptors. B cells and T cells each express immune receptors; B cells express immunoglobulins and BCRs, and T cells express TCRs. Both types of immune receptors can comprise two polypeptide chains. Immunoglobulins can comprise

variable heavy (VH) and variable light (VL) chains. There can be two types of TCRs: one consisting of an alpha and a beta chain, and one consisting of a delta and a gamma chain. Polypeptides in an immune receptor can comprise constant region and a variable region. Variable regions can result from recombination and end joint rearrangement of gene fragments on the chromosome of a B or T cell. In B cells additional diversification of variable regions can occur by somatic hypermutation.

[0198] The immune system has a large repertoire of receptors, and any given receptor pair expressed by a lymphocyte can be encoded by a pair of separate, unique transcripts. Knowing the sequences of pairs of immune receptor chains expressed in a single cell can be used to ascertain the immune repertoire of a given individual or population of cells.

[0199] In some embodiments, the methods, devices, and systems of the disclosure can be used for antibody sequencing and pairing. The methods, devices, and systems of the disclosure can be used for sequencing antibody heavy and light chains (e.g., in B cells), and/or pairing the heavy and light chains. A single cell can be contained in a single partition (e.g., well) with a single solid support (e.g., bead). The cell can be lysed.

[0200] The bead can comprise a stochastic label that can bind to a specific location within a heavy and/or light chain of an antibody (e.g., in a B cell). The heavy and light chain molecules associated with solid support can be subjected to the molecular biology methods of the disclosure, including reverse transcription, amplification, and sequencing. Antibody heavy and light chains that comprise the same cellular label can be considered to be from the same single cell, thereby pairing heavy and light chains of the antibody.

[0201] The method disclosed herein can allow 3'-based, internal-based, and/or 5'-based sequence determination. This method can enable provide flexibility to sequence determination. In some embodiments, the method can enable immune repertoire profiling of both T cells and B cells on a Rhapsody™ system, for samples such as mouse and human samples. In some embodiments, 3', internal, and/or 5' expression profiling of V(D)J can be performed. In some embodiments, both phenotypic markers and V(D)J sequence of T cell and B cells in single cell platforms can be investigated. In some embodiments, 3', internal, and 5' information of their transcripts can be captured in a single experiment. The method disclosed herein can allow V(D)J detection of both T cells and B cells (e.g., hypermutation).

[0202] The methods and systems described herein can be used with methods and systems using antibodies associated with (e.g., attached to or conjugated with) oligonucleotides (also referred to herein as AbOs or AbOligos). Embodiments of using AbOs to determine protein expression profiles in single cells and tracking sample origins have been described in U.S. Patent Application No. 15/715,028, published as U.S. Patent Application Publication No.

2018/0088112, and U.S. Patent Application No. 15/937,713; the content of each is incorporated by reference herein in its entirety. In some embodiments, the method disclosed herein allows V(D)J profiling of T cells and B cells, 3' targeted, 5' targeted, 3' whole transcriptome amplification (WTA), 5' WTA, protein expression profiling with AbO, and/or sample multiplexing on a single experiment.

[0203] In some embodiments, the step of extending the random primers is conducted at an approximately constant temperature. In some embodiments, the step of extending the random primers is conducted at an invariant temperature. In some embodiments, the step of extending the random primers begins at a first extension temperature. In some embodiments, the step of extending the random primers is conducted at one or more different temperatures than the first extension temperature (e.g., a second extension temperature and/or a third extension temperature). The second extension temperature and/or third extension temperature can higher or lower than the first extension temperature. In some embodiments, the first extension temperature and/or the second extension temperature is about 30°C, 31°C, 32°C, 33°C, 34°C, 35°C, 36°C, 37°C, 38°C, 39°C, 40°C, 41°C, 42°C, 43°C, 44°C, 45°C, 46°C, 47°C, 48°C, 49°C, 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, 60°C, 61°C, 62°C, 63°C, 64°C, 65°C, 66°C, 67°C, 68°C, 69°C, 70°C, 71°C, 72°C, 73°C, 74°C, 75°C, 76°C, 77°C, 78°C, 79°C, 80°C, or a number or a range between any two of these values. In some embodiments, the first extension temperature is about 37°C. In some embodiments, the second extension temperature is about 55°C. In some embodiments, the second extension temperature is about 45°C.

[0204] The number of cycles of random priming and extension can be different in different implementations. In some embodiments, the number of cycles of random priming and extension can comprise, comprise about, comprise at least, or comprise 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, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 60, 70, 80, 90, 100, or a number or a range between any two of these values, cycles of random priming and extension.

[0205] The random primers can comprise a random sequence of nucleotides. The random sequence of nucleotides can be about 4 to about 30 nucleotides in length. In some embodiments, said random sequence of nucleotides is 6 or 9 nucleotides in length. The random sequence of nucleotides can have different lengths in different implementations. In some embodiments, the random sequence of nucleotides within the random primers is, is about, at least, or is at most, 1, 2, 3, 4, 5, 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, or a number or a range between any two of these values, nucleotides in length. The random primers can have different concentrations during the random priming step in different implementations. In some embodiments, the random primer is, is about,

is at least, or is at most, 1, 2, 3, 4, 5, 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, or a number or a range between any two of these values, uM in concentration during the random priming.

[0206] Methods for manufacturing oligonucleotide barcodes and barcoding particles has been described in, for example, US 2015/0299784, WO 2015/031691, Fu et al, PNAS U.S.A. 2011 May 31;108(22):9026-31, and in U.S. Patent Application Number 17/336,055, filed on June 1, 2021, entitled "OLIGONUCLEOTIDES AND BEADS FOR 5 PRIME GENE EXPRESSION ASSAY," the content of these applications and publications is incorporated hereby in its entirety.

[0207] 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.

[0208] One skilled in the art will appreciate that, for this and other processes and methods disclosed herein, the functions performed in the processes and methods can be implemented in differing order. Furthermore, the outlined steps and operations are only provided as examples, and some of the steps and operations can be optional, combined into fewer steps and operations, or expanded into additional steps and operations without detracting from the essence of the disclosed embodiments.

[0209] 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.

[0210] 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.”

[0211] 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.

[0212] 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 subranges and combinations of subranges 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,” and the like include the number recited and refer to ranges which can be subsequently broken down into subranges 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 cells refers to groups having 1, 2, or 3 cells. Similarly, a group having 1-5 cells refers to groups having 1, 2, 3,

4, or 5 cells, and so forth.

[0213] From the foregoing, it will be appreciated that various embodiments of the present disclosure have been described herein for purposes of illustration, and that various modifications may be made without departing from the scope and spirit of the present disclosure. Accordingly, the various embodiments disclosed herein are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

WHAT IS CLAIMED IS:

1. A method for labeling nucleic acid targets in a sample, comprising:

contacting copies of a nucleic acid target with a first plurality of oligonucleotide barcodes, wherein each oligonucleotide barcode of the first plurality of oligonucleotide barcodes comprises a first universal sequence, a first molecular label, and a first target-binding region capable of hybridizing to the nucleic acid target;

extending the first plurality of oligonucleotide barcodes hybridized to the copies of the nucleic acid target to generate a plurality of barcoded nucleic acid molecules each comprising the first universal sequence, the first molecular label, and a sequence complementary to at least a portion of the nucleic acid target;

contacting the barcoded nucleic acid molecules with a second plurality of oligonucleotide barcodes for hybridization,

wherein each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second universal sequence, a cleavage domain, and a blocking group,

wherein the blocking group is capable of preventing extension of the oligonucleotide barcode, wherein the cleavage domain is positioned 5' of the blocking group, and wherein a cleaving enzyme is capable of cleaving the oligonucleotide barcode at a point within or adjacent to the cleavage domain when the cleavage domain is hybridized to a barcoded nucleic acid molecule;

contacting the second plurality of oligonucleotide barcodes hybridized to the barcoded nucleic acid molecules with the cleaving enzyme, thereby removing the blocking group from said oligonucleotide barcodes;

extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes hybridized to the barcoded nucleic acid molecules to generate a plurality of extended barcoded nucleic acid molecules.

2. The method of claim 1, wherein each extended barcoded nucleic acid molecule of the plurality of extended barcoded nucleic acid molecules comprise a sequence of at least a portion of the nucleic acid target.

3. The method of any one of claims 1-2, wherein the cleaving enzyme is an RNase H enzyme and/or an RNase H2 enzyme, optionally the RNase H2 enzyme is a *Pyrococcus abyssi* RNase H2 enzyme.

4. The method of any one of claims 1-3, wherein the cleaving enzyme is a hot start cleaving enzyme which is thermostable and has reduced activity at lower temperatures.

5. The method of any one of claims 1-4, wherein the hot start cleaving enzyme is Pyrococcus abyssi RNase H2 comprising (a) a G12A amino acid substitution; (b) a P13T amino acid substitution; (c) a G169A amino acid substitution; or (d) a combination thereof.

6. The method of any one of claims 1-5, wherein the cleaving enzyme is chemically modified.

7. The method of any one of claims 1-6, wherein the cleaving enzyme is a chemically modified hot start cleaving enzyme which is thermostable and has reduced activity at lower temperatures, optionally the cleaving enzyme is reversibly inactivated through interaction with an antibody at lower temperatures.

8. The method of any one of claims 1-7, wherein the cleavage domain comprises one or more ribonucleotides that are capable of being cleaved by a RNase H enzyme.

9. The method of any one of claims 1-8, wherein the cleavage domain comprises one or more of the following moieties: a DNA residue, an abasic residue, a modified nucleoside, or a modified phosphate internucleotide linkage.

10. The method of any one of claims 1-9, wherein the cleavage domain comprises at least one RNA base.

11. The method of any one of claims 1-10, wherein the cleavage domain comprises one or more 2'-modified nucleosides, optionally the one or more modified nucleosides are 2'-fluoronucleosides.

12. The method of any one of claims 1-11, wherein the blocking group is attached to the 3' -terminal nucleotide of the oligonucleotide barcode.

13. The method of any one of claims 1-12, wherein the blocking group is at or near 3' terminal of the oligonucleotide barcode.

14. The method of any one of claims 1-13, wherein the blocking group is a 2', 3'-dideoxynucleotide, a ribonucleotide residue, a 2', 3' SH nucleotide, or a 2'-O-PO3 nucleotide.

15. The method of any one of claims 1-14, wherein the blocking group comprises a non-nucleotide modification.

16. The method of any one of claims 1-15, wherein the blocking group further comprises a napthyl-azo compound, a spacer and/or a biotin.

17. The method of any one of claims 1-16, wherein extending the extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that has strand displacement activity.

18. The method of claim 17, wherein extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that has

strand displacement activity is capable of generating extended barcoded nucleic acid molecules comprising a complement of the first molecular label and a complement of the first universal sequence.

19. The method of any one of claims 1-18, wherein extending the extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes of the second plurality of oligonucleotide barcodes with a DNA polymerase that does not have strand displacement activity.

20. The method of any one of claims 1-19, wherein the polymerase is selected from the group comprising Phi29 DNA polymerase, E.coli DNA polymerase I, Bsu DNA polymerase, Bst DNA polymerase, Taq DNA polymerase, VENT™ DNA polymerase, DEEPVENT™ DNA polymerase, LongAmp® Taq DNA polymerase, LongAmp® Hot Start Taq DNA polymerase, Crimson LongAmp® Taq DNA polymerase, Crimson Taq DNA polymerase, OneTaq® DNA polymerase, OneTaq® Quick- Load® DNA polymerase, Hemo KlenTaq® DNA polymerase, REDTaq® DNA polymerase, Phusion® DNA polymerase, Phusion® High-Fidelity DNA polymerase, Platinum Pfx DNA polymerase, AccuPrime Pfx DNA polymerase, Klenow fragment, Pwo DNA polymerase, Pfu DNA polymerase, T4 DNA polymerase, T7 DNA polymerase, derivatives thereof, or any combination thereof.

21. The method of any one of claims 1-20, wherein extending the 3' ends of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes using a mesophilic DNA polymerase, a thermophilic DNA polymerase, a psychrophilic DNA polymerase, or any combination thereof.

22. The method of any one of claims 1-21, wherein extending the 3' ends of oligonucleotide barcodes comprises extending the 3' ends of oligonucleotide barcodes using a DNA polymerase lacking at least one of 5' to 3' exonuclease activity and 3' to 5' exonuclease activity, and optionally the DNA polymerase comprises a Klenow Fragment.

23. The method of any one of claims 1-22, wherein extending the first plurality of oligonucleotide barcodes comprises extending the first plurality of oligonucleotide barcodes using a reverse transcriptase.

24. The method of any one of claims 1-23, wherein the reverse transcriptase is capable of terminal transferase activity.

25. The method of any one of claims 1-24, wherein the reverse transcriptase with strand displacement activity is a PrimeScript reverse transcriptase, M-MuLV reverse transcriptase, SmartScribe reverse transcriptase, Maxima H Minus Reverse Transcriptase, and/or Superscript II reverse transcriptase.

26. The method of any one of claims 1-25, wherein the reverse transcriptase comprises a viral reverse transcriptase, optionally the viral reverse transcriptase is a murine leukemia virus (MLV) reverse transcriptase or a Moloney murine leukemia virus (MMLV) reverse transcriptase.

27. The method of any one of claims 1-26, wherein the each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second molecular label, wherein at least 10 of the second plurality of oligonucleotide barcodes comprise different second molecular label sequences, optionally each second molecular label comprises at least 6 nucleotides, further optionally the second molecular label sequence is a random sequence.

28. The method of any one of claims 1-27, wherein the second plurality of oligonucleotide barcodes are hybridized to the barcoded nucleic acid molecules via hybridization between the second molecular label and the sequence complementary to at least a portion of the nucleic acid target.

29. The method of any one of claims 1-28, wherein the each oligonucleotide barcode of the second plurality of oligonucleotide barcodes comprises a second target-binding region.

30. The method of any one of claims 1-29, wherein the first target-binding region and/or the second target-binding region comprises a poly(dA) region, a poly(dT) region, a random sequence, a gene-specific sequence, or any combination thereof.

31. The method of any one of claims 1-30, wherein oligonucleotide barcode of the second plurality of oligonucleotide barcodes are hybridized to the barcoded nucleic acid molecules via hybridization between the second target-binding region and the sequence complementary to at least a portion of the nucleic acid target.

32. The method of any one of claims 1-31, wherein at least 10 of the second plurality of oligonucleotide barcodes comprise different second target-binding regions, optionally at least two of the target-binding regions are capable of binding the complements of different nucleic acid targets, further optionally at least two of the target-binding regions are capable of hybridizing different regions of the complement of the same nucleic acid target.

33. The method of any one of claims 1-32, wherein two or more oligonucleotide barcodes of the second plurality of oligonucleotide barcodes are capable of hybridizing to different regions of the complement of the same nucleic acid target to generate two or more extended barcoded nucleic acid molecules.

34. The method of any one of claims 1-33, wherein said two or more extended barcoded nucleic acid molecules are capable of being generated by two or more oligonucleotide barcodes of the second plurality of oligonucleotide barcodes hybridizing to different regions of the complement of the same nucleic acid target.

35. The method of any one of claims 1-34, wherein said the two or more extended barcoded nucleic acid molecules collectively comprise the sequence of the entire at least about 50% of the nucleic acid target.

36. The method of any one of claims 1-35, the method comprising denaturing the plurality of barcoded nucleic acid molecules.

37. The method of any one of claims 1-36, the method comprising denaturing the plurality of extended barcoded nucleic acid molecules.

38. The method of any one of claims 1-37, further comprising determining the copy number of the nucleic acid target in the sample based on:

the number of first molecular labels with distinct sequences associated with the plurality of barcoded nucleic acid molecules, or products thereof.

39. The method of any one of claims 1-38, further comprising determining the copy number of the nucleic acid target in the sample based on:

the number of first molecular labels with distinct sequences, second molecular labels with distinct sequences, or a combination thereof, associated with the plurality of extended barcoded nucleic acid molecules, or products thereof.

40. The method of any one of claims 1-39, wherein determining the copy number of the nucleic acid target comprises determining the copy number of each of a plurality of nucleic acid targets in the sample based on:

the number of first molecular labels with distinct sequences associated with barcoded nucleic acid molecules of the plurality of barcoded nucleic acid molecules, or products thereof, comprising a sequence of the each of the plurality of nucleic acid targets; and/or

the number of first molecular labels with distinct sequences, second molecular labels with distinct sequences, or a combination thereof, associated with extended barcoded nucleic acid molecules of the plurality of extended barcoded nucleic acid molecules comprising a sequence of the each of the plurality of nucleic acid targets.

41. The method of any one of claims 1-40, wherein the sequence of the each of the plurality of nucleic acid targets comprises a subsequence of the each of the plurality of nucleic acid targets.

42. The method of any one of claims 1-41, wherein the sequence of the nucleic acid target in the plurality of barcoded nucleic acid molecules comprises a subsequence of the nucleic acid target.

43. The method of any one of claims 1-42, wherein the nucleic acid target comprises mRNA.

44. The method of any one of claims 1-43, wherein the sample comprises a single cell, optionally an immune cell, and further optionally a B cell or a T cell.

45. The method of any one of claims 1-44, wherein the sample comprises a plurality of cells, a plurality of single cells, a tissue, a tumor sample, or any combination thereof.

46. The method of any one of claims 44-45, wherein single cell comprises a circulating tumor cell.

47. The method of any one of claims 1-46, wherein:

the first universal sequence of each oligonucleotide barcode of the first plurality of oligonucleotide barcodes is 5' of the first molecular label and the first target-binding region; and/or

the second universal sequence of each oligonucleotide barcode of the second plurality of oligonucleotide barcodes is 5' of the second molecular label and/or the second target-binding region.

48. The method of any one of claims 1-47,

comprising amplifying the plurality of barcoded nucleic acid molecules using an amplification primer and a primer comprising the first universal sequence, or a portion thereof, thereby generating a first plurality of single-labeled nucleic acid molecules comprising the sequence of the nucleic acid target, or a portion thereof,

wherein determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of first molecular labels with distinct sequences associated with the first plurality of single-labeled nucleic acid molecules, or products thereof.

49. The method of any one of claims 1-48,

comprising amplifying the plurality of extended barcoded nucleic acid molecules using an amplification primer and a primer comprising the second universal sequence, or a portion thereof, thereby generating a second plurality of single-labeled nucleic acid molecules comprising the sequence of the nucleic acid target, or a portion thereof,

wherein determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of second molecular labels with distinct sequences associated with the second plurality of single-labeled nucleic acid molecules, or products thereof.

50. The method of any one of claims 48-49, wherein the amplification primer comprises a fourth universal sequence.

51. The method of any one of claims 48-50, wherein the amplification primer is a target-specific primer.

52. The method of claim 51, wherein the target-specific primer specifically hybridizes to an immune receptor, a constant region of an immune receptor, a variable region of an immune receptor, a diversity region of an immune receptor, and/or the junction of a variable region and diversity region of an immune receptor.

53. The method of claim 52, wherein the immune receptor is a T cell receptor (TCR) and/or a B cell receptor (BCR) receptor, and optionally

the TCR comprises TCR alpha chain, TCR beta chain, TCR gamma chain, TCR delta chain, or any combination thereof; and

the BCR receptor comprises BCR heavy chain and/or BCR light chain.

54. The method of any one of claims 1-53, comprising:

hybridizing random primers to the plurality of barcoded nucleic acid molecules and extending the random primers to generate a first plurality of extension products, wherein the random primers comprise a third universal sequence, or a complement thereof; and

amplifying the first plurality of extension products using a primer capable of hybridizing to the third universal sequence, or a complement thereof, and a primer capable of hybridizing to the first universal sequence, or a complement thereof, thereby generating a third plurality of single-labeled nucleic acid molecules, or products thereof.

55. The method of any one of claims 1-54, wherein determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid target in the sample based on the number of first molecular labels with distinct sequences associated with the third plurality of single-labeled nucleic acid molecules, or products thereof.

56. The method of any one of claims 1-55, comprising:

hybridizing random primers to the plurality of extended barcoded nucleic acid molecules and extending the random primers to generate a second plurality of extension products, wherein the random primers comprise a third universal sequence, or a complement thereof; and

amplifying the second plurality of extension products using a primer capable of hybridizing to the third universal sequence, or a complement thereof, and a primer capable of hybridizing to the second universal sequence, or a complement thereof, thereby generating a fourth plurality of single-labeled nucleic acid molecules, or products thereof.

57. The method of any one of claims 1-56, wherein determining the copy number of the nucleic acid target in the sample comprises: determining the copy number of the nucleic acid

target in the sample based on the number of second molecular labels with distinct sequences associated with the fourth plurality of single-labeled nucleic acid molecules, or products thereof.

58. The method of any one of claims 1-57, wherein the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence are the same.

59. The method of any one of claims 1-58, wherein the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence are different.

60. The method of any one of claims 1-59, wherein the first universal sequence, the second universal sequence, third universal sequence, and/or the fourth universal sequence comprise the binding sites of sequencing primers and/or sequencing adaptors, complementary sequences thereof, and/or portions thereof.

61. The method of claim 60, wherein the sequencing adaptors comprise a P5 sequence, a P7 sequence, complementary sequences thereof, and/or portions thereof.

62. The method of any one of claims 60-61, wherein the sequencing primers comprise a Read 1 sequencing primer, a Read 2 sequencing primer, complementary sequences thereof, and/or portions thereof.

63. The method of any one of claims 1-62, comprising obtaining sequence information of the plurality of barcoded nucleic acid molecules, or products thereof.

64. The method of claim 63, wherein obtaining the sequence information comprises attaching sequencing adaptors to the plurality of extended barcoded nucleic acid molecules, or products thereof.

65. The method of any one of claims 1-64, comprising obtaining sequence information of the plurality of extended barcoded nucleic acid molecules, or products thereof.

66. The method of claim 65, wherein obtaining the sequence information comprises attaching sequencing adaptors to the extended barcoded nucleic acid molecules, barcoded nucleic acid molecules, products thereof, or any combination thereof.

67. The method of any one of claims 1-66, comprising obtaining sequence information of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof,

68. The method of claim 67, wherein obtaining the sequence information comprises attaching sequencing adaptors to one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof.

69. The method of any one of claims 67-68, wherein obtaining sequence information of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof, comprises:

obtaining sequencing data comprising a plurality of sequencing reads of one or more of the first, second, third, and fourth pluralities of single-labeled nucleic acid molecules, or products thereof,

wherein each of the plurality of sequencing reads comprise (1) a cell label sequence, (2) a molecular label sequence, and/or (3) a subsequence of the nucleic acid target.

70. The method of claim 69, comprising:

for each unique cell label sequence, which indicates a single cell of the sample:

aligning each of the plurality of sequencing reads of the nucleic acid target to generate an aligned sequence of the nucleic acid target.

71. The method of claim 70, wherein the aligned sequence of the nucleic acid target comprises at least 50% of the cDNA sequence of the nucleic acid target, at least 70% of the cDNA sequence of the nucleic acid target, at least 90% of the cDNA sequence of the nucleic acid target, or the full length of the cDNA sequence of the nucleic acid target.

72. The method of any one of claims 1-71, wherein the nucleic acid target is an immune receptor, optionally the immune receptor comprises BCR light chain, BCR heavy chain, TCR alpha chain, TCR beta chain, TCR gamma chain, TCR delta chain, or any combination thereof.

73. The method of any one of claims 70-72, wherein the aligned sequence of the nucleic acid target comprises the complementarity determining region 1 (CDR1), the complementarity determining region 2 (CDR2), the complementarity determining region 3 (CDR3), the variable region, the full length of the variable region, or a combination thereof.

74. The method of any one of claims 70-73, wherein the aligned sequence of the nucleic acid target comprises the variable region, the diversity region, the junction of a variable region diversity region and/or the constant region, or any combination thereof.

75. The method of any one of claims 63-74, wherein:

obtaining the sequence information comprises obtaining the sequence information of the BCR light chain and the BCR heavy chain of a single cell;

the sequence information of the BCR light chain and the BCR heavy chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the CDR3, or any combination thereof, of the BCR light chain and/or the BCR heavy chain;

the method comprises pairing the BCR light chain and the BCR heavy chain of the single cell based on the obtained sequence information; and/or

the sample comprises a plurality of single cells, the method comprising pairing the BCR light chain and the BCR heavy chain of at least 50% of said single cells based on the obtained sequence information.

76. The method of any one of claims 63-75, wherein:

obtaining the sequence information comprises obtaining the sequence information of the TCR alpha chain and the TCR beta chain of a single cell;

the sequence information of the TCR alpha chain and the TCR beta chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the CDR3, or any combination thereof, of the TCR alpha chain and/or the TCR beta chain;

the method comprises pairing the TCR alpha chain and the TCR beta chain of the single cell based on the obtained sequence information; and/or

the sample comprises a plurality of single cells, the method comprising pairing the TCR alpha chain and the TCR beta chain of at least 50% of said single cells based on the obtained sequence information.

77. The method of any one of claims 63-76, wherein:

obtaining the sequence information comprises obtaining the sequence information of the TCR gamma chain and the TCR delta chain of a single cell;

the sequence information of the TCR gamma chain and the TCR delta chain comprises the sequence of the complementarity determining region 1 (CDR1), the CDR2, the CDR3, or any combination thereof, of the TCR gamma chain and/or the TCR delta chain;

the method comprises pairing the TCR gamma chain and the TCR delta chain of the single cell based on the obtained sequence information; and/or

the sample comprises a plurality of single cells, the method comprising pairing the TCR gamma chain and the TCR delta chain of at least 50% of said single cells based on the obtained sequence information.

78. The method of any one of claims 1-77, wherein the complement of the molecular label comprises a reverse complementary sequence of the molecular label or a complementary sequence of the molecular label.

79. The method of any one of claims 1-78, wherein the plurality of barcoded nucleic acid molecules comprises barcoded deoxyribonucleic acid (DNA) molecules, barcoded ribonucleic acid (RNA) molecules, or a combination thereof.

80. The method of any one of claims 1-79, wherein the nucleic acid target comprises a nucleic acid molecule, optionally the nucleic acid molecule comprises ribonucleic acid (RNA), messenger RNA (mRNA), microRNA, small interfering RNA (siRNA), RNA degradation product, RNA comprising a poly(A) tail, or any combination thereof, and further optionally the mRNA encodes an immune receptor.

81. The method of any one of claims 1-80, wherein the nucleic acid target comprises a cellular component binding reagent, and/or the nucleic acid molecule is associated with the cellular component binding reagent, optionally the method further comprising dissociating the nucleic acid molecule and the cellular component binding reagent.

82. The method of any one of claims 1-81, wherein at least 10 of the first and/or second pluralities of oligonucleotide barcodes comprise different molecular label sequences.

83. The method of any one of claims 1-82, wherein each molecular label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides.

84. The method of any one of claims 1-83, wherein the first and/or second pluralities of oligonucleotide barcodes are associated with a solid support.

85. The method of any one of claims 1-84, wherein the first and/or second pluralities of oligonucleotide barcodes associated with the same solid support each comprise an identical sample label.

86. The method of any one of claims 1-85, wherein each sample label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides.

87. The method of any one of claims 1-86, wherein the first and/or second pluralities of oligonucleotide barcodes each comprise a cell label.

88. The method of any one of claims 1-87, wherein each cell label of the first and/or second pluralities of oligonucleotide barcodes comprises at least 6 nucleotides.

89. The method of any one of claims 1-88, wherein oligonucleotide barcodes of the first and/or second pluralities of oligonucleotide barcodes associated with the same solid support comprise the same cell label.

90. The method of any one of claims 1-89, wherein oligonucleotide barcodes of the first and/or second pluralities of oligonucleotide barcodes associated with different solid supports comprise different cell labels.

91. The method of any one of claims 1-90, comprising extending the oligonucleotide barcodes in the presence of one or more of ethylene glycol, polyethylene glycol, 1,2-propanediol, dimethyl sulfoxide (DMSO), glycerol, formamide, 7-deaza-GTP, acetamide, tetramethylammonium chloride salt, betaine, or any combination thereof.

92. The method of any one of claims 1-91, wherein the solid support comprises a synthetic particle, a planar surface, or a combination thereof.

93. The method of any one of claims 1-92, wherein the sample comprises a single cell, the method comprising associating a synthetic particle comprising the first and second pluralities of oligonucleotide barcodes with the single cell in the sample.

94. The method of any one of claims 1-93, comprising lysing the single cell after associating the synthetic particle with the single cell, optionally lysing the single cell comprises heating the sample, contacting the sample with a detergent, changing the pH of the sample, or any combination thereof.

95. The method of any one of claims 1-94, wherein the synthetic particle and the single cell are in the same partition, and optionally the partition is a well or a droplet.

96. The method of any one of claims 1-95, wherein at least one oligonucleotide barcode of the first and/or second pluralities of oligonucleotide barcodes is immobilized or partially immobilized on the synthetic particle, or at least one oligonucleotide barcode of the first and/or second pluralities of oligonucleotide barcodes is enclosed or partially enclosed in the synthetic particle.

97. The method of any one of claims 1-96, wherein:

the synthetic particle is disruptable, optionally a disruptable hydrogel particle;

the synthetic particle comprises a bead, optionally the bead is 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; and/or

the synthetic particle comprises 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.

98. The method of any one of claims 1-97,

wherein each oligonucleotide barcode of the first and/or second pluralities of oligonucleotide barcodes comprises a linker functional group,

wherein the synthetic particle comprises a solid support functional group, and

wherein the support functional group and the linker functional group are associated with each other, and optionally the linker functional group and the support functional group are

individually selected from the group consisting of C6, biotin, streptavidin, primary amine(s), aldehyde(s), ketone(s), and any combination thereof.

99. A solid support associated with one or both of a first and second pluralities of oligonucleotide barcodes of any one of claims 1-98.

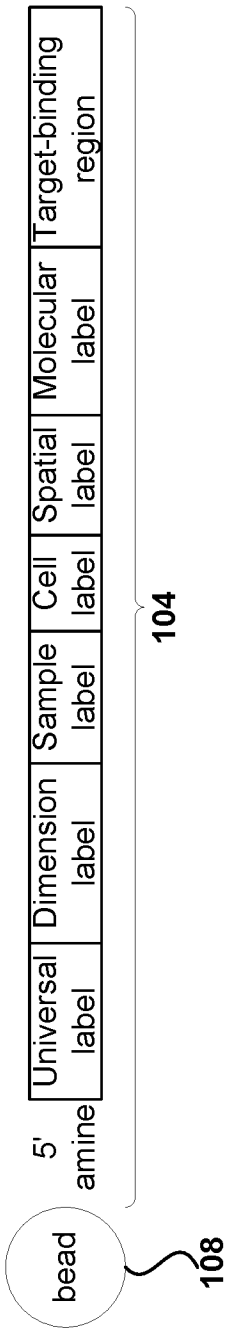

bead
108
5' amine
Universal label
Dimension label
Sample label
Cell label
Spatial label
Molecular label
Target-binding region
104

FIG. 1

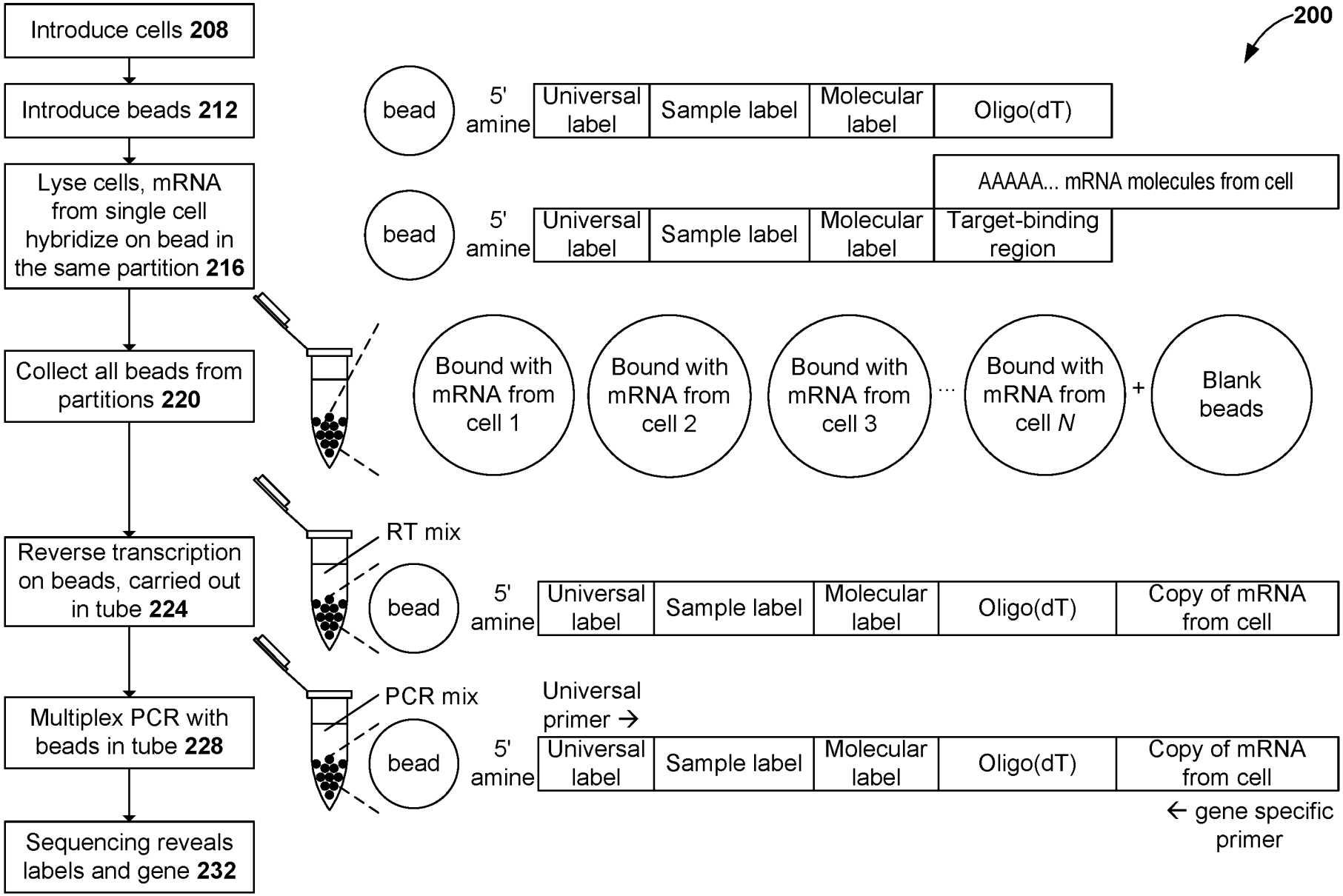

Introduce cells 208
Introduce beads 212
Lyse cells, mRNA from single cell hybridize on bead in the same partition 216
Collect all beads from partitions 220
Reverse transcription on beads, carried out in tube 224
Multiplex PCR with beads in tube 228
Sequencing reveals labels and gene 232
200
bead
5' amine
Universal label
Sample label
Molecular label
Oligo(dT)
AAAAA... mRNA molecules from cell
Target-binding region
Bound with mRNA from cell 1
Bound with mRNA from cell 2
Bound with mRNA from cell 3
Bound with mRNA from cell N
Blank beads
RT mix
PCR mix
Universal primer →
Copy of mRNA from cell
← gene specific primer

FIG. 2

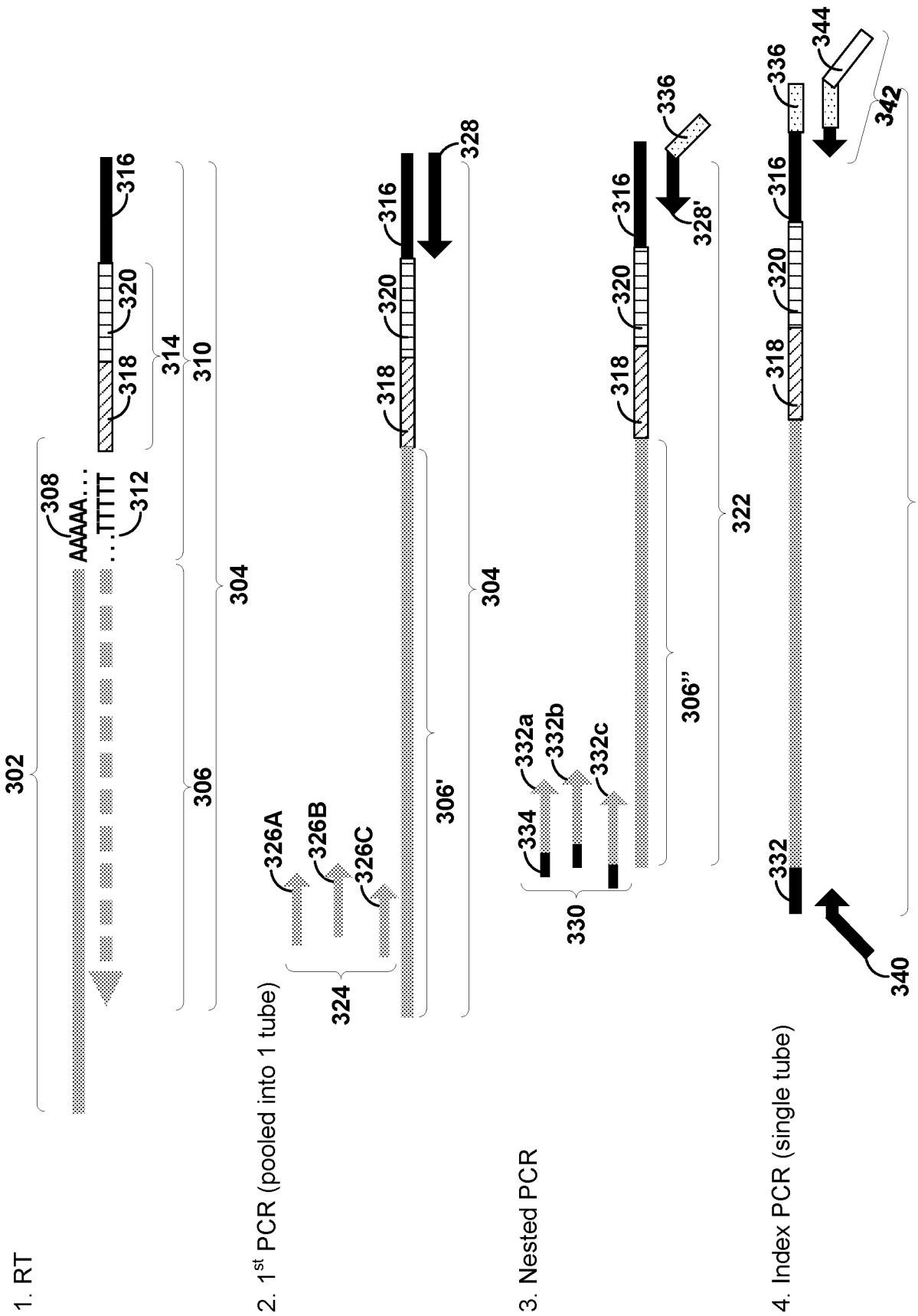

1. RT
2. 1st PCR (pooled into 1 tube)
3. Nested PCR
4. Index PCR (single tube)
302
308
AAAAA...
...TTTTT
312
316
320
318
314
310
304
306
306'
326A
326B
326C
324
328
330
334
332a
332b
332c
306''
322
328'
336
332
340
338
342
344

FIG. 3

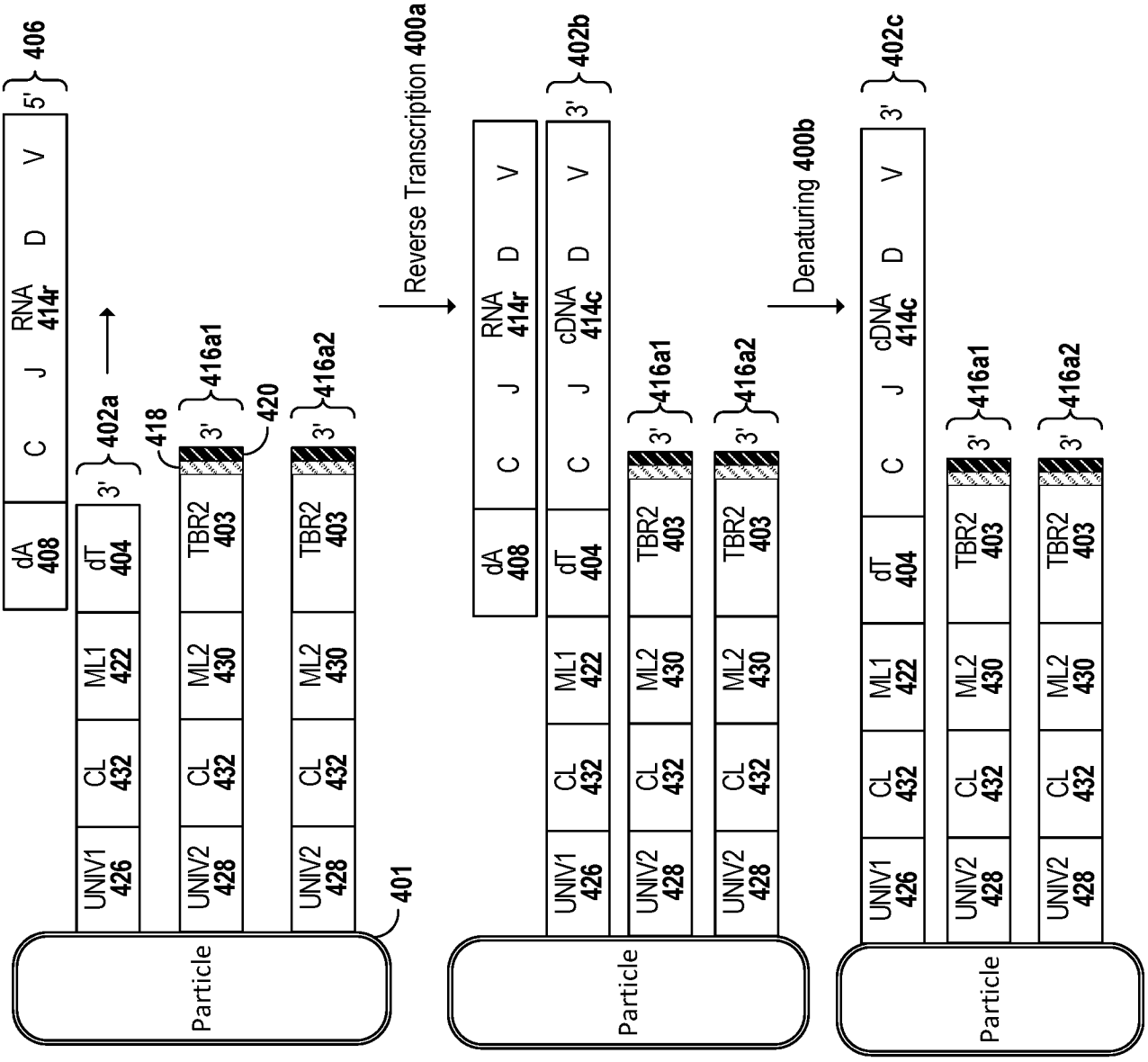

Particle
UNIV1 426
CL 432
ML1 422
dT 404
dA 408
3' 402a
C J RNA 414r D V 5' 406
418
3'
TBR2 403
416a1
420
ML2 430
416a2
401
Reverse Transcription 400a
cDNA 414c
402b
Denaturing 400b
402c
Intermolecular Hybridization 400c

FIG. 4A

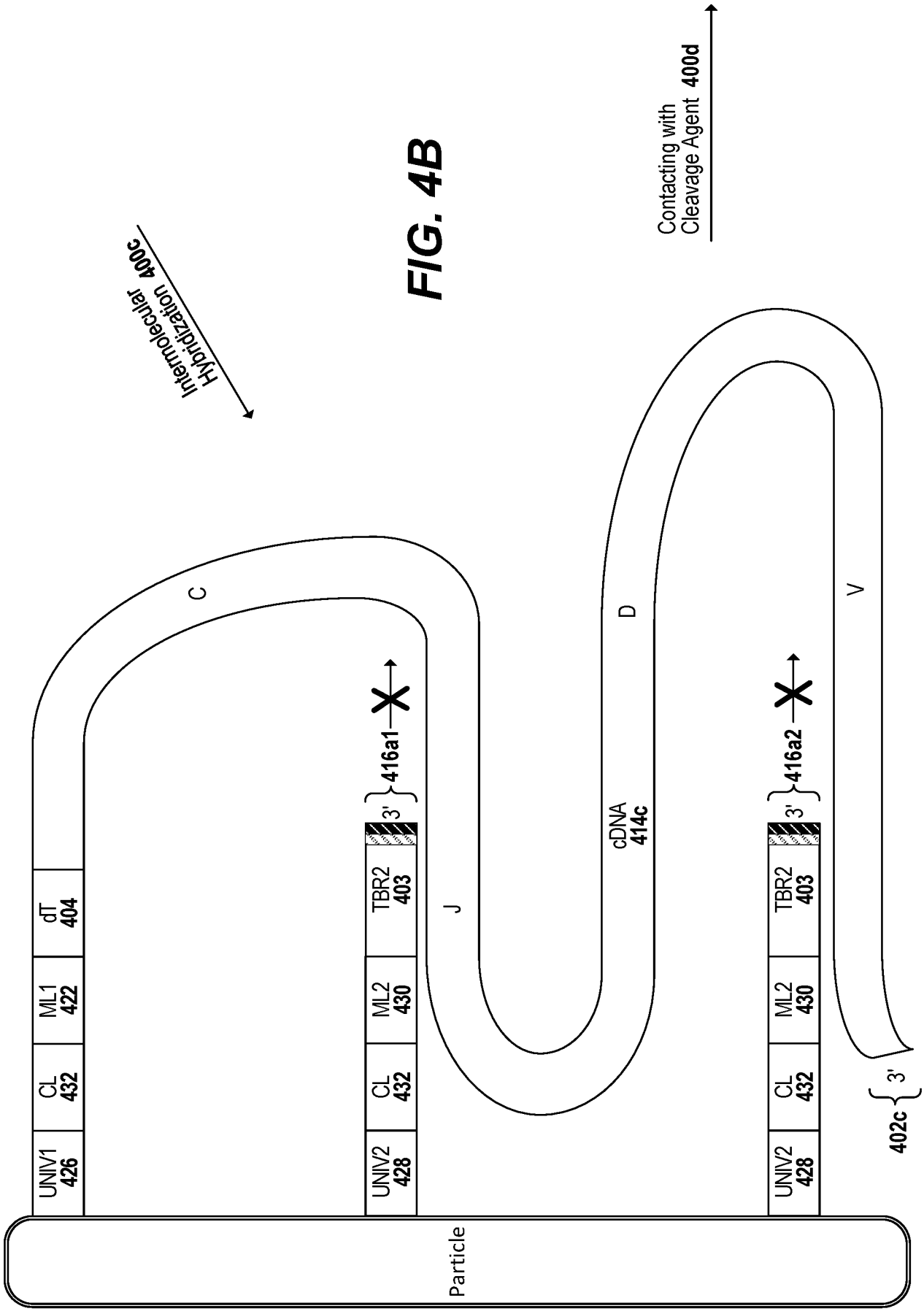

Intermolecular Hybridization 400c
Contacting with Cleavage Agent 400d
UNIV1 426
CL 432
ML1 422
dT 404
UNIV2 428
CL 432
ML2 430
TBR2 403
3'
416a1
C
J
cDNA 414c
D
UNIV2 428
CL 432
ML2 430
TBR2 403
3'
416a2
V
3'
402c
Particle

FIG. 4B

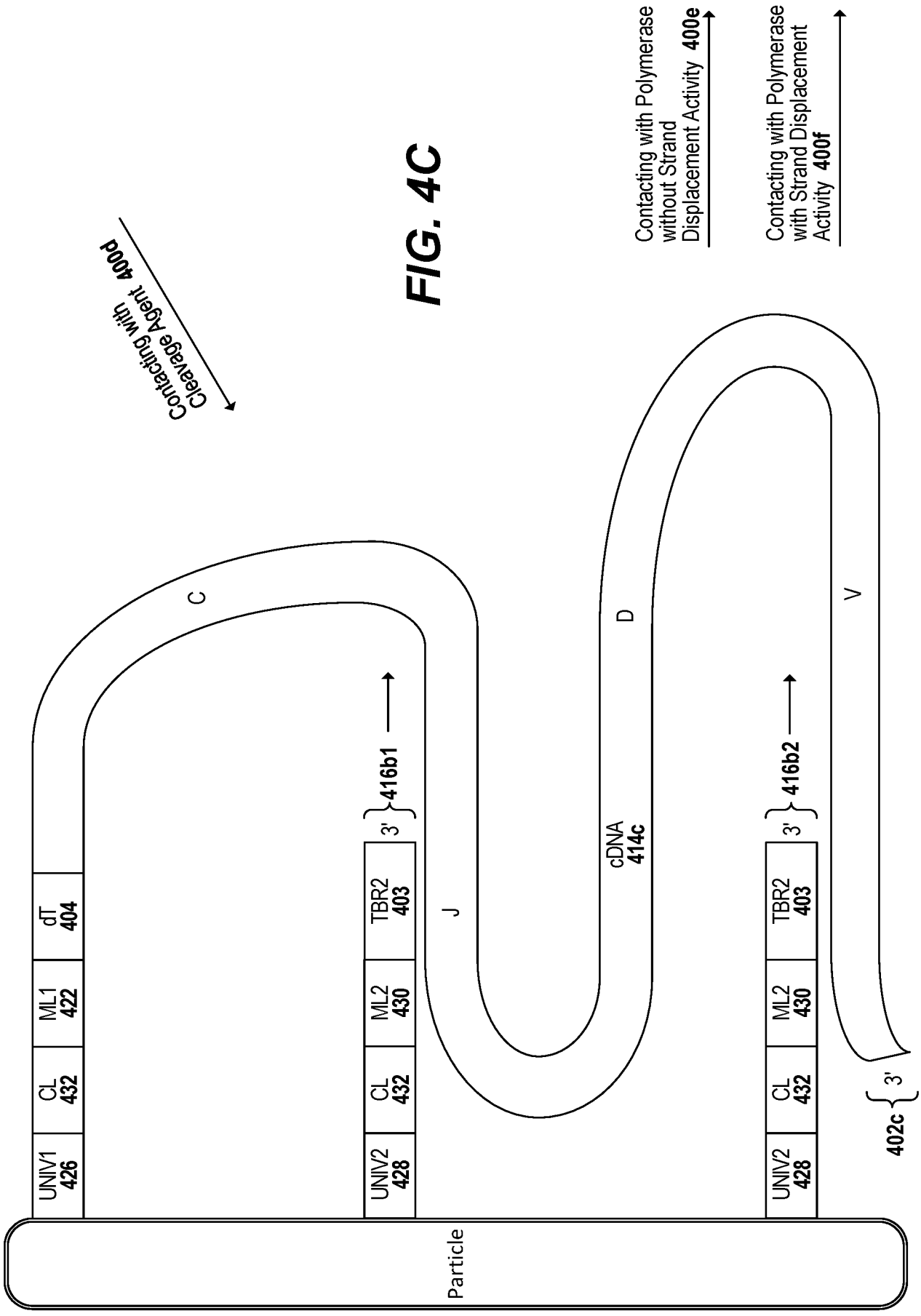

Particle
UNIV1 426
CL 432
ML1 422
dT 404
C
UNIV2 428
CL 432
ML2 430
TBR2 403
3'
416b1
J
cDNA 414c
D
UNIV2 428
CL 432
ML2 430
TBR2 403
3'
416b2
V
402c
3'
Contacting with Cleavage Agent 400d
Contacting with Polymerase without Strand Displacement Activity 400e
Contacting with Polymerase with Strand Displacement Activity 400f

FIG. 4C

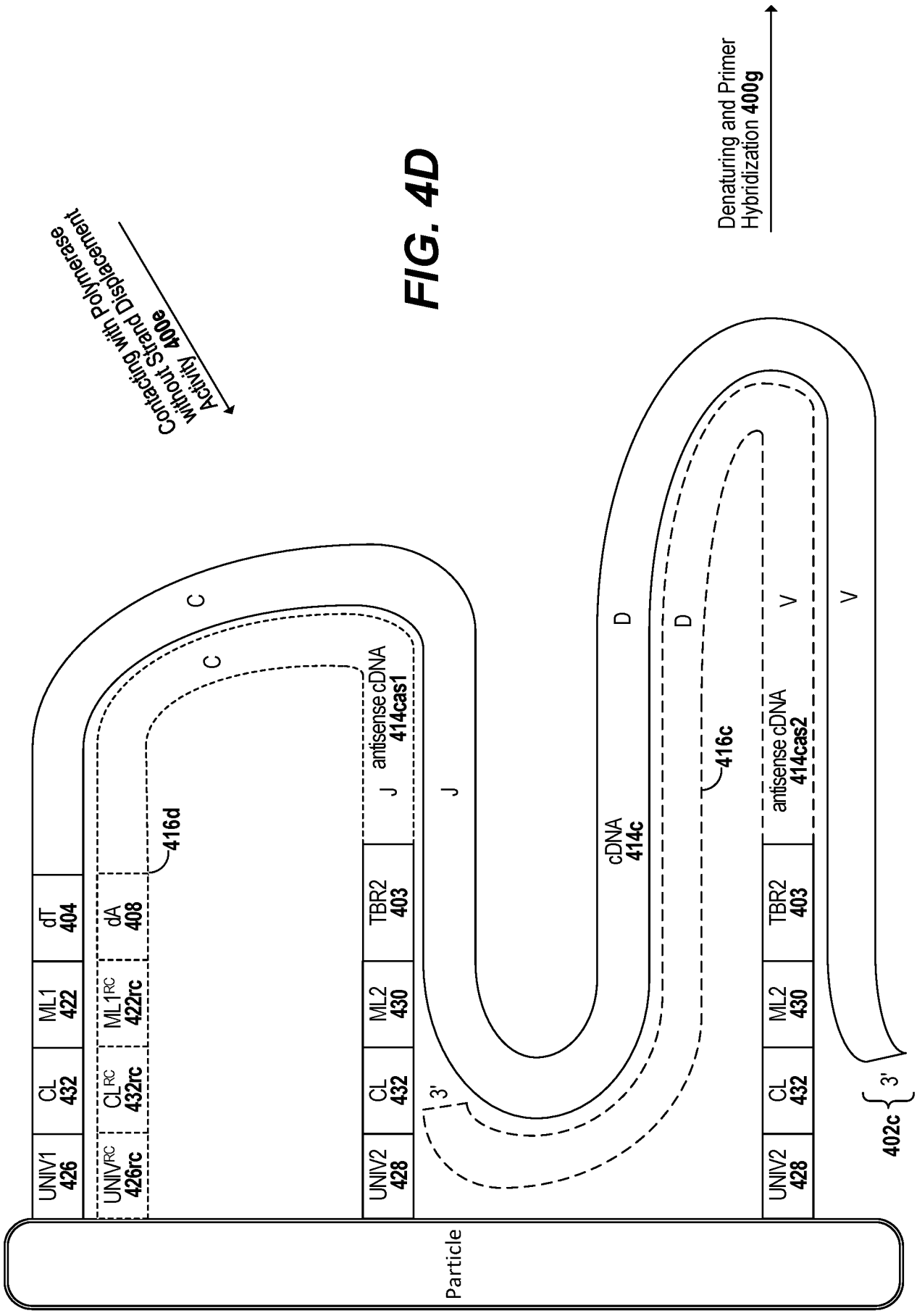

Particle
UNIV1 426
CL 432
ML1 422
dT 404
UNIV^RC 426rc
CL^RC 432rc
ML1^RC 422rc
dA 408
416d
UNIV2 428
CL 432
ML2 430
TBR2 403
J
antisense cDNA 414cas1
C
C
C
J
cDNA 414c
D
D
416c
antisense cDNA 414cas2
V
V
UNIV2 428
CL 432
ML2 430
TBR2 403
3'
402c
3'
Contacting with Polymerase Without Strand Displacement Activity 400e
Denaturing and Primer Hybridization 400g

FIG. 4D

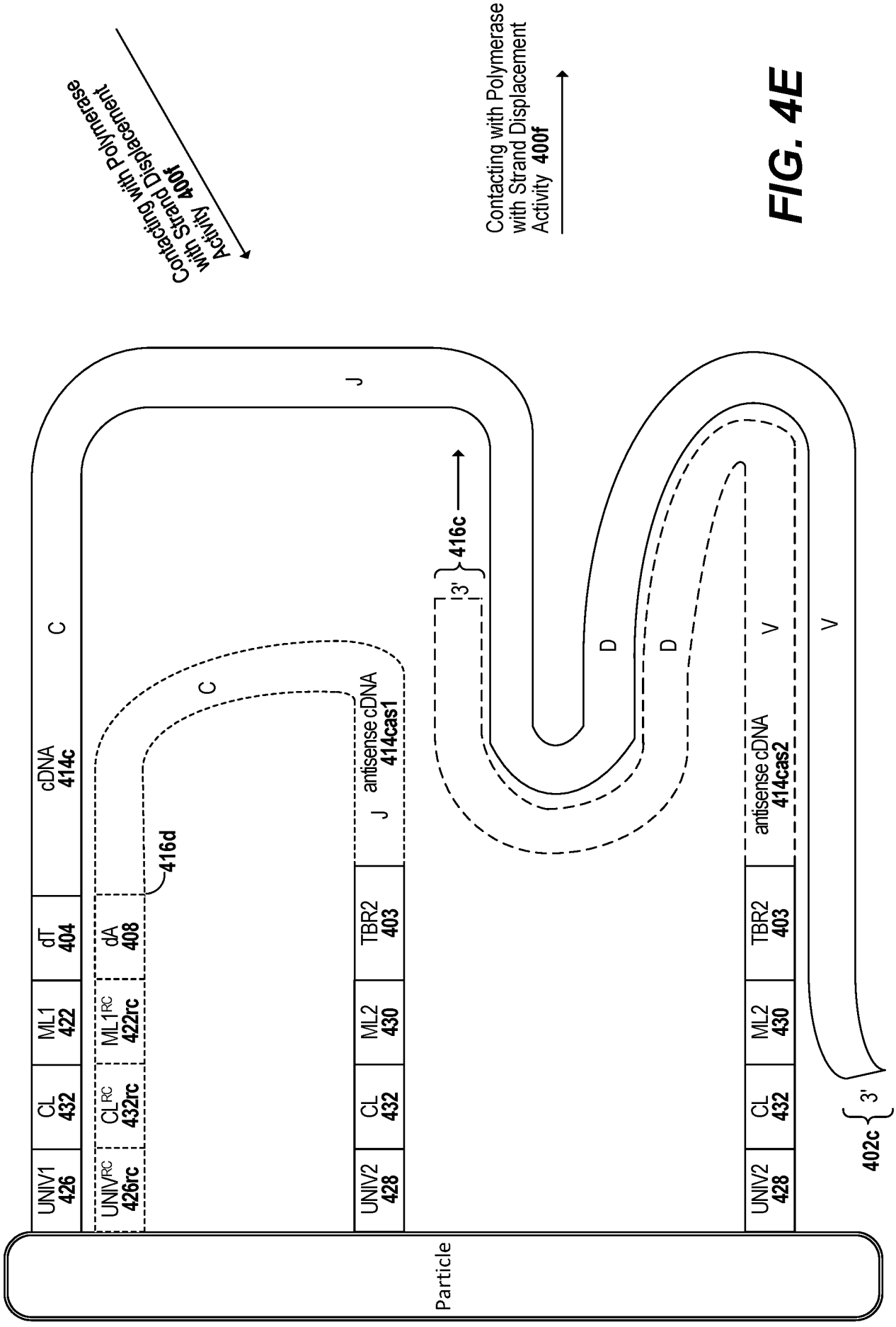

Contacting with Polymerase with Strand Displacement Activity 400f
Contacting with Polymerase with Strand Displacement Activity 400f
Particle
UNIV1 426
CL 432
ML1 422
dT 404
cDNA 414c
C
J
UNIV^{RC} 426rc
CL^{RC} 432rc
ML1^{RC} 422rc
dA 408
416d
C
J antisense cDNA 414cas1
UNIV2 428
CL 432
ML2 430
TBR2 403
3'
416c
D
D
V
V
antisense cDNA 414cas2
UNIV2 428
CL 432
ML2 430
TBR2 403
402c
3'

FIG. 4E

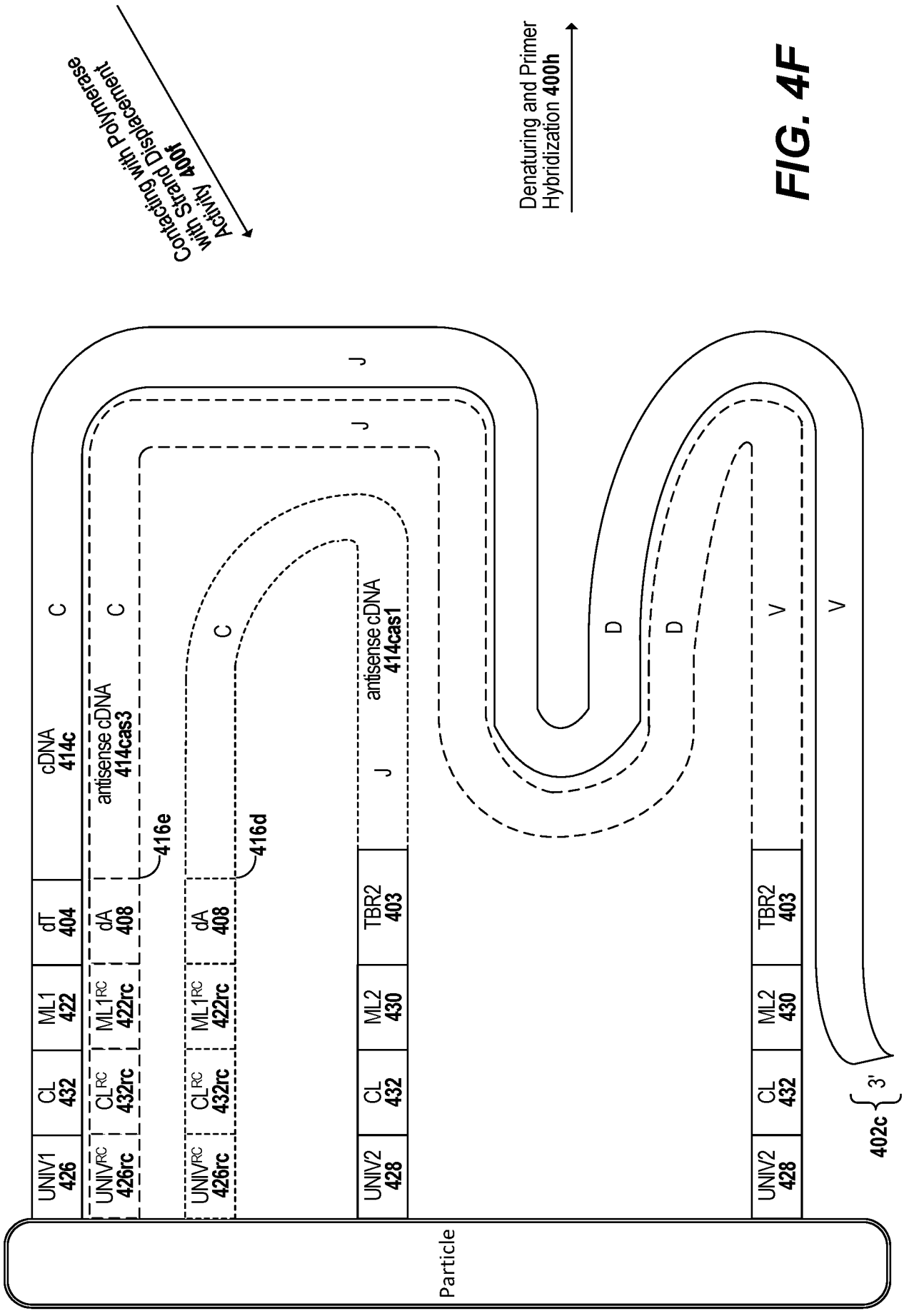

Contacting With Polymerase with Strand Displacement Activity 400f
Denaturing and Primer Hybridization 400h
Particle
UNIV1 426
CL 432
ML1 422
dT 404
cDNA 414c
C
UNIV^RC 426rc
CL^RC 432rc
ML1^RC 422rc
dA 408
antisense cDNA 414cas3
C
416e
416d
J
antisense cDNA 414cas1
TBR2 403
ML2 430
UNIV2 428
D
V
402c
3'

FIG. 4F

FIG. 4G

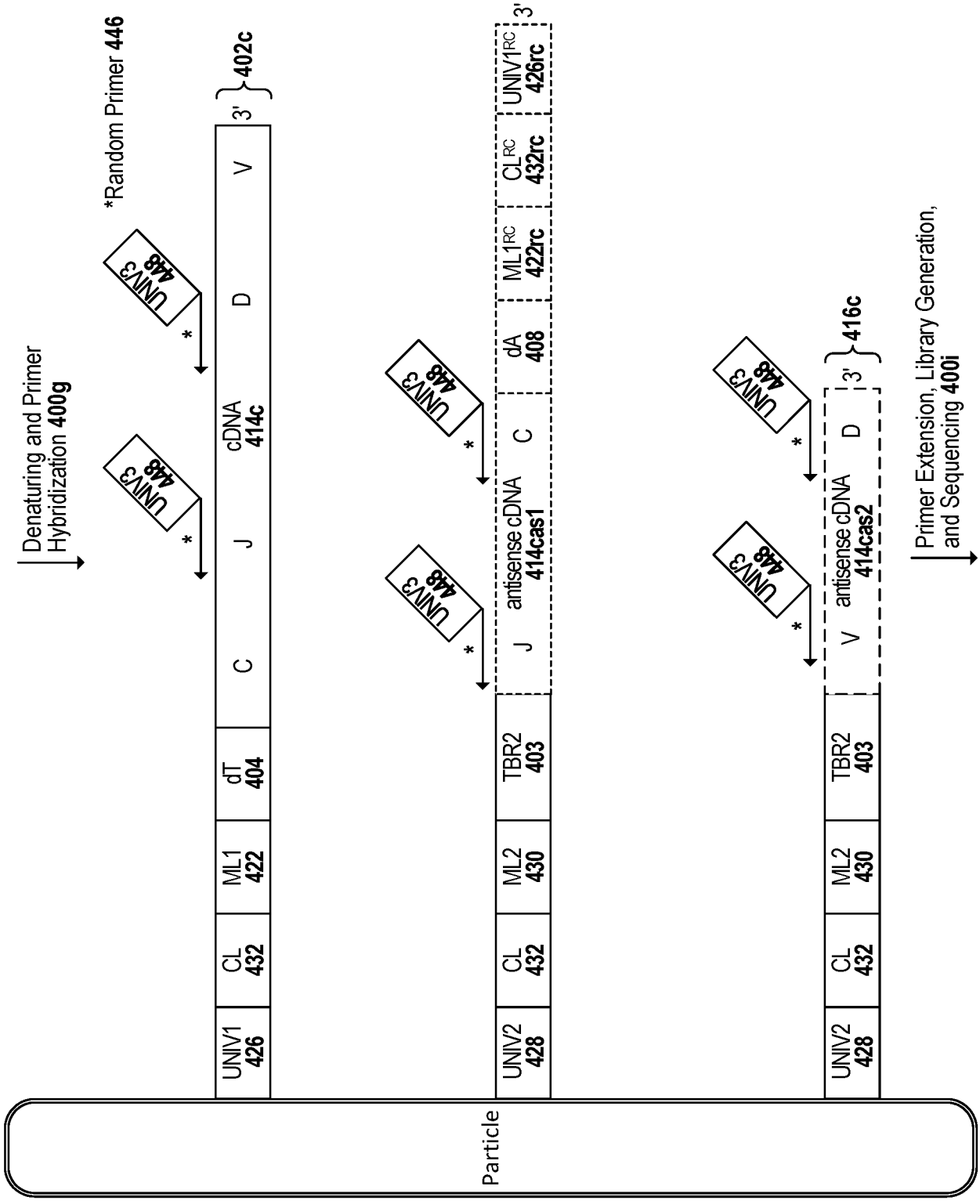
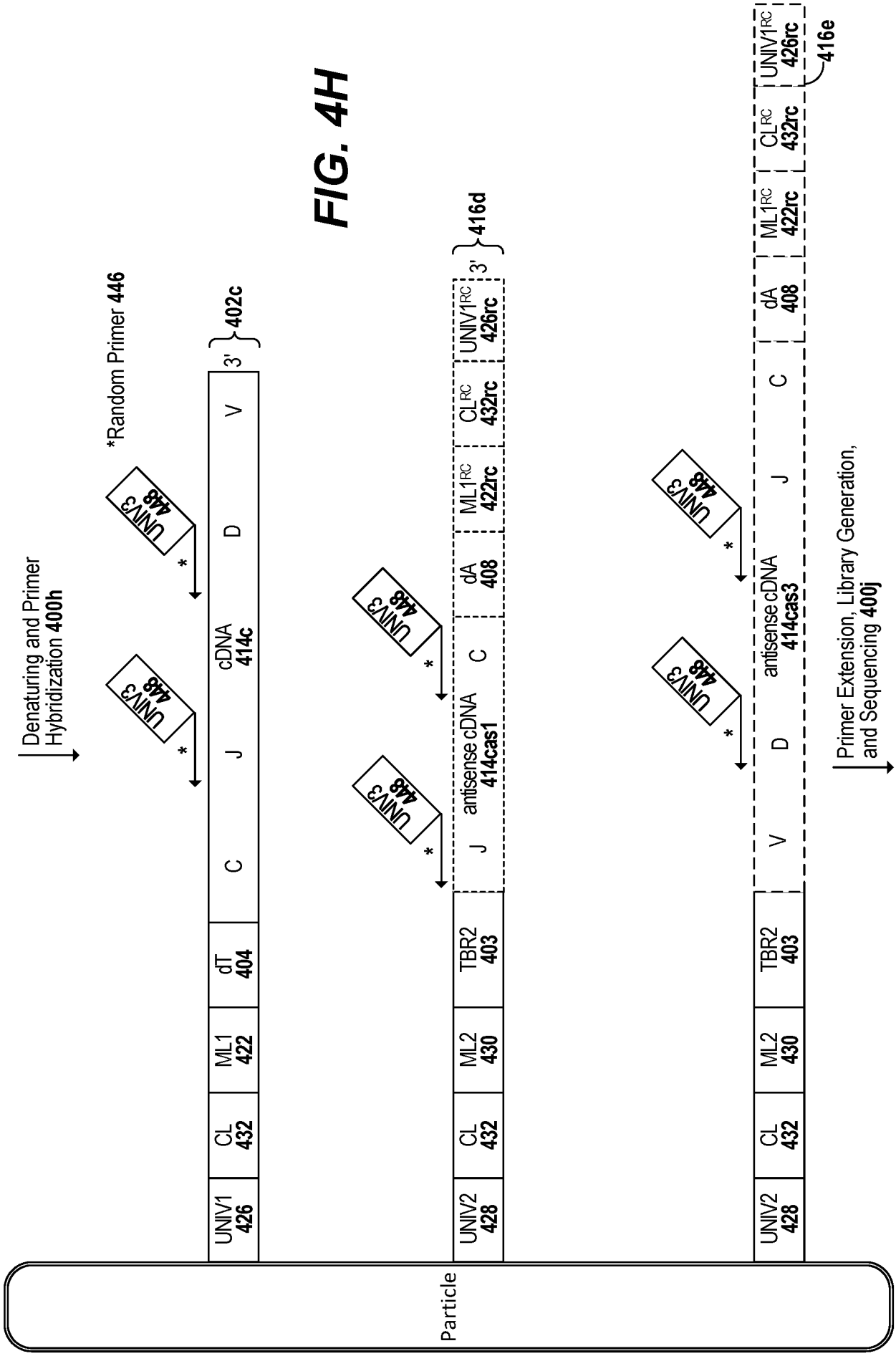

Denaturing and Primer Hybridization 400g
*Random Primer 446
UNIV3 448
UNIV1 426
CL 432
ML1 422
dT 404
C
J
cDNA 414c
D
V
3' 402c
Particle
UNIV2 428
CL 432
ML2 430
TBR2 403
J
antisense cDNA 414cas1
C
dA 408
ML1RC 422rc
CLRC 432rc
UNIV1RC 426rc
3' 416d
V
antisense cDNA 414cas2
D
3' 416c
Primer Extension, Library Generation, and Sequencing 400i

FIG. 4H


Denaturing and Primer Hybridization 400h
*Random Primer 446
UNIV3 448
UNIV1 426
CL 432
ML1 422
dT 404
C
J
cDNA 414c
D
V
3'
402c
Particle
UNIV2 428
CL 432
ML2 430
TBR2 403
J
antisense cDNA 414cas1
C
dA 408
ML1RC 422rc
CLRC 432rc
UNIV1RC 426rc
3'
416d
V
D
antisense cDNA 414cas3
J
C
dA 408
ML1RC 422rc
CLRC 432rc
UNIV1RC 426rc
3'
416e
Primer Extension, Library Generation, and Sequencing 400j

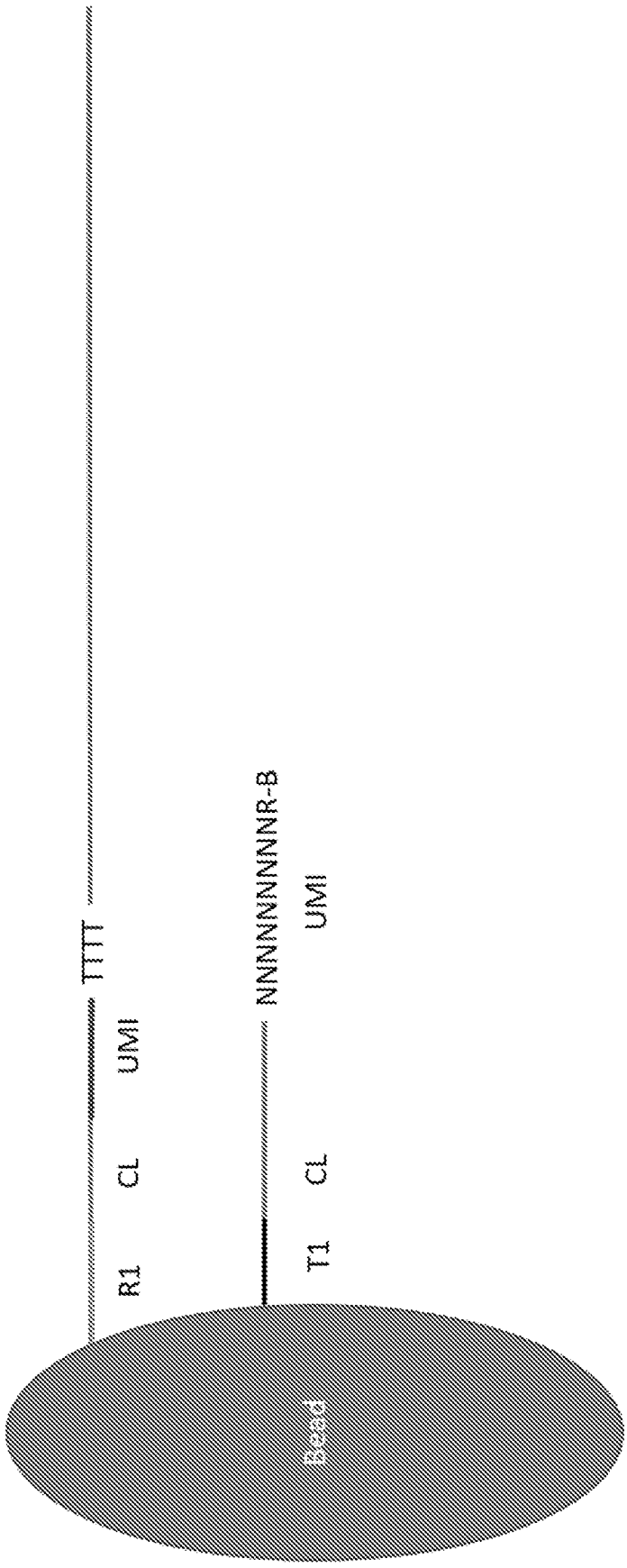


FIG. 5A

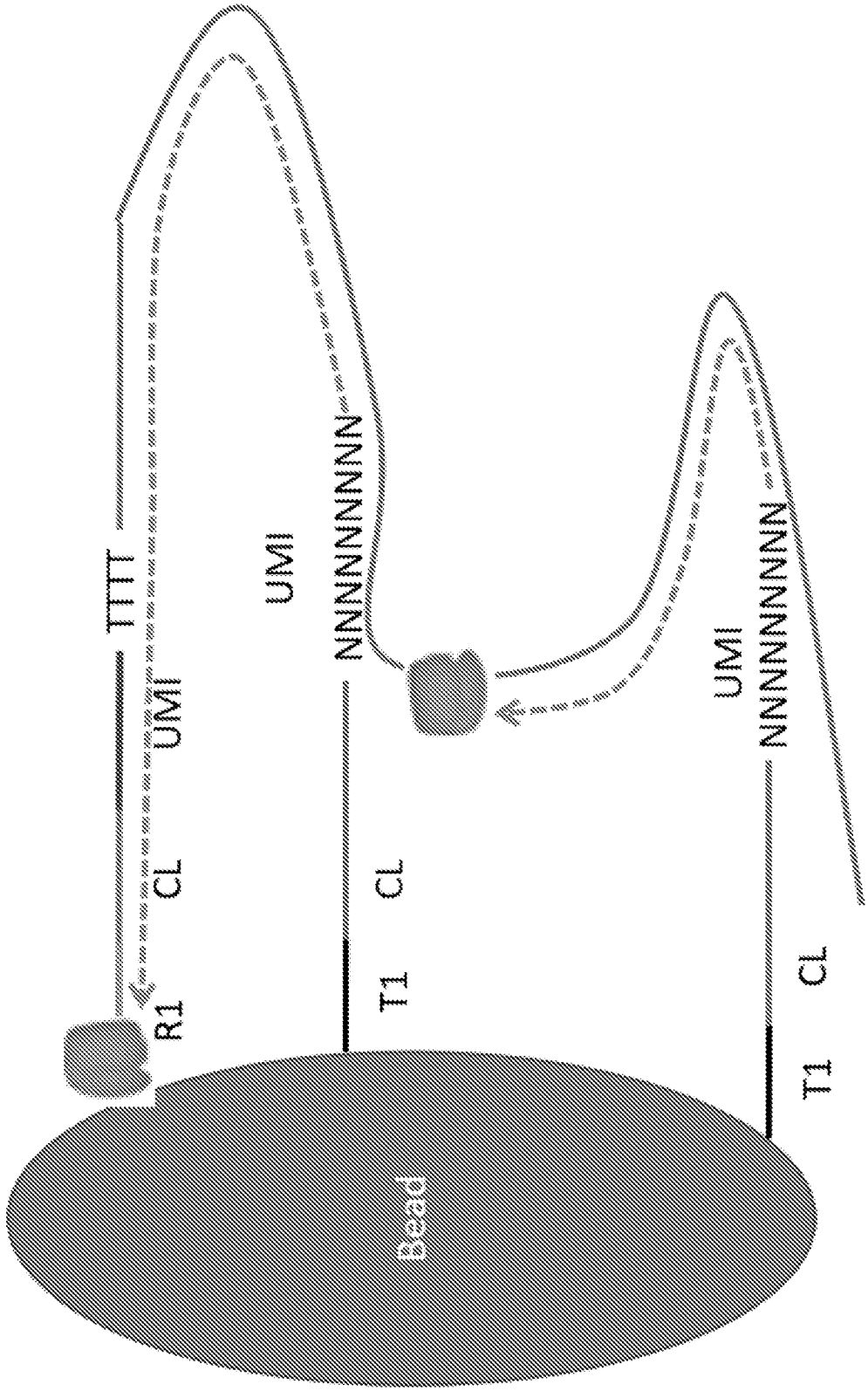


FIG. 5B

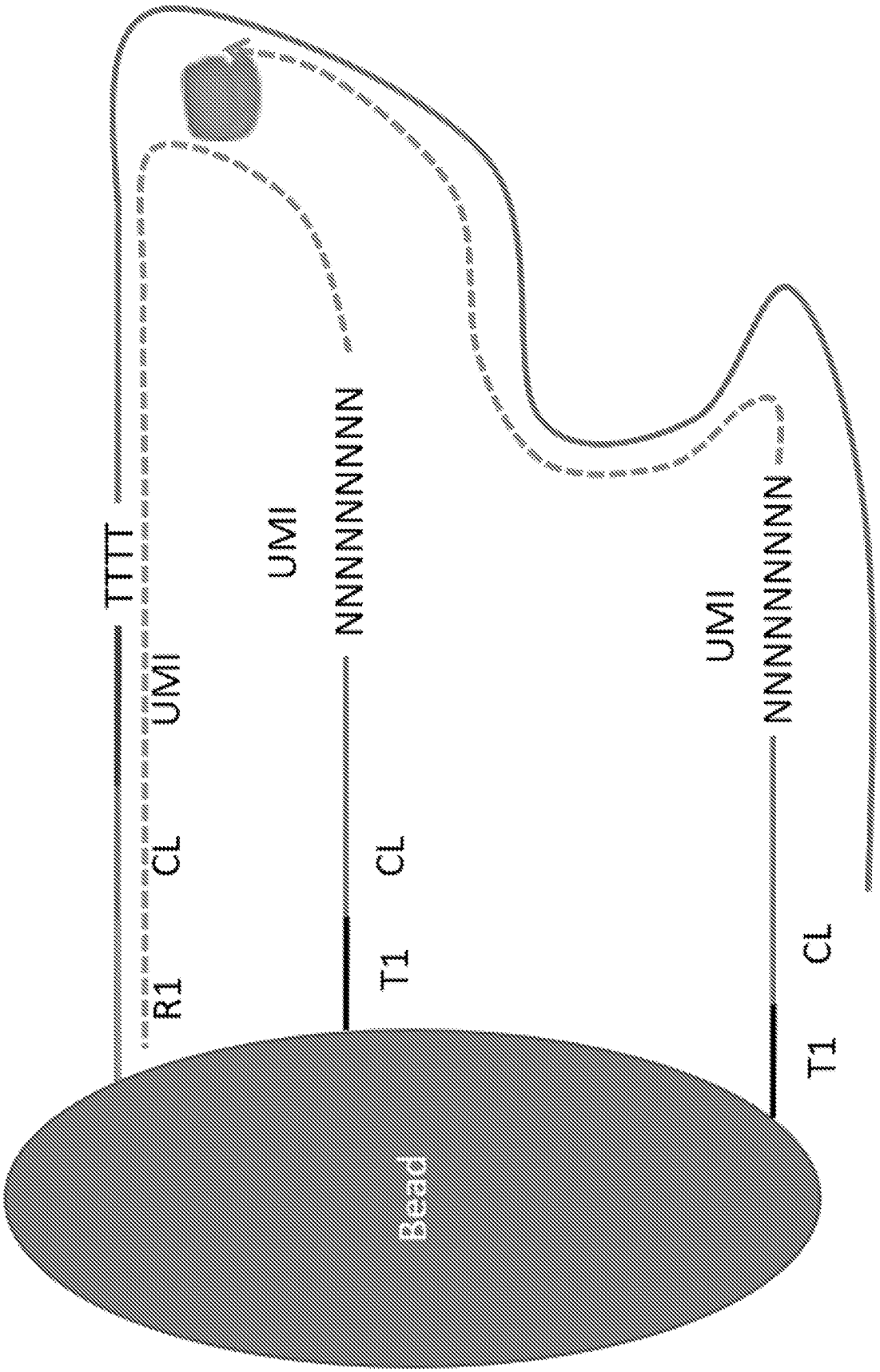


FIG. 5C

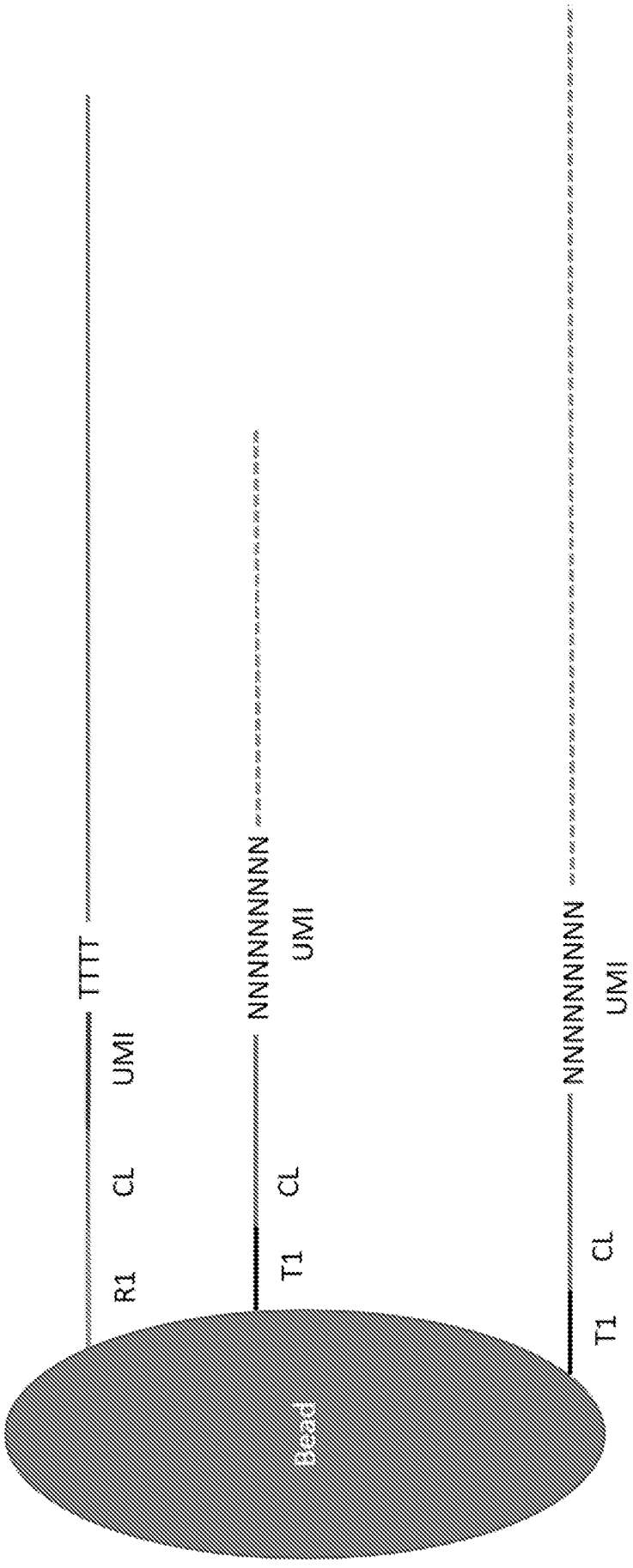


FIG. 5D

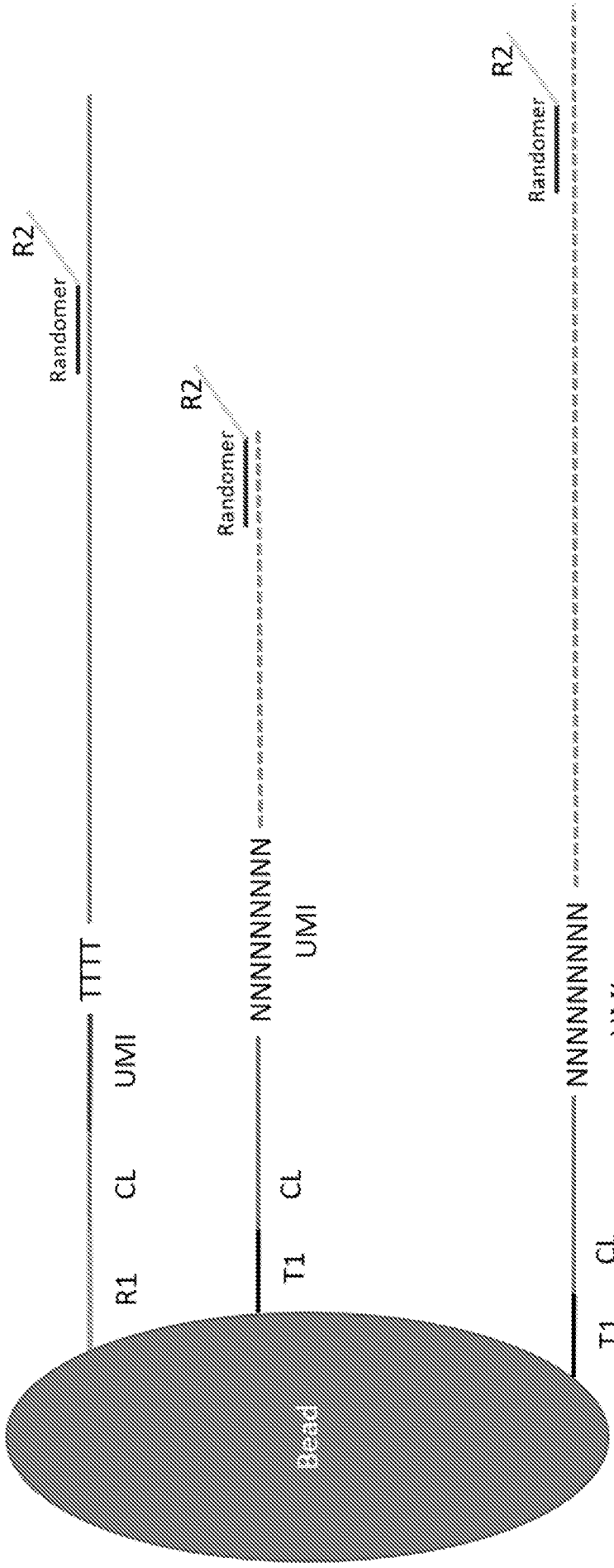

R2
Randomer
TTTT
UMI
CL
R1
bead
R2
Randomer
NNNNNNNN
UMI
CL
T1
R2
Randomer
NNNNNNNN
UMI
CL
T1

FIG. 5E

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2023/062070

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6806
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, CHEM ABS Data, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2021/139970 A1 (CHANG CHRISTINA [US] ET AL) 13 May 2021 (2021-05-13)	99
A	abstract; claims 1,13 paragraphs [0216], [0217]	1-98
A	US 2015/167060 A1 (CHUN JONG YOON [KR] ET AL) 18 June 2015 (2015-06-18) abstract; claims 1, 15	1-99

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 June 2023	23/06/2023

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Celler, Jakub

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2023/062070

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2021139970 A1	13-05-2021	CN 114729350 A	08-07-2022
		EP 4055160 A1	14-09-2022
		JP 2023500679 A	10-01-2023
		US 2021139970 A1	13-05-2021
		WO 2021092386 A1	14-05-2021
US 2015167060 A1	18-06-2015	EP 2839031 A1	25-02-2015
		JP 6096885 B2	15-03-2017
		JP 2015514415 A	21-05-2015
		KR 20140139535 A	05-12-2014
		US 2015167060 A1	18-06-2015
		WO 2013157821 A1	24-10-2013