



(51) International Patent Classification:

C12N 15/10 (2006.01) C40B 50/18 (2006.01)  
C12P 19/34 (2006.01) C40B 70/00 (2006.01)  
C40B 40/06 (2006.01)

(21) International Application Number:

PCT/US2017/067387

(22) International Filing Date:

19 December 2017 (19.12.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/436,288 19 December 2016 (19.12.2016) US

(71) Applicant: **BIO-RAD LABORATORIES, INC.**  
[US/US]; 1000 Alfred Nobel Drive, Hercules, California  
94547 (US).

(72) Inventors: **LEBOFSKY, Ronald**; c/o Bio-Rad Laborato-  
ries, Inc., 1000 Alfred Nobel Drive, Hercules, California  
94547-1803 (US). **AGRESTI, Jeremy**; c/o Bio-Rad Labo-  
ratories, Inc., 1000 Alfred Nobel Drive, Hercules, Califor-  
nia 94547-1803 (US).

(74) Agent: **HINSCH, Matthew E.** et al.; Kilpatrick  
Townsend and Stockton LLP, Mailstop: IP Docketing - 22,  
1100 Peachtree Street, Suite 2800, Atlanta, Georgia 30309  
(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, CZ, DE, DJ, DK, DM, DO,  
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,  
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,  
KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME,  
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,  
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,  
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,  
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,  
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,  
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,  
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,  
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,

(54) Title: DROPLET TAGGING CONTIGUITY PRESERVED TAGMENTED DNA

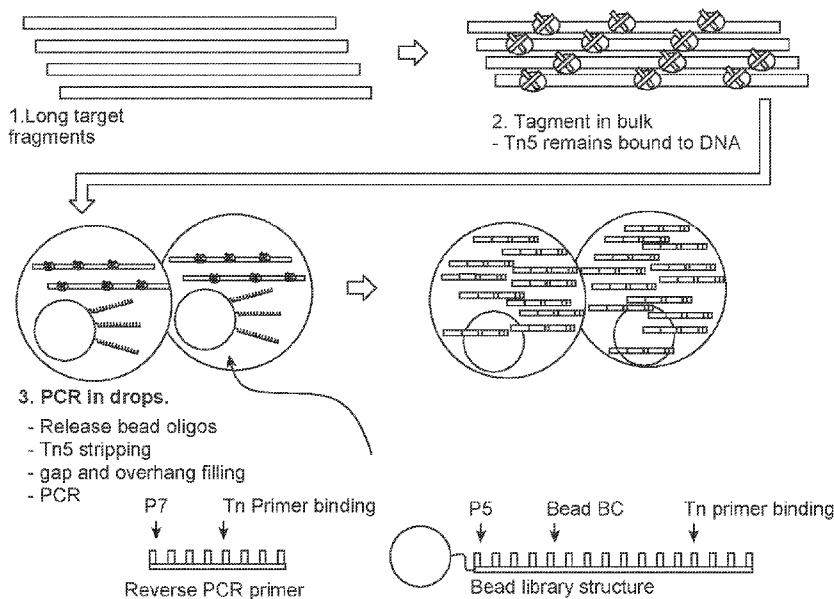


FIG. 1

(57) Abstract: Methods and compositions for maintaining DNA contiguity for sequencing is provided.

WO 2018/118971 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,  
KM, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

**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))*

## Droplet tagging contiguity preserved tagmented DNA

### 5 CROSS-REFERENCE TO RELATED PATENT APPLICATIONS

[0001] The present application claims benefit of priority to US Provisional Patent Application No. 62/436,288, filed December 19, 2016, which is incorporated by reference.

### BACKGROUND OF THE INVENTION

10 [0002] Haplotype information can be valuable in many genetic analyses. However, it can be difficult to obtain information regarding haplotypes from many sequencing methods because contiguity is not maintained. Amini *et al. Nature Genetics* 46(12):1343-1349 describes one method of maintaining contiguity, but the method involves many separate reactions, each of which require significant enzyme (e.g., tagmentase).

### 15 BRIEF SUMMARY OF THE INVENTION

[0003] In some embodiments, a method of determining haplotype genomic sequence is provided. In some embodiments, the method comprises providing fragments of genomic DNA;

20 reacting the fragments with an adapter-loaded tagmentase that creates DNA fragments defined by breakpoints in the fragments and inserts an adaptor at the break points, wherein the reacting is under conditions such that the tagmentase binds the breakpoints to form linked DNA segments in the form of DNA segment-first adaptor-tagmentase-second adaptor-(DNA segment-first adaptor-tagmentase-second adaptor)*n*-DNA segment, where *n* is any integer and “-” indicates a covalent linkage;

25 encapsulating the linked DNA segments into partitions, said partitions comprising:

a bead, the bead having a forward primer oligonucleotide linked to the bead by a 5' end of the forward primer oligonucleotide, the forward primer oligonucleotide having a bead-specific barcode and a 3' end specific for and complementary to the first or second adaptor;

a reverse primer oligonucleotide having a 3' end complementary to the first or second adaptor, wherein the forward primer 3' end and the reverse primer 3' end are complementary to different adaptors selected from the first adaptor and the second adaptor;

5 displacing the tagmentase from the segments in the partitions; performing amplification wherein the forward primer and reverse primer oligonucleotide generate amplicons from the DNA segments, such that amplicons within a partition are barcoded with the bead barcode; combining the partitions to form a reaction mixture containing the amplicons; and nucleotide sequencing the amplicons.

10 [0004] In some embodiments, the partitions comprise an amount of agent that displaces the tagmentase from the segments but does not inhibit polymerization. In some embodiments, the agent is a polymerase. In some embodiments, the agent is a detergent. In some embodiments, heat displaces the tagmentase.

15 [0005] In some embodiments combined with any embodiment listed above, prior to amplification, the single stranded regions of inserted adaptors are filled in by DNA polymerase. In some embodiments combined with any embodiment listed above, prior to amplification, single stranded regions of fragmented target nucleic acids are filled in by DNA polymerase combined with any embodiment listed above. In some embodiments, the forward primer oligonucleotide is released from the bead and amplification occurs in solution.

20 [0006] In some embodiments combined with any embodiment listed above, the agent is a detergent. In some embodiments combined with any embodiment listed above, the detergent is sodium dodecyl sulfate (SDS). In some embodiments combined with any embodiment listed above, the concentration of SDS is 0.005-0.05% (e.g., 0.01-0.04%, e.g., 0.01-0.02%).

25 [0007] In some embodiments combined with any embodiment listed above, the fragments are on average between 5-10 Mb.

[0008] In some embodiments combined with any embodiment listed above, the partitions are droplets in an emulsion.

[0009] In some embodiments combined with any embodiment listed above, the encapsulating encapsulates on average 0.02-3 (e.g., 0.05-1, 0.08-0.5, e.g., 0.1, 1, 2, or 3) beads into partitions.

[0010] In some embodiments combined with any embodiment listed above, the genomic DNA is from a single cell. In some embodiments combined with any embodiment listed above, the genomic DNA is from a mammal or plant.

[0011] In some embodiments combined with any embodiment listed above, the first adaptor and the second adaptor have different sequences. In some embodiments combined with any embodiment listed above, the first adaptor and the second adaptor have 5' overhang sequences and the 5' overhang sequences are less than 50% identical.

[0012] In some embodiments combined with any embodiment listed above, the first adaptor and the second adaptor have identical sequences.

[0013] In some embodiments combined with any embodiment listed above, the first adaptor and the second adaptor are linked by a linking sequence when loaded on the tagmentase such the DNA segments are linked by both the tagmentase and the linking sequence. In some embodiments combined with any embodiment listed above, the linking sequence comprises a restriction recognition sequence and the linking sequence is cleaved by a restriction enzyme after the encapsulating and before the performing. In some embodiments combined with any embodiment listed above, the linking sequence comprises one or more uracils and the linking sequence is cleaved by a uracil-DNA N-glycosylase after the encapsulating and before the performing. In some embodiments combined with any embodiment listed above, the linking sequence comprises one or more ribonucleotide and the linking sequence is cleaved at the ribonucleotide after the encapsulating and before the performing.

[0014] In some embodiments, combined with any embodiment listed above n is an integer selected from 0-10,000.

[0015] In some embodiments combined with any embodiment listed above, the length of the DNA segment-first adaptor-tagmentase-second adaptor-(DNA segment-first adaptor-tagmentase-second adaptor)<sub>n</sub>-DNA segment is between 5 kb-10Mb, e.g., 1Mb-10Mb.

[0016] In some embodiments combined with any embodiment listed above, at least 10,000 different linked DNA segments are encapsulated into different partitions.

[0017] In some embodiments combined with any embodiment listed above, the tagmentase is linked to a solid support. For example, in some embodiments the solid support is a bead that is  
5 different from the bead linked to the forward primer.

[0018] Also provided is a plurality of partitions. In some embodiments, the partitions comprise:

a bead, the bead having a forward primer oligonucleotide linked to the bead by a 5' end of the forward primer oligonucleotide, the forward primer oligonucleotide having a bead-specific  
10 barcode and a 3' end specific for and complementary to a first or second adaptor; and

a reverse primer oligonucleotide having a 3' end complementary to the first or second adaptor, wherein the forward primer 3' end and the reverse primer 3' end are complementary to different adaptors selected from the first adaptor and the second adaptor.

[0019] In some embodiments, the partitions comprise an amount of agent that displaces  
15 tagmentase from DNA but does not inhibit polymerization.

[0020] In some embodiments, the partitions further comprise: tagmentase; and genomic DNA, wherein segments of the DNA have ends defined by adaptors inserted at breakpoints created by the tagmentase in the form first adaptor-DNA segment- second adaptor.

[0021] In some embodiments, the agent is a detergent. In some embodiments, the detergent is  
20 sodium dodecyl sulfate (SDS). In some embodiments, the concentration of SDS is 0.005-0.05% (e.g., 0.01-0.04%, e.g., 0.01-0.02%).

[0022] In some embodiments, the partitions are droplets in an emulsion.

[0023] In some embodiments, the partitions comprise on average 0.02-3 (e.g., 0.05-1, 0.08-0.5, e.g., 0.1, 1, 2, or 3) beads.

[0024] In some embodiments, the genomic DNA is from a single cell. In some embodiments,  
25 the genomic DNA is from a mammal or plant.

[0025] In some embodiments, the first adaptor and the second adaptor have different sequences. In some embodiments, the first adaptor and the second adaptor are less than 50% identical.

5 [0026] In some embodiments, the first adaptor and the second adaptor have identical sequences.

[0027] In some embodiments, the first adaptor and the second adaptor are linked by a linking sequence in a form comprising first adaptor-DNA segment-second adaptor-(linker sequence)-first adaptor-DNA segment-second adaptor)<sub>n</sub>, where n is any integer 1 or greater.

10 [0028] In some embodiments, the linking sequence comprises a restriction recognition sequence and optionally the partition further comprises a restriction enzyme the restriction recognition sequence.

[0029] In some embodiments, the linking sequence comprises one or more uracils and optionally the partition further comprises a uracil-DNA N-glycosylase.

[0030] In some embodiments, the linking sequence comprises one or more ribonucleotide.

15 [0031] In some embodiments, n is an integer selected from 1-10,000.

[0032] In some embodiments, the length of the first adaptor-DNA segment-second adaptor-(linker sequence-first adaptor-DNA segment-second adaptor)<sub>n</sub> is between 5 kb-10Mb, e.g., 1Mb-10Mb.

[0033] In some embodiments, the plurality comprises at least 10,000 different partitions.

20 [0034] In some embodiments, the tagmentase is linked to a solid support. For example, in some embodiments the solid support is a bead that is different from the bead linked to the forward primer.

25 [0035] Also provided is a plurality of partitions comprising DNA segments having contiguity maintained by tagmentase. In some embodiments, the partitions comprise linked DNA segments in the form of DNA segment-first adaptor-tagmentase-second adaptor-(DNA segment-first adaptor-tagmentase-second adaptor)<sub>n</sub>-DNA segment, where n is any integer and “-” indicates a covalent linkage and wherein the linked DNA maintains contiguity compared to genomic DNA.

[0036] In some embodiments, the partitions are droplets in an emulsion.

[0037] In some embodiments, the partitions comprise on average 0.02-3 (e.g., 0.05-1, 0.08-0.5, e.g., 0.1, 1, 2, or 3) beads.

[0038] In some embodiments, the genomic DNA is from a single cell. In some embodiments,  
5 the genomic DNA is from a mammal or plant.

[0039] In some embodiments, the first adaptor and the second adaptor have different sequences. In some embodiments, the first adaptor and the second adaptor are less than 50% identical.

[0040] In some embodiments, the first adaptor and the second adaptor have identical  
10 sequences.

[0041] In some embodiments, the linking sequence comprises a restriction recognition sequence and optionally the partition further comprises a restriction enzyme the restriction recognition sequence.

[0042] In some embodiments, the linking sequence comprises one or more uracils and  
15 optionally the partition further comprises a uracil-DNA N-glycosylase.

[0043] In some embodiments, the linking sequence comprises one or more ribonucleotide.

[0044] In some embodiments, n is an integer selected from 1-10,000.

[0045] In some embodiments, the length of the first adaptor-DNA segment-second adaptor-  
(linker sequence-first adaptor-DNA segment-second adaptor)<sup>n</sup> is between 5 kb-10Mb, e.g., 1Mb-  
20 10Mb.

[0046] In some embodiments, the plurality comprise at least 10,000 different partitions.

#### BRIEF DESCRIPTION OF THE DRAWINGS

25 [0047] FIG. 1 illustrates the first part of a possible workflow. DNA fragments are reacted with adaptor-loaded tagmentase, encapsulated in a partition with a bead linked to a barcoded primer and second primer and amplified in the partition.



- [0048] FIG. 2 illustrates the second part of a possible workflow (continued from FIG. 1). Amplicons from the partitions are mixed (for example, if droplets in an emulsion are used, the emulsion is broken and the aqueous droplets pooled). The amplicons are then sequenced. FIG. 2 displays several of many configurations of primers for sequencing.
- 5 [0049] FIG. 3 depicts illustrates one possible for generating bead-linked and barcoded oligonucleotides as can be used in the methods described herein.
- [0050] FIG. 4 is a continuation of FIG. 3.
- [0051] FIG. 5 depicts exemplary tagmentase (Tn5) adaptors and amplification primers that can be used to amplify DNA segments having the adaptors.
- 10 [0052] FIG. 6 illustrates various possible non-limiting options for the bead-linked oligonucleotide primer.
- [0053] FIG. 7 illustrates additional possible non-limiting options for the bead-linked oligonucleotide primer.
- [0054] FIG. 8 depicts a possible option for using heteroadaptor-loaded tagmentase.
- 15 [0055] FIG. 9 is a continuation of FIG. 8.
- [0056] FIG. 10 depicts a possible option (different from FIG. 8) for using heteroadaptor-loaded tagmentase.
- [0057] FIG. 11 depicts a possible option (different from FIG. 8 or 10) for using heteroadaptor-loaded tagmentase.
- 20 [0058] FIG. 12 depicts a possible option (different from FIG. 8, 10, or 11) for using heteroadaptor-loaded tagmentase.
- [0059] FIG. 13 depicts a possible option for using homoadaptor-loaded tagmentase.
- [0060] FIG. 14 depicts some possible options for primer sequences when using homoadaptor-loaded tagmentase.
- 25 [0061] FIG. 15 depicts a possible option for using homoadaptor-loaded tagmentase.
- [0062] FIG. 16 is a continuation of FIG. 15.

[0063] FIG. 17 is a continuation of FIG. 15 and illustrates how one can use PCR suppression to preferentially inhibit homo-tailed amplicons produced by polymerase extensions within the context of using a homoadaptered Tn5 transposase.

5 [0064] FIG. 18 depicts a possible option (different from FIG. 15) for using homoadaptor-loaded tagmentase.

[0065] FIG. 19 is a continuation of FIG. 18.

[0066] FIG. 20 is a continuation of FIG. 18 and illustrates how one can use PCR suppression to preferentially inhibit homo-tailed amplicons produced by polymerase extensions within the context of using a homoadaptered Tn5 transposase.

10 [0067] FIG. 21 depicts possible embodiments of the method as well as illustrating the mechanism of Tn5 action.

[0068] FIG. 22 depicts an option in which the two adaptors loaded on the tagmentase are linked by a linking sequence. The linking sequence is included with the adaptors inserted by the tagmentase and thus the linking sequences function to maintain contiguity of DNA segments.

15 [0069] FIG. 23 is a continuation of FIG. 22 and depicts various possible options for cleaving the linking sequences, thereby releasing separate DNA segments into the partition.

[0070] FIG. 24 depicts an exemplary adaptor for use as a homoadaptor.

[0071] FIG. 25 illustrates possible oligonucleotide primers for use with the homoadaptor depicted in FIG. 24.

20 [0072] FIG. 26 is a continuation of FIG. 25 and depicts how the primers can be used.

[0073] FIG. 27 is a continuation of FIG. 26.

[0074] FIG. 28 is a continuation of FIG. 25 and depicts an alternative for how the primers can be used.

[0075] FIG. 29 is a continuation of FIG. 28.

[0076] Figure 30A-C: Experiment demonstrating CPTseq in droplets. Figure 30A depicts the workflow. Figure 30B depicts accumulation of product. Figure 30C depicts a cartoon sequencing read and QC results.

5 [0077] Figure 31: Expected whole genome analysis and phasing data from the data sets produced by the sequencing run shown in Figure 30A-C.

[0078] Figure 32: Illustration of a ddSEQ chip or cartridge with indications of where to pipette the DNA and enzyme mix. The oil is pipetted in the row2 on the bottom. The DNA and enzyme mix inlets are primed first as indicated. The output wells where the emulsion results is indicated in the top row.

10

#### DEFINITIONS

[0079] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Generally, the nomenclature used herein and the laboratory procedures in cell  
15 culture, molecular genetics, organic chemistry, and nucleic acid chemistry and hybridization described below are those well-known and commonly employed in the art. Standard techniques are used for nucleic acid and peptide synthesis. The techniques and procedures are generally performed according to conventional methods in the art and various general references (see generally, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, 2d ed.  
20 (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is incorporated herein by reference), which are provided throughout this document. The nomenclature used herein and the laboratory procedures in analytical chemistry, and organic synthetic described below are those well-known and commonly employed in the art.

[0080] The term "amplification reaction" refers to any *in vitro* means for multiplying the  
25 copies of a target sequence of nucleic acid in a linear or exponential manner. Such methods include but are not limited to polymerase chain reaction (PCR); DNA ligase chain reaction (see U.S. Pat. Nos. 4,683,195 and 4,683,202; PCR Protocols: A Guide to Methods and Applications (Innis *et al.*, eds, 1990)) (LCR); QBeta RNA replicase and RNA transcription-based amplification reactions (*e.g.*, amplification that involves T7, T3, or SP6 primed RNA

polymerization), such as the transcription amplification system (TAS), nucleic acid sequence based amplification (NASBA), and self-sustained sequence replication (3SR), isothermal amplification reactions (*e.g.*, single-primer isothermal amplification (SPIA)); as well as others known to those of skill in the art.

5 [0081] "Amplifying" refers to a step of submitting a solution to conditions sufficient to allow for amplification of a polynucleotide if all of the components of the reaction are intact. Components of an amplification reaction include, *e.g.*, primers, a polynucleotide template, polymerase, nucleotides, and the like. The term "amplifying" typically refers to an "exponential" increase in target nucleic acid. However, "amplifying" as used herein can also refer to linear increases in the numbers of a select target sequence of nucleic acid, such as is obtained with  
10 cycle sequencing or linear amplification. In an exemplary embodiment, amplifying refers to PCR amplification using a first and a second amplification primer.

[0082] The term "amplification reaction mixture" refers to an aqueous solution comprising the various reagents used to amplify a target nucleic acid. These include enzymes, aqueous buffers,  
15 salts, amplification primers, target nucleic acid, and nucleoside triphosphates. Amplification reaction mixtures may also further include stabilizers and other additives to optimize efficiency and specificity. Depending upon the context, the mixture can be either a complete or incomplete amplification reaction mixture

[0083] "Polymerase chain reaction" or "PCR" refers to a method whereby a specific segment  
20 or subsequence of a target double-stranded DNA, is amplified in a geometric progression. PCR is well known to those of skill in the art; see, *e.g.*, U.S. Pat. Nos. 4,683,195 and 4,683,202; and PCR Protocols: A Guide to Methods and Applications, Innis *et al.*, eds, 1990. Exemplary PCR reaction conditions typically comprise either two or three step cycles. Two step cycles have a denaturation step followed by a hybridization/elongation step. Three step cycles comprise a  
25 denaturation step followed by a hybridization step followed by a separate elongation step.

[0084] A "primer" refers to a polynucleotide sequence that hybridizes to a sequence on a target nucleic acid and serves as a point of initiation of nucleic acid synthesis. Primers can be of a variety of lengths and are often less than 50 nucleotides in length, for example 12-30 nucleotides, in length. The length and sequences of primers for use in PCR can be designed based on  
30 principles known to those of skill in the art, see, *e.g.*, Innis *et al.*, *supra*. Primers can be DNA,

RNA, or a chimera of DNA and RNA portions. In some cases, primers can include one or more modified or non-natural nucleotide bases. In some cases, primers are labeled.

[0085] A nucleic acid, or a portion thereof, "hybridizes" to another nucleic acid under conditions such that non-specific hybridization is minimal at a defined temperature in a physiological buffer (*e.g.*, pH 6-9, 25-150 mM chloride salt). In some cases, a nucleic acid, or portion thereof, hybridizes to a conserved sequence shared among a group of target nucleic acids. In some cases, a primer, or portion thereof, can hybridize to a primer binding site if there are at least about 6, 8, 10, 12, 14, 16, or 18 contiguous complementary nucleotides, including "universal" nucleotides that are complementary to more than one nucleotide partner.

Alternatively, a primer, or portion thereof, can hybridize to a primer binding site if there are fewer than 1 or 2 complementarity mismatches over at least about 12, 14, 16, or 18 contiguous complementary nucleotides. In some embodiments, the defined temperature at which specific hybridization occurs is room temperature. In some embodiments, the defined temperature at which specific hybridization occurs is higher than room temperature. In some embodiments, the defined temperature at which specific hybridization occurs is at least about 37, 40, 42, 45, 50, 55, 60, 65, 70, 75, or 80 °C. In some embodiments, the defined temperature at which specific hybridization occurs is 37, 40, 42, 45, 50, 55, 60, 65, 70, 75, or 80 °C.

[0086] A "template" refers to a polynucleotide sequence that comprises the polynucleotide to be amplified, flanked by or a pair of primer hybridization sites. Thus, a "target template" comprises the target polynucleotide sequence adjacent to at least one hybridization site for a primer. In some cases, a "target template" comprises the target polynucleotide sequence flanked by a hybridization site for a "forward" primer and a "reverse" primer.

[0087] As used herein, "nucleic acid" means DNA, RNA, single-stranded, double-stranded, or more highly aggregated hybridization motifs, and any chemical modifications thereof.

Modifications include, but are not limited to, those providing chemical groups that incorporate additional charge, polarizability, hydrogen bonding, electrostatic interaction, points of attachment and functionality to the nucleic acid ligand bases or to the nucleic acid ligand as a whole. Such modifications include, but are not limited to, peptide nucleic acids (PNAs), phosphodiester group modifications (*e.g.*, phosphorothioates, methylphosphonates), 2'-position sugar modifications, 5-position pyrimidine modifications, 8-position purine modifications,

modifications at exocyclic amines, substitution of 4-thiouridine, substitution of 5-bromo or 5-iodo-uracil; backbone modifications, methylations, unusual base-pairing combinations such as the isobases, isocytidine and isoguanidine and the like. Nucleic acids can also include non-natural bases, such as, for example, nitroindole. Modifications can also include 3' and 5' modifications including but not limited to capping with a fluorophore (e.g., quantum dot) or another moiety.

[0088] A "polymerase" refers to an enzyme that performs template-directed synthesis of polynucleotides, e.g., DNA and/or RNA. The term encompasses both the full length polypeptide and a domain that has polymerase activity. DNA polymerases are well-known to those skilled in the art, including but not limited to DNA polymerases isolated or derived from *Pyrococcus furiosus*, *Thermococcus litoralis*, and *Thermotoga maritime*, or modified versions thereof. Additional examples of commercially available polymerase enzymes include, but are not limited to: Klenow fragment (New England Biolabs® Inc.), Taq DNA polymerase (QIAGEN), 9<sup>o</sup>N<sup>TM</sup> DNA polymerase (New England Biolabs® Inc.), Deep Vent<sup>TM</sup> DNA polymerase (New England Biolabs® Inc.), Manta DNA polymerase (Enzymatics®), *Bst* DNA polymerase (New England Biolabs® Inc.), and phi29 DNA polymerase (New England Biolabs® Inc.).

[0089] Polymerases include both DNA-dependent polymerases and RNA-dependent polymerases such as reverse transcriptase. At least five families of DNA-dependent DNA polymerases are known, although most fall into families A, B and C. Other types of DNA polymerases include phage polymerases. Similarly, RNA polymerases typically include eukaryotic RNA polymerases I, II, and III, and bacterial RNA polymerases as well as phage and viral polymerases. RNA polymerases can be DNA-dependent and RNA-dependent.

[0090] As used herein, the term "partitioning" or "partitioned" refers to separating a sample into a plurality of portions, or "partitions." Partitions are generally physical, such that a sample in one partition does not, or does not substantially, mix with a sample in an adjacent partition. Partitions can be solid or fluid. In some embodiments, a partition is a solid partition, e.g., a microchannel. In some embodiments, a partition is a fluid partition, e.g., a droplet. In some embodiments, a fluid partition (e.g., a droplet) is a mixture of immiscible fluids (e.g., water and oil). In some embodiments, a fluid partition (e.g., a droplet) is an aqueous droplet that is surrounded by an immiscible carrier fluid (e.g., oil).

[0091] As used herein a “barcode” is a short nucleotide sequence (*e.g.*, at least about 4, 6, 8, 10, or 12, nucleotides long) that identifies a molecule to which it is conjugated. Barcodes can be used, *e.g.*, to identify molecules in a partition. Such a partition-specific barcode should be unique for that partition as compared to barcodes present in other partitions. For example, 5 partitions containing target RNA from single-cells can subject to reverse transcription conditions using primers that contain a different partition-specific barcode sequence in each partition, thus incorporating a copy of a unique “cellular barcode” into the reverse transcribed nucleic acids of each partition. Thus, nucleic acid from each cell can be distinguished from nucleic acid of other cells due to the unique “cellular barcode.” In some cases, the cellular barcode is provided by a 10 “particle barcode” that is present on oligonucleotides conjugated to a particle, wherein the particle barcode is shared by (*e.g.*, identical or substantially identical amongst) all, or substantially all, of the oligonucleotides conjugated to that particle. Thus, cellular and particle barcodes can be present in a partition, attached to a particle, or bound to cellular nucleic acid as multiple copies of the same barcode sequence. Cellular or particle barcodes of the same 15 sequence can be identified as deriving from the same cell, partition, or particle. Such partition-specific, cellular, or particle barcodes can be generated using a variety of methods, which methods result in the barcode conjugated to or incorporated into a solid or hydrogel support (*e.g.*, a solid bead or particle or hydrogel bead or particle). In some cases, the partition-specific, cellular, or particle barcode is generated using a split and mix (also referred to as split and pool) 20 synthetic scheme as described herein. A partition-specific barcode can be a cellular barcode and/or a particle barcode. Similarly, a cellular barcode can be a partition specific barcode and/or a particle barcode. Additionally, a particle barcode can be a cellular barcode and/or a partition-specific barcode.

[0092] In other cases, barcodes uniquely identify the molecule to which it is conjugated. For 25 example, by performing reverse transcription using primers that each contain a unique “molecular barcode.” In still other examples, primers can be utilized that contain “partition-specific barcodes” unique to each partition, and “molecular barcodes” unique to each molecule. After barcoding, partitions can then be combined, and optionally amplified, while maintaining virtual partitioning. Thus, *e.g.*, the presence or absence of a target nucleic acid (*e.g.*, reverse 30 transcribed nucleic acid) comprising each barcode can be counted (*e.g.* by sequencing) without the necessity of maintaining physical partitions.

[0093] The length of the barcode sequence determines how many unique samples can be differentiated. For example, a 1 nucleotide barcode can differentiate 4, or fewer, different samples or molecules; a 4 nucleotide barcode can differentiate  $4^4$  or 256 samples or less; a 6 nucleotide barcode can differentiate 4096 different samples or less; and an 8 nucleotide barcode can index 65,536 different samples or less. Additionally, barcodes can be attached to both strands either through barcoded primers for both first and second strand synthesis, through ligation, or in a tagmentation reaction.

[0094] Barcodes are typically synthesized and/or polymerized (*e.g.*, amplified) using processes that are inherently inexact. Thus, barcodes that are meant to be uniform (*e.g.*, a cellular, particle, or partition-specific barcode shared amongst all barcoded nucleic acid of a single partition, cell, or bead) can contain various N-1 deletions or other mutations from the canonical barcode sequence. Thus, barcodes that are referred to as “identical” or “substantially identical” copies refer to barcodes that differ due to one or more errors in, *e.g.*, synthesis, polymerization, or purification errors, and thus contain various N-1 deletions or other mutations from the canonical barcode sequence. Moreover, the random conjugation of barcode nucleotides during synthesis using *e.g.*, a split and pool approach and/or an equal mixture of nucleotide precursor molecules as described herein, can lead to low probability events in which a barcode is not absolutely unique (*e.g.*, different from all other barcodes of a population or different from barcodes of a different partition, cell, or bead). However, such minor variations from theoretically ideal barcodes do not interfere with the high-throughput sequencing analysis methods, compositions, and kits described herein. Therefore, as used herein, the term “unique” in the context of a particle, cellular, partition-specific, or molecular barcode encompasses various inadvertent N-1 deletions and mutations from the ideal barcode sequence. In some cases, issues due to the inexact nature of barcode synthesis, polymerization, and/or amplification, are overcome by oversampling of possible barcode sequences as compared to the number of barcode sequences to be distinguished (*e.g.*, at least about 2-, 5-, 10-fold or more possible barcode sequences). For example, 10,000 cells can be analyzed using a cellular barcode having 9 barcode nucleotides, representing 262,144 possible barcode sequences. The use of barcode technology is well known in the art, see for example Katsuyuki Shiroguchi, *et al.* Proc Natl Acad Sci U S A., 2012 Jan 24;109(4):1347-52; and Smith, AM *et al.*, Nucleic Acids Research Can 11, (2010). Further



methods and compositions for using barcode technology include those described in U.S. 2016/0060621.

[0095] A “transposase” or “tagmentase” means an enzyme that is capable of forming a functional complex with a transposon end-containing composition and catalyzing insertion or transposition of the transposon end-containing composition into the double-stranded target DNA with which it is incubated in an in vitro transposition reaction.

[0096] The term “transposon end” means a double-stranded DNA that exhibits only the nucleotide sequences (the “transposon end sequences”) that are necessary to form the complex with the transposase that is functional in an in vitro transposition reaction. A transposon end forms a “complex” or a “synaptic complex” or a “transposome complex” or a “transposome composition with a transposase or integrase that recognizes and binds to the transposon end, and which complex is capable of inserting or transposing the transposon end into target DNA with which it is incubated in an in vitro transposition reaction. A transposon end exhibits two complementary sequences consisting of a “transferred transposon end sequence” or “transferred strand” and a “non-transferred transposon end sequence,” or “non transferred strand” For example, one transposon end that forms a complex with a hyperactive Tn5 transposase (e.g., EZ-Tn5™ Transposase, EPICENTRE Biotechnologies, Madison, Wis., USA) that is active in an in vitro transposition reaction comprises a transferred strand that exhibits a “transferred transposon end sequence” as follows:

5' AGATGTGTATAAGAGACAG 3',

and a non-transferred strand that exhibits a “non-transferred transposon end sequence” as follows:

5' CTGTCTCTTATACACATCT 3'.

[0097] The 3'-end of a transferred strand is joined or transferred to target DNA in an in vitro transposition reaction. The non-transferred strand, which exhibits a transposon end sequence that is complementary to the transferred transposon end sequence, is not joined or transferred to the target DNA in an in vitro transposition reaction.

[0098] In some embodiments, the transferred strand and non-transferred strand are covalently joined. For example, in some embodiments, the transferred and non-transferred strand sequences are provided on a single oligonucleotide, e.g., in a hairpin configuration. As such, although the free end of the non-transferred strand is not joined to the target DNA directly by the transposition  
5 reaction, the non-transferred strand becomes attached to the DNA fragment indirectly, because the non-transferred strand is linked to the transferred strand by the loop of the hairpin structure.

## DETAILED DESCRIPTION OF THE INVENTION

### *Introduction*

10 [0099] The inventors have discovered an efficient method for maintaining contiguity of sequences such that sequences on the same haplotype can be determined and associated with each other. For example, genomic DNA can be treated with adaptor-loaded tagmentase such that the tagmentase randomly creates breakpoints in the DNA and inserts a first adaptor and a second  
15 adaptor on either side of the breakpoint. The adaptors when loaded onto a tagmentase have a common double-stranded portion and optionally a specific portion that is single-stranded and that can be the same in the case of homoadaptors or different in the case of heteroadaptors. Single-stranded adaptors are introduced to breakpoints by the tagmentase, wherein the single-stranded adaptor include the common sequence and, if present, the specific portion. The introduced  
20 adaptor sequences will have a 5' single-stranded overhang that can be filled by a polymerase. Conditions are selected such that the tagmentase does not release from the DNA, and thus forms a bridge linking DNA segments that have the same relationship (haplotype) as occurred in the genomic DNA. Thus the tagmentase step can occur in bulk (before partitions are formed). The  
25 resulting tagmentase-bridged DNA segments can be added to partitions such that one set of bridged DNA segments (e.g., on average) are in partitions. Contiguity is therefore maintained until the bridged segments have been separated into different partitions. Once in partitions, the contiguous DNA segments can be barcoded with a partition-specific barcode, thereby allowing for later identification of contiguous DNA after sequencing in bulk (after partitions contents are merged).

[0100] The partitions will further include an agent that separates the DNA segments. In  
30 embodiments in which the tagmentase itself bridges the DNA segments, an agent can be

included in the partitions that disassociates the tagmentase from the DNA segments. In embodiments in which there is a linking sequence linking the tagmentase-added adaptors, an agent can be included in partitions that cleaves the linking sequence.

5 [0101] The partitions will further include a bead linked to a first oligonucleotide having a free 3' end. The oligonucleotide will include a barcode specific for the particular bead to which it is linked, and also includes a 3' end sequence that is specific for and complementary to an adaptor sequence on the DNA segment. The partitions can in some embodiments further include a second oligonucleotide that functions as a reverse primer, having a 3' end specific for and complementary to the adaptor sequence on the opposite end of the DNA segment compared to  
10 the adaptor at which the first oligonucleotide hybridizes.

[0102] Prior to amplification, there is a gap filling reaction such that the complement to the 5' overhang of the transposase adapter is synthesized. Gap filling does not involve ligation. Amplification is performed in the partitions. Thus, each partition only contains contiguous DNA from the genomic DNA. The resulting amplicons will contain the bead-specific barcode.

15 [0103] After amplification, the contents of the partitions can be merged to form a mixture of amplicons. The resulting mixture of amplicons can be nucleotide sequenced by any method desired by the user. DNA from the same haplotype will be readily identifiable as having the same bead-specific barcode.

20 [0104] Performing the method in partitions (e.g., droplets) allows for: 1) release of bead oligonucleotides, which makes the thermodynamics of finding binding partners more favorable and 2) performing an overhang + gap fill to create a contiguous bottom strand that is an extendible template for the released oligonucleotide primers. This also adds to efficiency multiple fold over ligation. This achieves much higher efficiency than methods performed on beads, thus better coverage, and thus fewer numbers of bead - reactions required.

## 25 *Genomic DNA*

[0105] Any genomic DNA can be used in the methods,. In some embodiments, the DNA is from a single cell or is from a single type of cell from an organism. In some embodiments, the genomic DNA is from a eukaryote, for example from a mammal, e.g. a human. In some embodiments, the DNA is from a plant or fungus. In some embodiments, the starting DNA is

purified as desired and used directly in the method. Alternatively, DNA can be treated to generate DNA fragments of a desired average size, for example using size-selection columns or gel purification. For example, in some embodiments, the starting DNA fragments are on average between 5kb-10 Mb.

## 5 *Tagmentase*

[0106] Heteroadapter-loaded tagmentases and homoadapter-loaded tagmentases can be used as described herein. Homoadapter-loaded tagmentases are tagmentases that contain adaptors of only one sequence, which adaptor is added to either end of a tagmentase-induced breakpoint in the genomic DNA. Heteroadapter loaded tagmentases are tagmentases that contain two different  
10 adaptors, such that a different adaptor sequence is added to the two DNA ends created by a tagmentase-induced breakpoint in the DNA. Adapter loaded tagmentases are further described, e.g., in U.S. Patent Publication Nos: 2010/0120098; 2012/0301925; and 2015/0291942 and U.S. Patent Nos: 5,965,443; U.S. 6,437,109; 7,083,980; 9,005,935; and 9,238,671, the contents of each of which are hereby incorporated by reference in the entirety for all purposes.

15 [0107] A tagmentase is an enzyme that is capable of forming a functional complex with a transposon end-containing composition and catalyzing insertion or transposition of the transposon end-containing composition into the double-stranded target DNA with which it is incubated in an in vitro transposition reaction. Exemplary transposases include but are not limited to modified TN5 transposases that are hyperactive compared to wildtype TN5, for  
20 example can have one or more mutations selected from E54K, M56A, or L372P. Wild-type Tn5 transposon is a composite transposon in which two near-identical insertion sequences (IS50L and IS50R) are flanking three antibiotic resistance genes (Reznikoff WS. *Annu Rev Genet* 42: 269–286 (2008)). Each IS50 contains two inverted 19-bp end sequences (ESs), an outside end (OE) and an inside end (IE). However, wild-type ESs have a relatively low activity and were replaced  
25 in vitro by hyperactive mosaic end (ME) sequences. A complex of the transposase with the 19-bp ME is thus all that is necessary for transposition to occur, provided that the intervening DNA is long enough to bring two of these sequences close together to form an active Tn5 transposase homodimer (Reznikoff WS., *Mol Microbiol* 47: 1199–1206 (2003)). Transposition is a very infrequent event in vivo, and hyperactive mutants were historically derived by introducing three  
30 missense mutations in the 476 residues of the Tn5 protein (E54K, M56A, L372P), which is

encoded by IS50R (Goryshin IY, Reznikoff WS. 1998. *J Biol Chem* 273: 7367–7374 (1998)).

Transposition works through a “cut-and-paste” mechanism, where the Tn5 excises itself from the donor DNA and inserts into a target sequence, creating a 9-bp duplication of the target (Schaller H. *Cold Spring Harb Symp Quant Biol* 43: 401–408 (1979); Reznikoff WS., *Annu Rev Genet* 42: 269–286 (2008)). In current commercial solutions (Nextera™ DNA kits, Illumina), free synthetic ME adaptors are end-joined to the 5'-end of the target DNA by the transposase (tagmentase). In some embodiments, the tagmentase is linked to a solid support (e.g., a bead that is different from the bead linked to the forward primer). An example commercial bead-linked tagmentase is Nextera™ DNA Flex (Illumina).

10 [0108] In some embodiments, the adaptor(s) is at least 19 nucleotides in length, e.g., 19-100 nucleotides. In some embodiments, the adapters are double stranded with a 5' end overhang, wherein the 5' overhand sequence is different between heteroadaptors, while the double stranded portion (typically 19 bp) is the same. In some embodiments, an adaptor comprises TCGTCGGCAGCGTC or GTCTCGTGGGCTCGG. In some embodiments involving the

15 heteroadaptor-loaded tagmentase, the tagmentase is loaded with a first adaptor comprising TCGTCGGCAGCGTC and a second adaptor comprising GTCTCGTGGGCTCGG. In some embodiments, the adapter comprises AGATGTGTATAAGAGACAG and the complement thereof (this is the mosaic end and this is the only specifically required cis active sequence for Tn5 transposition). In some embodiments, the adapter comprises

20 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG with the complement for AGATGTGTATAAGAGACAG or GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG with the complement for AGATGTGTATAAGAGACAG. In some embodiments involving the heteroadaptor-loaded tagmentase, the tagmentase is loaded with a first adaptor comprising

25 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG with the complement for AGATGTGTATAAGAGACAG and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG with the complement for AGATGTGTATAAGAGACAG.

[0109] In some embodiments, the adaptors have a 19bp double stranded region and a 5' 15 bp single stranded overhang. The sequence of the 15 bp is different between heteroadaptors whereas the double-stranded region has a common sequence between adaptors (homo or heteroadaptors).

30 See, e.g., FIG. 5.

[0110] In some embodiments, whether the tagmentase is loaded with hetero or homo adaptors, the pair of adaptors can be linked via a linking nucleotide sequence. This aspect is depicted in FIG. 21-22. The linking sequence can be any nucleotide sequence linking the two adaptors. The linking sequence can be, in some embodiments, between 2 nucleotides to 5 kb long. In some  
5 embodiments, the linking sequence can contain one or more restriction recognition sequence such that the linking sequence can be cleaved later by a restriction enzyme added to the partitions. To avoid cleavage within the DNA segments themselves, it can be beneficial to select a rare cutting restriction enzyme, for example a restriction enzyme having a recognition sequence having 8 or more nucleotides.

10 [0111] In other embodiments, the linking sequence can contain one or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, or 10) uracils. The linking sequence can subsequently be cleaved in the presence of uracil-DNA N-glycosylase (e.g., “UNG”), which can be included in the partition.

[0112] In other embodiments, the linking sequence can contain one or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, or 10) ribonucleotides. The linking sequence can subsequently be cleaved in the presence  
15 of base or RNase, which can be included in the partition.

[0113] Conditions for tagmentation are selected such that the tagmentase creates breakpoints in the DNA and such that adaptors loaded on the tagmentase are added to either end of the breakpoint. The tagmentase introduces a single-stranded adaptor sequence on either end of the breakpoint, forming a 5' overhang. The 5' overhang is then filled (“gap filled”) by a polymerase  
20 to create a double-stranded sequence at either end of the DNA segments. See, e.g., FIG. 9. The non-transferred bottom strand is thus rendered contiguous and complementary to the transferred top strand. This contiguous bottom strand is now compatible with polymerase extension reactions, e.g., PCR. Thus “gap filling” is the process after tagmentation that renders the bottom strand (the one that is not transferred) contiguous with the top strand at the end of the DNA  
25 segments. Gap filling refers to reconstitution of the bottom strand. This is done preferentially by a DNA polymerase that extends back from the 3' of the bottom non-transferred strand that is upstream of the gap. The polymerase that gap fills can either have 5' to 3' exo activity or strand displacing activity to help overcome the non-transferred mosaic end. Neither gap filling nor adding of the barcode involves ligation.

[0114] Conditions are also selected such that the tagmentase remains bound to the DNA breakpoints thereby maintaining contiguity. Tagmentase has been observed to remain bound to DNA until a detergent such as SDS is added to the reaction (Amini *et al. Nature Genetics* 46(12):1343-1349).

## 5 *Partitions*

[0115] Any type of partition can be used in the methods described herein. While the method has been exemplified using droplets it should be understood that other types of partitions can also be used.

[0116] In some embodiments, prior to contact to the segmented DNA, the partitions will  
10 contain a first oligonucleotide linked to a bead and optionally a second reverse oligonucleotide primer. The partitions can also include a sufficient amount of an agent to remove tagmentase from the DNA or to cleave the linking sequences, if present. As described herein, subsequently, the segmented and linked DNA can be added to partitions. The number of segmented and linked DNA molecules, while ideally in some embodiments, can be one per partition, can be selected as  
15 needed depending for example on the number of partitions and other factors.

[0117] Methods and compositions for partitioning are described, for example, in published patent applications WO 2010/036,352, US 2010/0173,394, US 2011/0092,373, and US 2011/0092,376, the contents of each of which are incorporated herein by reference in the entirety. The plurality of mixture partitions can be in a plurality of emulsion droplets, or a  
20 plurality of microwells, *etc.*

[0118] In some embodiments, the primers and other reagents can be partitioned into a plurality of mixture partitions, and then linked DNA segments can be introduced into the plurality of mixture partitions. Methods and compositions for delivering reagents to one or more mixture partitions include microfluidic methods as known in the art; droplet or microcapsule merging, coalescing, fusing, bursting, or degrading (*e.g.*, as described in U.S. 2015/0027,892; US 25 2014/0227,684; WO 2012/149,042; and WO 2014/028,537); droplet injection methods (*e.g.*, as described in WO 2010/151,776); and combinations thereof.

[0119] As described herein, the mixture partitions can be picowells, nanowells, or microwells. The mixture partitions can be pico-, nano-, or micro- reaction chambers, such as pico, nano, or

microcapsules. The mixture partitions can be pico-, nano-, or micro- channels. The mixture partitions can be droplets, *e.g.*, emulsion droplets.

[0120] In some embodiments, the partitions are droplets. In some embodiments, a droplet comprises an emulsion composition, *i.e.*, a mixture of immiscible fluids (*e.g.*, water and oil). In some embodiments, a droplet is an aqueous droplet that is surrounded by an immiscible carrier fluid (*e.g.*, oil). In some embodiments, a droplet is an oil droplet that is surrounded by an immiscible carrier fluid (*e.g.*, an aqueous solution). In some embodiments, the droplets described herein are relatively stable and have minimal coalescence between two or more droplets. In some embodiments, less than 0.0001%, 0.0005%, 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, or 10% of droplets generated from a sample coalesce with other droplets. The emulsions can also have limited flocculation, a process by which the dispersed phase comes out of suspension in flakes. In some cases, such stability or minimal coalescence is maintained for up to 4, 6, 8, 10, 12, 24, or 48 hours or more (*e.g.*, at room temperature, or at about 0, 2, 4, 6, 8, 10, or 12 °C). In some embodiments, the droplet is formed by flowing an oil phase through an aqueous sample or reagents.

[0121] The oil phase can comprise a fluorinated base oil which can additionally be stabilized by combination with a fluorinated surfactant such as a perfluorinated polyether. In some embodiments, the base oil comprises one or more of a HFE 7500, FC-40, FC-43, FC-70, or another common fluorinated oil. In some embodiments, the oil phase comprises an anionic fluorosurfactant. In some embodiments, the anionic fluorosurfactant is Ammonium Krytox (Krytox-AS), the ammonium salt of Krytox FSH, or a morpholino derivative of Krytox FSH. Krytox-AS can be present at a concentration of about 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1.0%, 2.0%, 3.0%, or 4.0% (w/w). In some embodiments, the concentration of Krytox-AS is about 1.8%. In some embodiments, the concentration of Krytox-AS is about 1.62%. Morpholino derivative of Krytox FSH can be present at a concentration of about 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1.0%, 2.0%, 3.0%, or 4.0% (w/w). In some embodiments, the concentration of morpholino derivative of Krytox FSH is about 1.8%. In some embodiments, the concentration of morpholino derivative of Krytox FSH is about 1.62%.

[0122] In some embodiments, the oil phase further comprises an additive for tuning the oil properties, such as vapor pressure, viscosity, or surface tension. Non-limiting examples include



perfluorooctanol and 1H,1H,2H,2H-Perfluorodecanol. In some embodiments, 1H,1H,2H,2H-Perfluorodecanol is added to a concentration of about 0.05%, 0.06%, 0.07%, 0.08%, 0.09%, 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1.0%, 1.25%, 1.50%, 1.75%, 2.0%, 2.25%, 2.5%, 2.75%, or 3.0% (w/w). In some embodiments, 1H,1H,2H,2H-Perfluorodecanol is added to a concentration of about 0.18% (w/w).

[0123] In some embodiments, the emulsion is formulated to produce highly monodisperse droplets having a liquid-like interfacial film that can be converted by heating into microcapsules having a solid-like interfacial film; such microcapsules can behave as bioreactors able to retain their contents through an incubation period. The conversion to microcapsule form can occur upon heating. For example, such conversion can occur at a temperature of greater than about 40°, 50°, 60°, 70°, 80°, 90°, or 95 °C. During the heating process, a fluid or mineral oil overlay can be used to prevent evaporation. Excess continuous phase oil can be removed prior to heating, or left in place. The microcapsules can be resistant to coalescence and/or flocculation across a wide range of thermal and mechanical processing.

[0124] Following conversion of droplets into microcapsules, the microcapsules can be stored at about -70°, -20°, 0°, 3°, 4°, 5°, 6°, 7°, 8°, 9°, 10°, 15°, 20°, 25°, 30°, 35°, or 40° C. In some embodiments, these capsules are useful for storage or transport of partition mixtures. For example, samples can be collected at one location, partitioned into droplets containing enzymes, buffers, and/or primers or other probes, optionally one or more polymerization reactions can be performed, the partitions can then be heated to perform microencapsulation, and the microcapsules can be stored or transported for further analysis.

[0125] In some embodiments, the sample is partitioned into, or into at least, 500 partitions, 1000 partitions, 2000 partitions, 3000 partitions, 4000 partitions, 5000 partitions, 6000 partitions, 7000 partitions, 8000 partitions, 10,000 partitions, 15,000 partitions, 20,000 partitions, 30,000 partitions, 40,000 partitions, 50,000 partitions, 60,000 partitions, 70,000 partitions, 80,000 partitions, 90,000 partitions, 100,000 partitions, 200,000 partitions, 300,000 partitions, 400,000 partitions, 500,000 partitions, 600,000 partitions, 700,000 partitions, 800,000 partitions, 900,000 partitions, 1,000,000 partitions, 2,000,000 partitions, 3,000,000 partitions, 4,000,000 partitions, 5,000,000 partitions, 10,000,000 partitions, 20,000,000 partitions, 30,000,000 partitions, 40,000,000 partitions, 50,000,000 partitions, 60,000,000 partitions, 70,000,000 partitions,

80,000,000 partitions, 90,000,000 partitions, 100,000,000 partitions, 150,000,000 partitions, or 200,000,000 partitions.

[0126] In some embodiments, the droplets that are generated are substantially uniform in shape and/or size. For example, in some embodiments, the droplets are substantially uniform in  
5 average diameter. In some embodiments, the droplets that are generated have an average diameter of about 0.001 microns, about 0.005 microns, about 0.01 microns, about 0.05 microns, about 0.1 microns, about 0.5 microns, about 1 microns, about 5 microns, about 10 microns, about 20 microns, about 30 microns, about 40 microns, about 50 microns, about 60 microns, about 70 microns, about 80 microns, about 90 microns, about 100 microns, about 150 microns, about 200  
10 microns, about 300 microns, about 400 microns, about 500 microns, about 600 microns, about 700 microns, about 800 microns, about 900 microns, or about 1000 microns. In some embodiments, the droplets that are generated have an average diameter of less than about 1000 microns, less than about 900 microns, less than about 800 microns, less than about 700 microns, less than about 600 microns, less than about 500 microns, less than about 400 microns, less than  
15 about 300 microns, less than about 200 microns, less than about 100 microns, less than about 50 microns, or less than about 25 microns. In some embodiments, the droplets that are generated are non-uniform in shape and/or size.

[0127] In some embodiments, the droplets that are generated are substantially uniform in volume. For example, the standard deviation of droplet volume can be less than about 1  
20 picoliter, 5 picoliters, 10 picoliters, 100 picoliters, 1 nL, or less than about 10 nL. In some cases, the standard deviation of droplet volume can be less than about 10-25% of the average droplet volume. In some embodiments, the droplets that are generated have an average volume of about 0.001 nL, about 0.005 nL, about 0.01 nL, about 0.02 nL, about 0.03 nL, about 0.04 nL, about 0.05 nL, about 0.06 nL, about 0.07 nL, about 0.08 nL, about 0.09 nL, about 0.1 nL, about 0.2 nL,  
25 about 0.3 nL, about 0.4 nL, about 0.5 nL, about 0.6 nL, about 0.7 nL, about 0.8 nL, about 0.9 nL, about 1 nL, about 1.5 nL, about 2 nL, about 2.5 nL, about 3 nL, about 3.5 nL, about 4 nL, about 4.5 nL, about 5 nL, about 5.5 nL, about 6 nL, about 6.5 nL, about 7 nL, about 7.5 nL, about 8 nL, about 8.5 nL, about 9 nL, about 9.5 nL, about 10 nL, about 11 nL, about 12 nL, about 13 nL, about 14 nL, about 15 nL, about 16 nL, about 17 nL, about 18 nL, about 19 nL, about 20 nL,  
30 about 25 nL, about 30 nL, about 35 nL, about 40 nL, about 45 nL, or about 50 nL.

[0128] As noted above, the partitions will include one or a few (e.g., 1, 2, 3, 4) beads per partition, where in each bead is linked to a first oligonucleotide primer having a free 3' end. The first oligonucleotide primer will have a bead-specific barcode and a 3' end that is complementary to an adaptor. In some embodiments, the barcode will be, e.g., 2-10 nucleotides in length, e.g., 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. The barcode can be continuous or discontinuous, i.e., broken up by other nucleotides. In some embodiments, the 3' end will be complementary to the entire adaptor sequence. In some embodiments, at least the 3'-most 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 of the oligonucleotide are complementary to a sequence in the adaptor. In some embodiments, the first oligonucleotide primer further comprises a universal or other additional sequence to assist with downstream manipulation or sequencing of the amplicon. For example, when Illumina-based sequencing is used the first oligonucleotide primer can have a 5' P5 or P7 sequence (optionally with the second oligonucleotide primer having the other of the two sequences). Optionally, the first oligonucleotide primer comprises a restriction or cleavage site to remove the first oligonucleotide primer from the bead when desired. In some embodiments, once the DNA segments are in the partitions with the bead-linked first oligonucleotide primer, the first oligonucleotide primer is cleaved from the bead prior to amplification.

[0129] The term "bead" refers to any solid support that can be in a partition, e.g., a small particle or other solid support. Exemplary beads can include hydrogel beads. In some cases, the hydrogel is in *sol* form. In some cases, the hydrogel is in *gel* form. An exemplary hydrogel is an agarose hydrogel. Other hydrogels include, but are not limited to, those described in, e.g., U.S. Patent Nos. 4,438,258; 6,534,083; 8,008,476; 8,329,763; U.S. Patent Appl. Nos. 2002/0,009,591; 2013/0,022,569; 2013/0,034,592; and International Patent Publication Nos. WO/1997/030092; and WO/2001/049240.

[0130] Methods of linking oligonucleotides to beads are described in, e.g., WO 2015/200541. In some embodiments, the oligonucleotide configured to link the hydrogel to the barcode is covalently linked to the hydrogel. Numerous methods for covalently linking an oligonucleotide to one or more hydrogel matrices are known in the art. As but one example, aldehyde derivatized agarose can be covalently linked to a 5'-amine group of a synthetic oligonucleotide.

[0131] As noted elsewhere herein, the partitions can also contain a second oligonucleotide primer, which can optionally be linked to the bead, or not. This primer can function as a reverse

primer for the first oligonucleotide primer such that the two oligonucleotides generate an amplicon in PCR. The second oligonucleotide primer will have a 3' end that is complementary to an adaptor sequence, i.e., the adaptor sequence at the opposite end of the DNA segment compared to the adaptor sequence targeted by the first oligonucleotide primer. In some  
5 embodiments, the 3' end will be complementary to the entire adaptor sequence. In some embodiments, at least the 3'-most 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 of the oligonucleotide are complementary to a sequence in the adaptor. The second oligonucleotide primer can also contain a universal or other additional sequence to assist with downstream manipulation or sequencing of the amplicon. For example, when Illumina-based sequencing is  
10 used the second oligonucleotide primer can have a 5' P5 or P7 sequence for binding to the Illumina flow cell (optionally with the first oligonucleotide primer having the other of the two sequences).

[0132] The partitions can also contain an agent that removes tagmentase from the DNA segments. In some embodiments, the agent is a detergent, e.g., an ionic or non-ionic detergent.  
15 An exemplary detergent is sodium dodecyl sulfate (SDS). For example, the inventors have determined that concentrations of 0.1 and 0.2% SDS are sufficient to remove the tagmentase and yet are sufficiently low to not interfere with amplification. Thus, in some embodiments, the partitions contain 0.02-0.4%, e.g., 0.05-0.3% SDS. In some embodiments (e.g., in which droplets are used) the agent is compatible with droplet formation. In some embodiments, the  
20 agent is a polymerase, e.g., the polymerase used for gap filling or that is otherwise used to amplify the DNA. Alternatively, in some embodiments, the tagmentase can be displaced by heat. For example, in some embodiments, heating to 70 C degrees or more (e.g., around 72°C) reduces affinity of the tagmentase for the DNA or displaces the tagmentase.

[0133] Prior to amplification, one can remove or cleave the first oligonucleotide primer from  
25 the bead. This can be achieved by any method as desired. Methods of cleaving include, but are not limited to altering the pH or contacting the oligonucleotides with UDG/ApeI or a restriction endonuclease. In some embodiments, the oligonucleotide is linked to the bead via one or more uracils (Us) and USER enzyme (e.g., from NEB) is used to cleave the Us incorporated in the oligo backbone. USER has 2 enzymes: UDG and Endonuclease VIII. In some cases, the  
30 oligonucleotides are attached to a solid support through a disulfide linkage (e.g., through a

disulfide bond between a sulfide of the solid support and a sulfide covalently attached to the 5' or 3' end, or an intervening nucleic acid, of the oligonucleotide). In such cases, the oligonucleotide can be cleaved from the solid support by contacting the solid support with a reducing agent such as a thiol or phosphine reagent, including but not limited to a beta mercaptoethanol, dithiothreitol (DTT), or tris(2-carboxyethyl)phosphine (TCEP). It can be advantageous to release the first oligonucleotide primer from the bead for a number of reasons. For example, thermodynamics of DNA interactions will greatly increase.

[0134] Amplification can be achieved within the partitions (before combining the contents to the partitions). Various digital amplification methods are known and can be used.

10 [0135] Following amplification, the contents of the partitions are combined and sequenced. Any method of nucleotide sequencing can be used as desired so long as at least some of the DNA segments sequence and the barcode sequence is determined. Methods for high throughput sequencing and genotyping are known in the art. For example, such sequencing technologies include, but are not limited to, pyrosequencing, sequencing-by-ligation, single molecule  
15 sequencing, sequence-by-synthesis (SBS), massive parallel clonal, massive parallel single molecule SBS, massive parallel single molecule real-time, massive parallel single molecule real-time nanopore technology, *etc.* Morozova and Marra provide a review of some such technologies in *Genomics*, 92: 255 (2008), herein incorporated by reference in its entirety.

[0136] Exemplary DNA sequencing techniques include fluorescence-based sequencing  
20 methodologies (*See, e.g.,* Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, Cold Spring Harbor, N.Y.; herein incorporated by reference in its entirety). In some embodiments, automated sequencing techniques understood in that art are utilized. In some embodiments, the present technology provides parallel sequencing of partitioned amplicons (PCT Publication No. WO 2006/0841,32, herein incorporated by reference in its entirety). In some embodiments, DNA  
25 sequencing is achieved by parallel oligonucleotide extension (*See, e.g.,* U.S. Pat. Nos. 5,750,341; and 6,306,597, both of which are herein incorporated by reference in their entireties). Additional examples of sequencing techniques include the Church polony technology (Mitra *et al.*, 2003, *Analytical Biochemistry* 320, 55-65; Shendure *et al.*, 2005 *Science* 309, 1728-1732; and U.S. Pat. Nos. 6,432,360; 6,485,944; 6,511,803; herein incorporated by reference in their entireties),  
30 the 454 picotiter pyrosequencing technology (Margulies *et al.*, 2005 *Nature* 437, 376-380; U.S.

Publication No. 2005/0130173; herein incorporated by reference in their entireties), the Solexa single base addition technology (Bennett *et al.*, 2005, *Pharmacogenomics*, 6, 373-382; U.S. Pat. Nos. 6,787,308; and 6,833,246; herein incorporated by reference in their entireties), the Lynx massively parallel signature sequencing technology (Brenner *et al.* (2000). *Nat. Biotechnol.* 18:630-634; U.S. Pat. Nos. 5,695,934; 5,714,330; herein incorporated by reference in their entireties), and the Adessi PCR colony technology (Adessi *et al.* (2000). *Nucleic Acid Res.* 28, E87; WO 2000/018957; herein incorporated by reference in its entirety).

[0137] Typically, high throughput sequencing methods share the common feature of massively parallel, high-throughput strategies, with the goal of lower costs in comparison to older sequencing methods (*See, e.g.,* Voelkerding *et al.*, *Clinical Chem.*, 55: 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbiol.*, 7:287-296; each herein incorporated by reference in their entirety). Such methods can be broadly divided into those that typically use template amplification and those that do not. Amplification-requiring methods include pyrosequencing commercialized by Roche as the 454 technology platforms (*e.g.,* GS 20 and GS FLX), the Solexa platform commercialized by Illumina, and the Supported Oligonucleotide Ligation and Detection (SOLiD) platform commercialized by Applied Biosystems. Non-amplification approaches, also known as single-molecule sequencing, are exemplified by the HeliScope platform commercialized by Helicos BioSciences, and platforms commercialized by VisiGen, Oxford Nanopore Technologies Ltd., Life Technologies/Ion Torrent, and Pacific Biosciences, respectively.

[0138] In pyrosequencing (Voelkerding *et al.*, *Clinical Chem.*, 55: 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbiol.*, 7:287-296; U.S. Pat. Nos. 6,210,891; and 6,258,568; each herein incorporated by reference in its entirety), template DNA is fragmented, end-repaired, ligated to adaptors, and clonally amplified *in-situ* by capturing single template molecules with beads bearing oligonucleotides complementary to the adaptors. Each bead bearing a single template type is compartmentalized into a water-in-oil microvesicle, and the template is clonally amplified using a technique referred to as emulsion PCR. The emulsion is disrupted after amplification and beads are deposited into individual wells of a picotitre plate functioning as a flow cell during the sequencing reactions. Ordered, iterative introduction of each of the four dNTP reagents occurs in the flow cell in the presence of sequencing enzymes and luminescent reporter such as

luciferase. In the event that an appropriate dNTP is added to the 3' end of the sequencing primer, the resulting production of ATP causes a burst of luminescence within the well, which is recorded using a CCD camera. It is possible to achieve read lengths greater than or equal to 400 bases, and  $10^6$  sequence reads can be achieved, resulting in up to 500 million base pairs (Mb) of sequence.

[0139] In the Solexa/Illumina platform (Voelkerding *et al.*, *Clinical Chem.*, 55: 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbial.*, 7:287-296; U.S. Pat. Nos. 6,833,246; 7,115,400; and 6,969,488; each herein incorporated by reference in its entirety), sequencing data are produced in the form of shorter-length reads. In this method, single-stranded fragmented DNA is end-repaired to generate 5'-phosphorylated blunt ends, followed by Klenow-mediated addition of a single A base to the 3' end of the fragments. A-addition facilitates addition of T-overhang adaptor oligonucleotides, which are subsequently used to capture the template-adaptor molecules on the surface of a flow cell that is studded with oligonucleotide anchors. The anchor is used as a PCR primer, but because of the length of the template and its proximity to other nearby anchor oligonucleotides, extension by PCR results in the "arching over" of the molecule to hybridize with an adjacent anchor oligonucleotide to form a bridge structure on the surface of the flow cell. These loops of DNA are denatured and cleaved. Forward strands are then sequenced with reversible dye terminators. The sequence of incorporated nucleotides is determined by detection of post-incorporation fluorescence, with each fluor and block removed prior to the next cycle of dNTP addition. Sequence read length ranges from 36 nucleotides to over 50 nucleotides, with overall output exceeding 1 billion nucleotide pairs per analytical run.

[0140] Sequencing nucleic acid molecules using SOLiD technology (Voelkerding *et al.*, *Clinical Chem.*, 55: 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbial.*, 7:287-296; U.S. Pat. Nos. 5,912,148; and 6,130,073; each herein incorporated by reference in their entirety) also involves fragmentation of the template, ligation to oligonucleotide adaptors, attachment to beads, and clonal amplification by emulsion PCR. Following this, beads bearing template are immobilized on a derivatized surface of a glass flow-cell, and a primer complementary to the adaptor oligonucleotide is annealed. However, rather than utilizing this primer for 3' extension, it is instead used to provide a 5' phosphate group for ligation to interrogation probes containing two probe-specific bases followed by 6 degenerate bases and one of four fluorescent labels. In

the SOLiD system, interrogation probes have 16 possible combinations of the two bases at the 3' end of each probe, and one of four fluors at the 5' end. Fluor color, and thus identity of each probe, corresponds to specified color-space coding schemes. Multiple rounds (usually 7) of probe annealing, ligation, and fluor detection are followed by denaturation, and then a second round of sequencing using a primer that is offset by one base relative to the initial primer. In this manner, the template sequence can be computationally re-constructed, and template bases are interrogated twice, resulting in increased accuracy. Sequence read length averages 35 nucleotides, and overall output exceeds 4 billion bases per sequencing run.

[0141] In certain embodiments, nanopore sequencing is employed (*See, e.g., Astier et al., J. Am. Chem. Soc.* 2006 Feb. 8; 128(5)1705-10, herein incorporated by reference). The theory behind nanopore sequencing has to do with what occurs when a nanopore is immersed in a conducting fluid and a potential (voltage) is applied across it. Under these conditions a slight electric current due to conduction of ions through the nanopore can be observed, and the amount of current is exceedingly sensitive to the size of the nanopore. As each base of a nucleic acid passes through the nanopore, this causes a change in the magnitude of the current through the nanopore that is distinct for each of the four bases, thereby allowing the sequence of the DNA molecule to be determined.

[0142] In certain embodiments, HeliScope by Helicos BioSciences is employed (Voelkerding *et al., Clinical Chem.*, 55. 641-658, 2009; MacLean *et al., Nature Rev. Microbial*, 7:287-296; U.S. Pat. Nos. 7,169,560; 7,282,337; 7,482,120; 7,501,245; 6,818,395; 6,911,345; and 7,501,245; each herein incorporated by reference in their entirety). Template DNA is fragmented and polyadenylated at the 3' end, with the final adenosine bearing a fluorescent label. Denatured polyadenylated template fragments are ligated to poly(dT) oligonucleotides on the surface of a flow cell. Initial physical locations of captured template molecules are recorded by a CCD camera, and then label is cleaved and washed away. Sequencing is achieved by addition of polymerase and serial addition of fluorescently-labeled dNTP reagents. Incorporation events result in fluor signal corresponding to the dNTP, and signal is captured by a CCD camera before each round of dNTP addition. Sequence read length ranges from 25-50 nucleotides, with overall output exceeding 1 billion nucleotide pairs per analytical run.



[0143] The Ion Torrent technology is a method of DNA sequencing based on the detection of hydrogen ions that are released during the polymerization of DNA (*See, e.g., Science* 327(5970): 1190 (2010); U.S. Pat. Appl. Pub. Nos. 2009/0026082; 2009/0127589; 2010/0301398; 2010/0197507; 2010/0188073; and 2010/0137143, incorporated by reference in their entireties for all purposes). A microwell contains a template DNA strand to be sequenced. Beneath the layer of microwells is a hypersensitive ISFET ion sensor. All layers are contained within a CMOS semiconductor chip, similar to that used in the electronics industry. When a dNTP is incorporated into the growing complementary strand a hydrogen ion is released, which triggers the hypersensitive ion sensor. If homopolymer repeats are present in the template sequence, multiple dNTP molecules will be incorporated in a single cycle. This leads to a corresponding number of released hydrogens and a proportionally higher electronic signal. This technology differs from other sequencing technologies in that no modified nucleotides or optics are used. The per base accuracy of the Ion Torrent sequencer is ~99.6% for 50 base reads, with ~100 Mb generated per run. The read-length is 100 base pairs. The accuracy for homopolymer repeats of 5 repeats in length is ~98%. The benefits of ion semiconductor sequencing are rapid sequencing speed and low upfront and operating costs.

[0144] Another exemplary nucleic acid sequencing approach that may be adapted for use with the present invention was developed by Stratos Genomics, Inc. and involves the use of Xpandomers. This sequencing process typically includes providing a daughter strand produced by a template-directed synthesis. The daughter strand generally includes a plurality of subunits coupled in a sequence corresponding to a contiguous nucleotide sequence of all or a portion of a target nucleic acid in which the individual subunits comprise a tether, at least one probe or nucleobase residue, and at least one selectively cleavable bond. The selectively cleavable bond(s) is/are cleaved to yield an Xpandomer of a length longer than the plurality of the subunits of the daughter strand. The Xpandomer typically includes the tethers and reporter elements for parsing genetic information in a sequence corresponding to the contiguous nucleotide sequence of all or a portion of the target nucleic acid. Reporter elements of the Xpandomer are then detected. Additional details relating to Xpandomer-based approaches are described in, for example, U.S. Pat. Pub No. 2009/0035777, which is incorporated herein in its entirety.

[0145] Other single molecule sequencing methods include real-time sequencing by synthesis using a VisiGen platform (Voelkerding *et al.*, *Clinical Chem.*, 55: 641-58, 2009; U.S. Pat. No. 7,329,492; and U.S. Patent Appl. Ser. Nos. 11/671,956; and 11/781,166; each herein incorporated by reference in their entirety) in which immobilized, primed DNA template is  
5 subjected to strand extension using a fluorescently-modified polymerase and fluorescent acceptor molecules, resulting in detectable fluorescence resonance energy transfer (FRET) upon nucleotide addition.

[0146] Another real-time single molecule sequencing system developed by Pacific Biosciences (Voelkerding *et al.*, *Clinical Chem.*, 55: 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbiol.*,  
10 7:287-296; U.S. Pat. Nos. 7,170,050; 7,302,146; 7,313,308; and 7,476,503; all of which are herein incorporated by reference) utilizes reaction wells 50-100 nm in diameter and encompassing a reaction volume of approximately 20 zeptoliters ( $10^{-21}$  L). Sequencing reactions are performed using immobilized template, modified phi29 DNA polymerase, and high local concentrations of fluorescently labeled dNTPs. High local concentrations and continuous  
15 reaction conditions allow incorporation events to be captured in real time by fluor signal detection using laser excitation, an optical waveguide, and a CCD camera.

[0147] In certain embodiments, the single molecule real time (SMRT) DNA sequencing methods using zero-mode waveguides (ZMWs) developed by Pacific Biosciences, or similar methods, are employed. With this technology, DNA sequencing is performed on SMRT chips,  
20 each containing thousands of zero-mode waveguides (ZMWs). A ZMW is a hole, tens of nanometers in diameter, fabricated in a 100 nm metal film deposited on a silicon dioxide substrate. Each ZMW becomes a nanophotonic visualization chamber providing a detection volume of just 20 zeptoliters ( $10^{-21}$  L). At this volume, the activity of a single molecule can be detected amongst a background of thousands of labeled nucleotides. The ZMW provides a  
25 window for watching DNA polymerase as it performs sequencing by synthesis. Within each chamber, a single DNA polymerase molecule is attached to the bottom surface such that it permanently resides within the detection volume. Phospholinked nucleotides, each type labeled with a different colored fluorophore, are then introduced into the reaction solution at high concentrations which promote enzyme speed, accuracy, and processivity. Due to the small size  
30 of the ZMW, even at these high concentrations, the detection volume is occupied by nucleotides

only a small fraction of the time. In addition, visits to the detection volume are fast, lasting only a few microseconds, due to the very small distance that diffusion has to carry the nucleotides. The result is a very low background.

[0148] Processes and systems for such real time sequencing that may be adapted for use with the invention are described in, for example, U.S. Pat. Nos. 7,405,281; 7,315,019; 7,313,308; 7,302,146; and 7,170,050; and U.S. Pat. Pub. Nos. 2008/0212960; 2008/0206764; 2008/0199932; 2008/0199874; 2008/0176769; 2008/0176316; 2008/0176241; 2008/0165346; 2008/0160531; 2008/0157005; 2008/0153100; 2008/0153095; 2008/0152281; 2008/0152280; 2008/0145278; 2008/0128627; 2008/0108082; 2008/0095488; 2008/0080059; 2008/0050747; 2008/0032301; 2008/0030628; 2008/0009007; 2007/0238679; 2007/0231804; 2007/0206187; 2007/0196846; 2007/0188750; 2007/0161017; 2007/0141598; 2007/0134128; 2007/0128133; 2007/0077564; 2007/0072196; and 2007/0036511; and Koriach *et al.* (2008) "Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures" PNAS 105(4): 1176-81, all of which are herein incorporated by reference in their entireties.

[0149] Upon competition of sequencing, sequences can be sorted by same barcode, wherein sequences having the same barcode came from the same partition and thus are contiguous. In some embodiments, sequences linked based on common barcode sequence can be determined and optionally SNPs can be detected per fragment per barcode. In some embodiments, one can detect fragment colocalization to a single barcode greater than chance (skewed distributions), thereby detecting a rearrangement.

### Example

[0150] The following examples are offered to illustrate, but not to limit, the claimed invention.

[0129] An experiment demonstrating CPTseq in droplets was performed. The workflow provided in Figure 30A covered 5 conditions tested for the DNA aqueous inlet: (1) Beads, (2-3) beads + P5 PCR primer (x2 replicates), (4) N70X only, and (5) no template control (NTC). The method protocol is provided below. Briefly, the DNA was tagmented (with Tagment DNA enzyme (TDE)) in a tube to generate CPT DNA, five emulsions made using that CPT DNA, PCR reagents and a variety of oligos that vary per condition. In the droplet, the oligos were released

from the beads for the bead conditions, the Tn5 displaced from the CPTDNA, the gaps filled and PCR performed for 10 cycles. The droplets were then broken, the DNA purified, and a second nested PCR was performed followed by a final purification. The products were run on a gel and results are illustrated in Figure 30B. The test conditions using the beads are shown for conditions 1-3. The absence of amplification for the negative control in condition 4 omitting the bead and all bead primers, but including the N70X primer, argues that the amplification in conditions 1-3 is bead dependent. The NTC showing no amplification was confirmed for condition 5.

Sequencing methods on Illumina sequencers for the bead sample is shown in Figure 30C together with the sequencing QC metrics. High cluster densities and the majority of bases with Q scores of greater than 30 argue that sequencing was successfully executed.

[0129] Figure 31 shows expected whole genome analysis and phasing data from the data sets produced by the sequencing run shown in Figure 30A-C, in the CPTseq in droplets expected results. Beads in bulk data (Zhang et al 2017 *Nature Biotechnology*) and 10X data (Zheng et al 2017 *Nature Biotechnology*) are provided for comparison.

[0151] Detailed protocol:

[0152] Equipment: Thermocycler (either C1000 touch or T100)

Tagment reaction (150 haploid genomes/uL)

Component	100	total volume						
	Part number	LOT NUMBER	Supplier	stock	unit	final	unit	MM (uL)
TD buffer				2	X	1	X	50.0
DNA, diluted in 10mM TRIS, pH8				1.25	ng/uL	0.5	ng/uL	40.0
TDE Enzyme				10	U/uL	1	U/uL	10.0
Total								100.0

Tagment Reaction

[0153] Mix thoroughly by pipetting.

[0154] Incubate in a thermocycler (Lid temperature: 55C, Sample Volume: 100uL).

55°C 10 minutes

4°C

[0155] Keep on ice. At this concentration of DNA, 150 haploid genomes/uL (for use in enzyme mix).

[0156] Oligos

[0157] P5-short AATGATACGGCGAC

[0158] P7-short CAAGCAGAAGACGG

5 [0159] N70X CAAGCAGAAGACGGCATAACGAGATXXXXXXXXXXGTCTCGTGGGCTCGG

[0160] Enzyme Mix:

	25	volume per reaction						
	8	total reactions						
Component	Part number	Supplier	stock	unit	final (2x)	unit	for 1 rxn	MM
Water	AM9937	Ambion	N/A	N/A	N/A	N/A	8.94	71.53
Drop Solution	N/A	Bio-Rad	15	x	1	x	1.67	13.33
Optiprep	D1556	Sigma	100	%	3	%	0.75	6.00
USER	M5505	NEB	1	U/uL	0.125	U/uL	3.13	25.00
DTT	N/A	Bio-Rad	300	mM	10	mM	0.83	6.67
HS Q5 Enzyme, concentrated	custom	NEB	10	U/uL	0.04	U/uL	0.10	0.80
HS Q5 Enzyme	M0491L	NEB	2	U/uL	0.02	U/uL	0.25	2.00
KOD Xtreme Hot Start DNA Pol	71975	Millipore/Sigma	1	U/uL	0.04	U/uL	1.00	8.00
DNA			150	hap genomes/uL	50	hap genomes/uL	8.33	66.67
Total							25.00	200.00

[0161] DNA Mix

	25	volume per reaction						
	8	total reactions						
Component	Part number	Supplier	stock	unit	final (2x)	unit	for 1 rxn	MM
Water	AM9937	Ambion	N/A	N/A	N/A	N/A	0.33	2.67
Drop	N/A	Bio-Rad	15	x	1	x	1.67	13.33

Solution								
Optiprep	D1556	Sigma	100	%	28	%	7.00	56.00
P5 short*	N/A	IDT	100000	nM	2000	nM	0.50	4.00
dNTPs			10	mM	0.8	mM	2.00	16.00
1 step beads (Full P5 adaptered)*	N/A	Bio-Rad	3200	beads/uL	1600	beads/uL	12.50	100.00
add 1uL to each sample after aliquoting mastermix								
N70X primer	N/A	IDT	50000	nM	2000	nM	1.00	8.00
Total							25.00	200.00

**[0162] Enzyme/DNA Mix:**

[0163] Make Enzyme mix according to table above.

[0164] Make DNA mix according to table above. Do not add N70X primer if multiplexing.

5 [0165] After aliquoting DNA mix, add 1uL of N70X oligo (50uM) to each sample. Keep both

[0166] DNA and Enzyme mix on ice until droplet making.

[0167] The above mixes can be modified. However, for successful drop generation, some constants are helpful.

10 [0168] Constants for successful drop generation:

1) Drop solution quantities

2) Optiprep concentrations

3) The amount of glycerol contained in the enzyme mixes

15 **[0169] Droplet Generation**

[0170] For all droplet generation and transfer steps, use Rainin pipettes and corresponding tips.

[0171] Rinse pipette tip with solution before loading chip. Depress pipette plunger only to first stop when loading cartridge to avoid bubbles

20 [0172] Insert cartridge into cartridge holder. Check that cartridge is fully inserted and lying flat against the bottom of the holder, then close the lever.

[0173] Using a reservoir and a P20 multichannel pipette, add 20uL of prime solution to each well of the second row of the cartridge.

[0174] Allow prime solution to remain in wells for 1 minute, then remove all solution with a multichannel pipette. Do not allow prime solution to remain in wells longer than 3 minutes.

25 [0175] Mix DNA solution by pipetting. Using a P20 single-channel pipette, load 20uL into the bottom of the B ports.

[0176] Mix Enzyme solution by pipetting. Using a P20 single-channel pipette, load 20uL into the bottom of the ports numbered 1-4. See FIG. 32.

[0177] Note: Do not leave any ports in the second row empty.

30 [0178] Pour EvaGreen oil into a reservoir. Using a P200 multi-channel pipette, load 80uL of oil into each well of the bottom row of the cartridge labeled OIL.

- [0179] Place loaded cartridge holder in ddSEQ Single-Cell Isolator, press silver button of top of machine to close.
- [0180] Once machine is finished running (all 3 indicator lights are solid green), remove cartridge holder.
- 5 [0181] Using a P50 multichannel pipette set at 43uL with Rainin tips, gently and slowly aspirate all encapsulated sample (~40uL) from the output wells. Transfer to a 96-well PCR plate.
- [0182] PCR: Incubate in a thermocycler (Lid temperature: 105C, Sample Volume: 50uL).
- [0183] Note: Use either Bio-Rad C1000 Touch (Deep Well) or T100 Thermal Cycler
- 10 37°C 30 min
- 72°C 5 min
- 98°C 30 sec
- 37°C
- 72°C 10 sec
- 98°C 30 sec
- 15 37°C 120 sec
- 72°C
- 98°C 5 min
- 4°C
- 20 [0184] **Emulsion Breakage and Cleanup – if pooling chips**
- [0185] Add 10uL of Droplet Disruptor to each sample (2 wells/sample)
- [0186] Add 40uL of water. Do not mix.
- [0187] Add 70uL Ampure beads (1X cleanup).
- [0188] Pipette mix in the aqueous layer only until the beads are evenly distributed. Do not mix
- 25 the aqueous layer with the oil layer at the bottom of the well.
- [0189] After mixing, each sample should have 2 distinct layers: an oil layer at the bottom of the well and a homogenous brown aqueous layer on top.
- [0190] Incubate mixture at room temperature for 5 minutes
- [0191] Place on magnetic stand and wait 5 minutes.
- 30 [0192] Keeping plate on the magnetic stand, remove and discard the supernatant, including oil.
- [0193] Wash 2 times on the magnet as follows:
- Add 200uL freshly prepared 80% EtOH to each well.
- Incubate for 30 seconds
- Remove and discard all supernatant from each well.
- 35 [0194] Using a P20 pipette, remove residual 80% EtOH from each well.
- [0195] Air-dry on the magnetic stand for ~10 minutes. Dry time can vary depending on temperature/humidity of room. Do not overdry pellet.
- [0196] Remove plate from the magnetic stand. Add 10uL RSB to each sample well. Pipette to mix.
- 40 [0197] Incubate at room temperature (not on magnetic stand) for 2 minutes.
- [0198] Place tube on magnetic stand, wait ~2 minutes or until solution is clear
- [0199] Combine the 2 wells for each sample into a single well by transferring 10uL of supernatant from each sample well to a new plate.
- [0200] The final volume for each sample (now combined) should be 20uL. Run 1uL on HS DNA
- 45 chip.

[0201] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, one of skill in the art will appreciate that certain changes and modifications may be practiced within the scope of the appended claims. In addition, each reference provided herein is incorporated by reference in its entirety to the same extent as if each reference was individually incorporated by reference. 5 Where a conflict exists between the instant application and a reference provided herein, the instant application shall dominate.



WHAT IS CLAIMED IS:

- 1                   1.       A method of determining haplotype genomic sequence information, the  
2 method comprising,  
3                   providing fragments of genomic DNA;  
4                   reacting the fragments with an adaptor-loaded tagmentase that creates DNA  
5 fragments defined by breakpoints in the fragments and inserts an adaptor at the break points,  
6 wherein the reacting is under conditions such that the tagmentase binds the breakpoints to form  
7 linked DNA segments in the form of DNA segment-first adaptor-tagmentase-second adaptor-  
8 (DNA segment-first adaptor-tagmentase-second adaptor)<sub>n</sub>-DNA segment, where n is any integer  
9 and "-" indicates a covalent linkage;  
10                  encapsulating the linked DNA segments into partitions, said partitions  
11 comprising:  
12                   a bead, the bead having a forward primer oligonucleotide linked to the  
13 bead by a 5' end of the forward primer oligonucleotide, the forward primer oligonucleotide  
14 having a bead-specific barcode and a 3' end specific for and complementary to the first or second  
15 adaptor;  
16                   a reverse primer oligonucleotide having a 3' end complementary to the  
17 first or second adaptor, wherein the forward primer 3' end and the reverse primer 3' end are  
18 complementary to different adaptors selected from the first adaptor and the second adaptor;  
19                   displacing the tagmentase from the segments in the partitions;  
20                   performing amplification wherein the forward primer and reverse primer  
21 oligonucleotide generate amplicons from the DNA segments, such that amplicons within a  
22 partition are barcoded with the bead barcode;  
23                   combining the partitions to form a reaction mixture containing the amplicons; and  
24                   nucleotide sequencing the amplicons.
- 1                   2.       The method of claim 1, wherein the partitions comprise an amount of  
2 agent that displaces the tagmentase from the segments but does not inhibit polymerization.
- 1                   3.       The method of claim 1, wherein prior to amplification, the single stranded  
2 regions of inserted adaptors are filled in by DNA polymerase.

- 1           4.       The method of claim 1, wherein prior to amplification, single stranded  
2 regions of fragmented target nucleic acids are filled in by DNA polymerase.
- 1           5.       The method of claim 1, wherein the forward primer oligonucleotide is  
2 released from the bead and amplification occurs in solution.
- 1           6.       The method of claim 1, wherein the agent is a detergent.
- 1           7.       The method of claim 6, wherein the detergent is sodium dodecyl sulfate  
2 (SDS).
- 1           8.       The method of claim 7, wherein the concentration of SDS is 0.005-0.05%  
2 (e.g., 0.01-0.04%, e.g., 0.01-0.02%).
- 1           9.       The method of claim 1, wherein the fragments are on average between 5-  
2 10 Mb.
- 1           10.      The method of claim 1, wherein the partitions are droplets in an emulsion.
- 1           11.      The method of claim 1, wherein the encapsulating encapsulates on average  
2 0.02-3 (e.g., 0.05-1, 0.08-0.5, e.g., 0.1, 1, 2, or 3) beads into partitions.
- 1           12.      The method of claim 1, wherein the genomic DNA is from a single cell.
- 1           13.      The method of claim 1, wherein the genomic DNA is from a mammal or  
2 plant.
- 1           14.      The method of claim 1, wherein the first adaptor and the second adaptor  
2 have different sequences.
- 1           15.      The method of claim 14, wherein the first adaptor and the second adaptor  
2 are less than 50% identical.
- 1           16.      The method of claim 1, wherein the first adaptor and the second adaptor  
2 have identical sequences.

1           17.     The method of claim 1, wherein the first adaptor and the second adaptor  
2 are linked by a linking sequence when loaded on the tagmentase such the DNA segments are  
3 linked by both the tagmentase and the linking sequence.

1           18.     The method of claim 17, wherein the linking sequence comprises a  
2 restriction recognition sequence and the linking sequence is cleaved by a restriction enzyme after  
3 the encapsulating and before the performing.

1           19.     The method of claim 17, wherein the linking sequence comprises one or  
2 more uracils and the linking sequence is cleaved by a uracil-DNA N-glycosylase after the  
3 encapsulating and before the performing.

1           20.     The method of claim 17, wherein the linking sequence comprises one or  
2 more ribonucleotide and the linking sequence is cleaved at the ribonucleotide after the  
3 encapsulating and before the performing.

1           21.     The method of claim 1, wherein n is an integer selected from 0-10,000.

1           22.     The method of claim 1, wherein the length of the DNA segment-first  
2 adaptor-tagmentase-second adaptor-(DNA segment-first adaptor-tagmentase-second adaptor)<sub>n</sub>-  
3 DNA segment is between 5 kb-10Mb, e.g., 1Mb-10Mb.

1           23.     The method of claim 1, wherein at least 10,000 different linked DNA  
2 segments are encapsulated into different partitions.

1           24.     A plurality of partitions, the partitions comprising:  
2                   a bead, the bead having a forward primer oligonucleotide linked to the  
3 bead by a 5' end of the forward primer oligonucleotide, the forward primer oligonucleotide  
4 having a bead-specific barcode and a 3' end specific for and complementary to a first or second  
5 adaptor;  
6                   a reverse primer oligonucleotide having a 3' end complementary to the first or  
7 second adaptor, wherein the forward primer 3' end and the reverse primer 3' end are  
8 complementary to different adaptors selected from the first adaptor and the second adaptor; and

9 an amount of agent that displaces tagmentase from DNA but does not inhibit  
10 polymerization.

1 25. The plurality of partitions, wherein the partitions further comprise:  
2 tagmentase; and  
3 genomic DNA, wherein segments of the DNA having ends defined by adaptors  
4 inserted at breakpoints created by the tagmentase in the form first adaptor-DNA segment-second  
5 adaptor.

1 26. The plurality of partitions of claim 24, wherein the agent is a detergent.

1 27. The plurality of partitions of claim 26, wherein the detergent is sodium  
2 dodecyl sulfate (SDS).

1 28. The plurality of partitions of claim 27, wherein the concentration of SDS  
2 is 0.005-0.05% (e.g., 0.01-0.04%, e.g., 0.01-0.02%).

1 29. The plurality of partitions of claim 24, wherein the partitions are droplets  
2 in an emulsion.

1 30. The plurality of partitions of claim 24, wherein the partitions comprise on  
2 average 0.02-3 (e.g., 0.05-1, 0.08-0.5, e.g., 0.1, 1, 2, or 3) beads.

1 31. The plurality of partitions of claim 24, wherein the genomic DNA is from  
2 a single cell.

1 32. The plurality of partitions of claim 24, wherein the genomic DNA is from  
2 a mammal or plant.

1 33. The plurality of partitions of claim 24, wherein the first adaptor and the  
2 second adaptor have different sequences.

1 34. The plurality of partitions of partitions of claim 33, wherein the first  
2 adaptor and the second adaptor are less than 50% identical.

1           35.     The plurality of partitions of claim 24, wherein the first adaptor and the  
2 second adaptor have identical sequences.

1           36.     The plurality of partitions of claim 24, wherein the first adaptor and the  
2 second adaptor are linked by a linking sequence in a form comprising first adaptor-DNA  
3 segment-second adaptor-(linker sequence)-first adaptor-DNA segment-second adaptor)<sub>n</sub>, where n  
4 is any integer 1 or greater.

1           37.     The plurality of partitions of claim 36, wherein the linking sequence  
2 comprises a restriction recognition sequence and optionally the partition further comprises a  
3 restriction enzyme the restriction recognition sequence.

1           38.     The plurality of partitions of claim 36, wherein the linking sequence  
2 comprises one or more uracils and optionally the partition further comprises a uracil-DNA N-  
3 glycosylase.

1           39.     The plurality of partitions of claim 36, wherein the linking sequence  
2 comprises one or more ribonucleotide.

1           40.     The plurality of partitions of claim 36, wherein n is an integer selected  
2 from 1-10,000.

1           41.     The plurality of partitions of claim 36, wherein the length of the first  
2 adaptor-DNA segment-second adaptor-(linker sequence-first adaptor-DNA segment-second  
3 adaptor)<sub>n</sub> is between 5 kb-10Mb, e.g., 1Mb-10Mb.

1           42.     The plurality of partitions of any of claims 24-41, comprising at least  
2 10,000 different partitions.

1           43.     A plurality of partitions comprising DNA segments having contiguity  
2 maintained by tagmentase, the partitions comprising

3           linked DNA segments in the form of DNA segment-first adaptor-tagmentase-  
4 second adaptor-(DNA segment-first adaptor-tagmentase-second adaptor)<sub>n</sub>-DNA segment, where

5 n is any integer and “-” indicates a covalent linkage and wherein the linked DNA maintains  
6 contiguity compared to genomic DNA.

1 44. The plurality of partitions of claim 43, wherein the partitions are droplets  
2 in an emulsion.

1 45. The plurality of partitions of claim 43, wherein the partitions comprise on  
2 average 0.02-3 (e.g., 0.05-1, 0.08-0.5, e.g., 0.1, 1, 2, or 3) beads.

1 46. The plurality of partitions of claim 43, wherein the genomic DNA is from  
2 a single cell.

1 47. The plurality of partitions of claim 43, wherein the genomic DNA is from  
2 a mammal or plant.

1 48. The plurality of partitions of claim 43, wherein the first adaptor and the  
2 second adaptor have different sequences.

1 49. The plurality of partitions of partitions of claim 48, wherein the first  
2 adaptor and the second adaptor are less than 50% identical.

1 50. The plurality of partitions of claim 43, wherein the first adaptor and the  
2 second adaptor have identical sequences.

1 51. The plurality of partitions of claim 43, wherein the linking sequence  
2 comprises a restriction recognition sequence and optionally the partition further comprises a  
3 restriction enzyme the restriction recognition sequence.

1 52. The plurality of partitions of claim 43, wherein the linking sequence  
2 comprises one or more uracils and optionally the partition further comprises a uracil-DNA N-  
3 glycosylase.

1 53. The plurality of partitions of claim 43, wherein the linking sequence  
2 comprises one or more ribonucleotide.

1                   54.     The plurality of partitions of claim 43, wherein n is an integer selected  
2 from 1-10,000.

1                   55.     The plurality of partitions of claim 43, wherein the length of the first  
2 adaptor-DNA segment-second adaptor-(linker sequence-first adaptor-DNA segment-second  
3 adaptor)<sub>n</sub> is between 5 kb-10Mb, e.g., 1Mb-10Mb.

1                   56.     The plurality of partitions of any of claims 43-55, comprising at least  
2 10,000 different partitions.

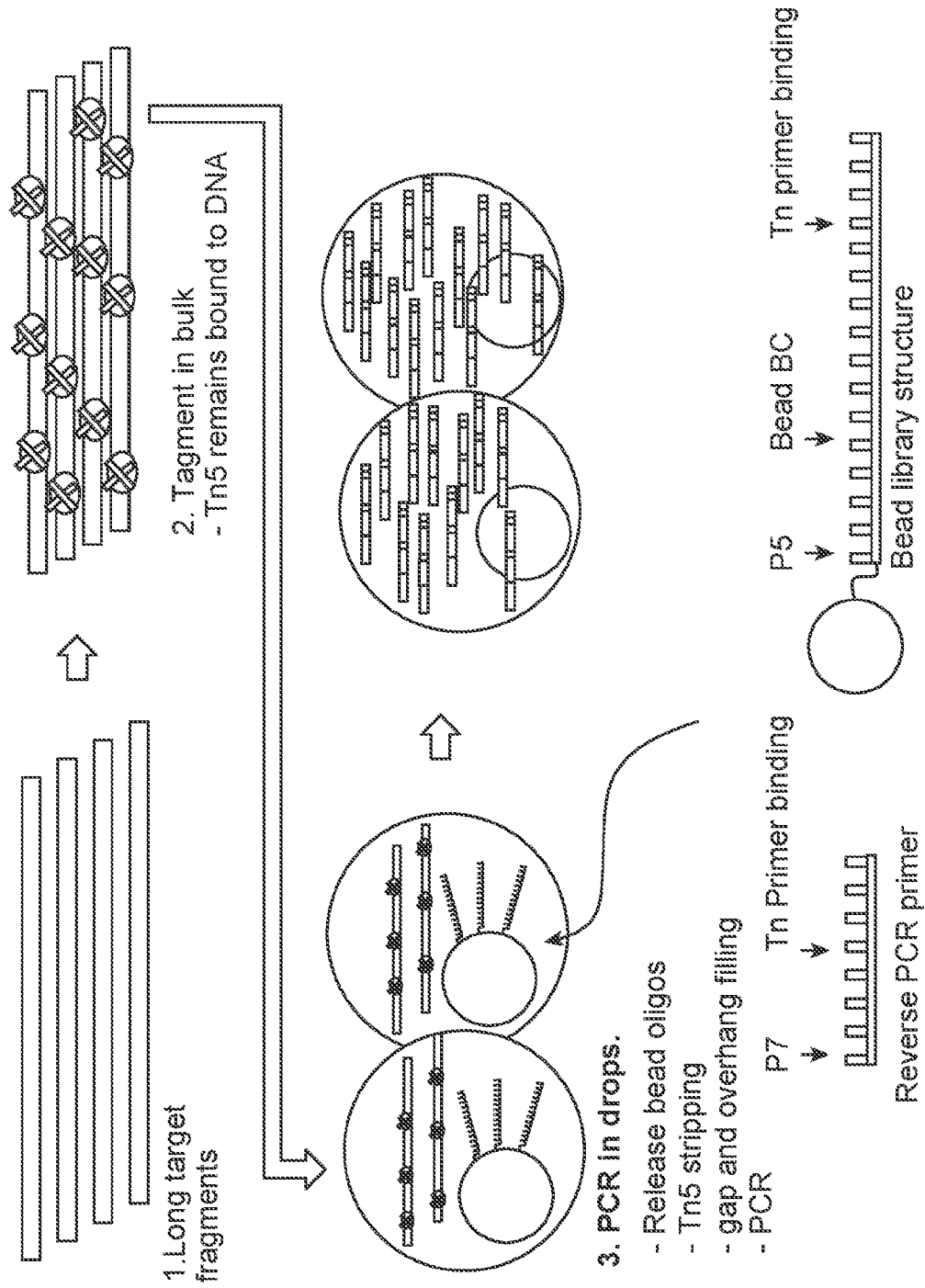


FIG. 1



5. Break emulsion, fragments pooled, sequence

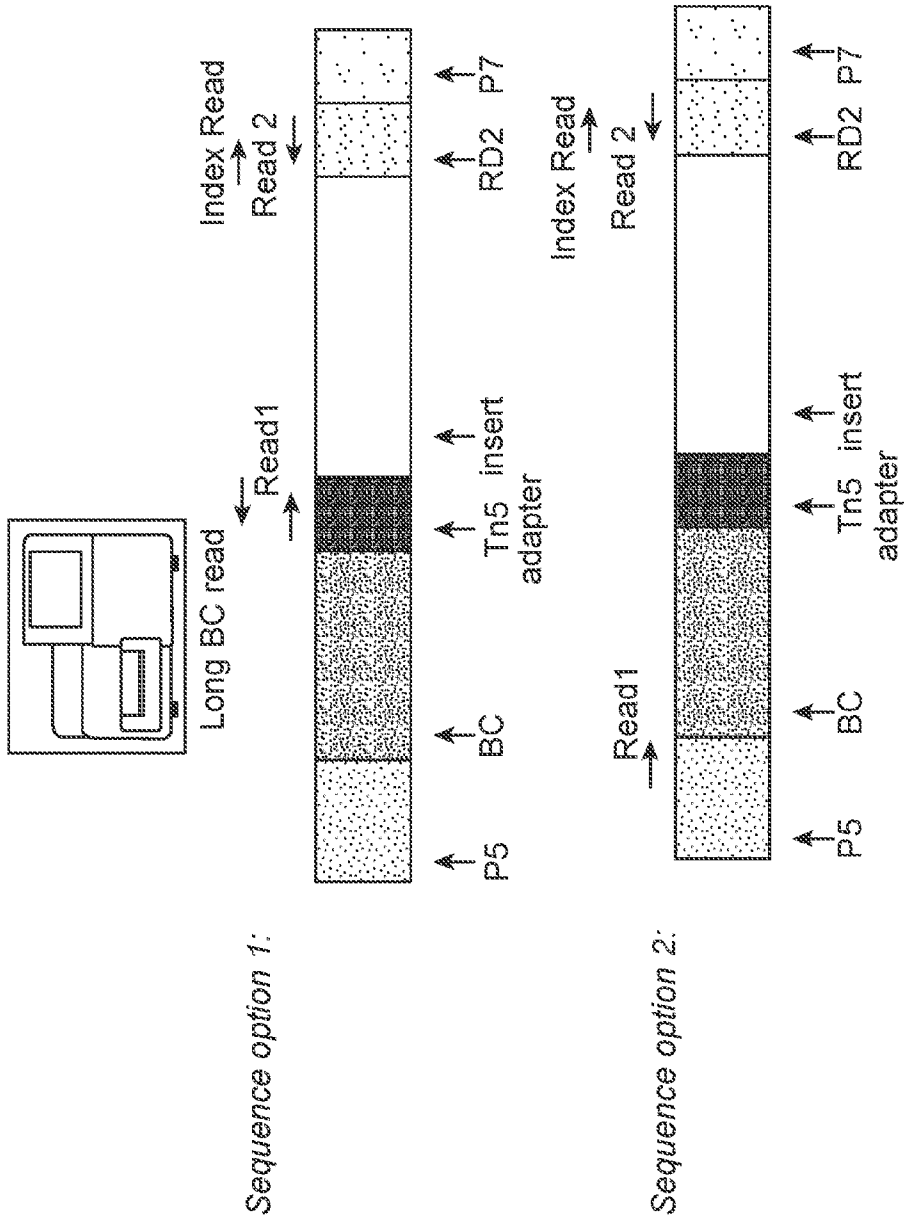


FIG. 2

1. Universal bead oligo and QC probe

Bead-**TTTTTTUUUUCTACAGCCCTGTCGGGGGAAAGCAGTGGTATCAACCCGACAGATAC**-3'  
**CATAGTTGGCTCICAT**-FAM

↙ Universal QC probe

2. Universal bead is split across 96-wells and incubated with BC block 1 (BC1) six different phase-shifting nucleotides (bold) to avoid in-phase clusters. "H" = defined hamming BC

Bead-**TTTTTTUUUUCTACAGCCCTGTCGGGGGAAAGCAGTGGTATCAACCCGACAGTAC**  
 TAGTTGGCTC**AT**CGHHHHHATCGGTAGCGTAACG  
 TAGTTGGCTC**AT**GCHHHHHATCGGTAGCGTAACG  
 TAGTTGGCTC**AT**GTHHHHHATCGGTAGCGTAACG  
 TAGTTGGCTC**AT**GAGHHHHATCGGTAGCGTAACG  
 TAGTTGGCTC**AT**GACHHHHHATCGGTAGCGTAACG  
 TAGTTGGCTC**AT**GATCCHHHHHATCGGTAGCGTAACG  
 TAGTTGGCTC**AT**GCGGAHHHHATCGGTAGCGTAACG

↙ BC Block 1 templates

3. Primer extension for BC block 1. Wash, pool

Bead-**TTTTTTUUUUCTACAGCCCTGTCGGGGGAAAGCAGTGGTATCAACCCGACAGTAC**[0-5]HHHHHTAGCCATCGCATTCG  
**TAGTTGGCTC**ATG[0-5]HHHHHATCGGTAGCGTAACG  
 3' -ATCGGTAGCGTAAC-FAM

↙ Block 1 QC probe

4. Split into 96 for Block 2 template oligos (96)

Bead-**TTTTTTUUUUCTACAGCCCTGTCGGGGGAAAGCAGTGGTATCAACCCGACAGTAC**[0-5]HHHHHTAGCCATCGCATTCG  
**A**TCGGTAGCGTAACGGHHHHHATGGTACTCGACTT

↙ BC Block 2 template

5. Primer extension, wash, Pool

Bead-**TTTTTTUUUUCTACAGCCCTGTCGGGGGAAAGCAGTGGTATCAACCCGACAGTAC**[0-5]HHHHHTAGCCATCGCATTCGHHHHHTACCAC**TGAGCTGAA**  
**A**TCGGTAGCGTAACGGHHHHHATGGTACTCGACTT  
 3' -ATGGTACTCGACTT-FAM

↙ Block 2 QC probe

FIG. 3

6. Split into 96 for block 3 template oligos (96)

...CGCAGAGTAC [0-5] HHHHHHTAGCCATCGCATTCGCHHHHHHTACCTCTGAGCTGAA  
ATGGAGACTCGACTTHHHHHHAGCAGCCGTCGCAG

BC Block 3 template, no UMI,  
NEXTERA Tn5 P5 side  
adapter seq

7. Primer extension, wash, pool.

...CGCAGAGTAC [0-5] HHHHHHTAGCCATCGCATTCGCHHHHHHTACCTCTGAGCTGA  
AHHHHHHHTCGTCGGCAGCGTC  
NEXTERA Tn5 P5 side adapter seq  
ATGGAGACTCGACTTHHHHHHAGCAGCCGTCGCAG

3' --AGCAGCCGTCGCAG

8. BC oligo on bead after synthesis --  
after cleavage 105-110 nt

Block 3 QC probe (we  
currently don't have)

Custom seq primer

Bead-TTTTTTUUUUCTACAGCCCTGTCGGGGAAGCAGTGGTATCAAGCCAGAGTAC [0-5] TCGTCGGCAGCGTCHHHHHHTAGCCATCGCATTCGCHHHHHHTACCCACTGAGCTGA  
AHHHHHH  
62-67 nt

FIG. 4

Primer definitions: Canonical nextera

14 bp P5 specific / 19 bp mosaic end / 15 bp P7 specific / 19 bp mosaic end /  
 (A14) (ME) (B15)  
 single stranded / double stranded / single stranded / double stranded /

Tn5 adapters

P5 side nextera adapter (what loaded onto tagmentase)

TCGTCGGCAGCGTCAGATGTGTATAAGACACAG  
LCLYCYCVLLVLLCLCLCLC

P7 side nextera adapter (what loaded onto tagmentase)

GTCTCGTGGGCTCGGAGATGTGTATAAGACACAG  
LCLYCYCVLLVLLCLCLCLC

N50x primer (PCR)

P5 grafting - INDEX- NEXTERA ADAPTER specific (A14)  
AATGATACGGCGACCCAGGAGATCTACAC [ 15 ] TCGTCGGCAGCGTC

N70x primer (PCR)

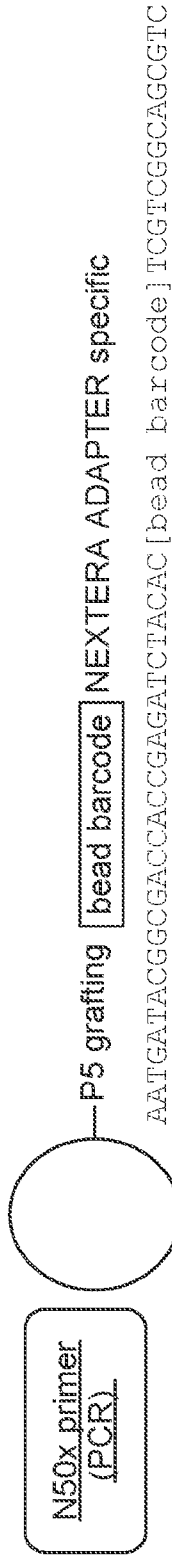
P7 grafting - INDEX- NEXTERA ADAPTER specific (B15)  
CAAGCAGAGACCGCATACGAGAT [ 17 ] GTCTCGTGGGCTCGG

FIG. 5

N50x primer (PCR) reminder from previous  
 AATGATACGGCCGACCACCGAGATCTACAC [1-5] TCGTCGGCAGCGTC

*Proposal for CPT seq in drops: Swap N50x primer (PCR) with gel bead oligos*

*Option 1:*



*Option 2:*

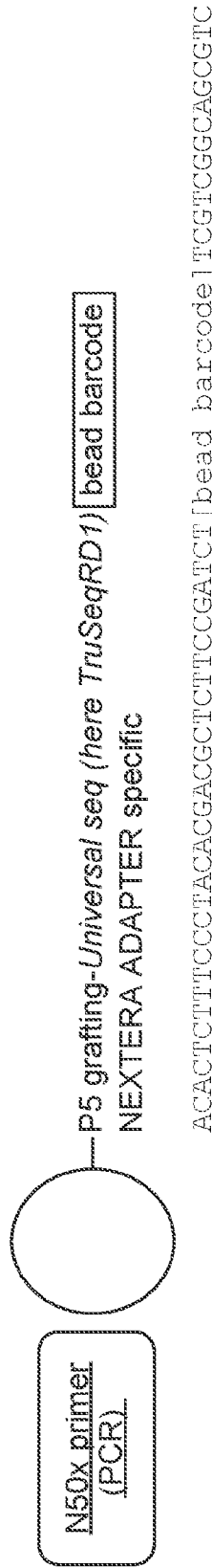


FIG. 6

N70x primer (PCR) reminder from previous  
 CAAGCAGAAGACGGGCATACGAGAT [17] GTCTCGTGGGCTCGG

*Proposal for CPT seq in drops: Swap N70X primer (PCR) with gel bead oligos*

Option 3:



Option 4:

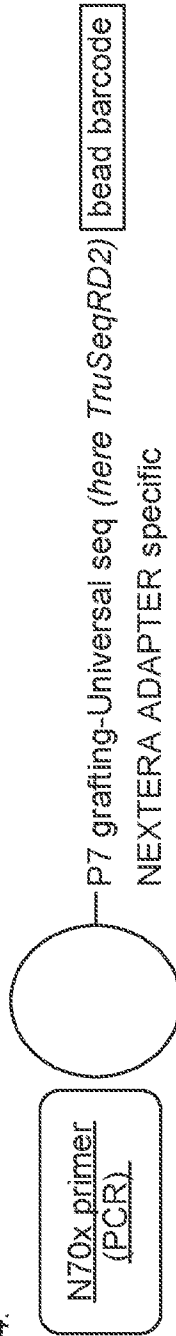


FIG. 7

Heteroadapter Tn5 biochemistry, bead option 1 and sequencing

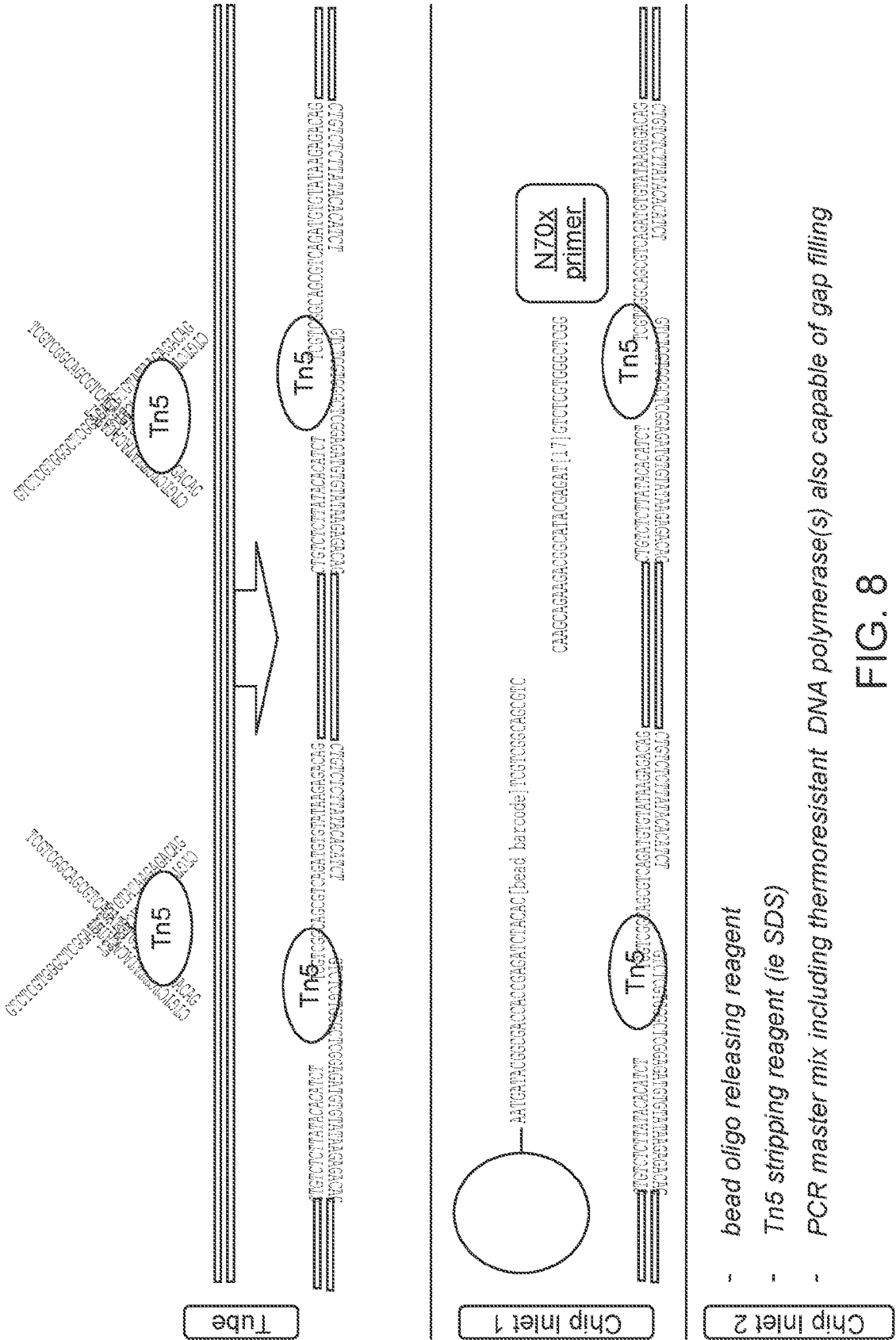


FIG. 8





**Heteroadapter Tn5 biochemistry, bead option 2 (N70X solution primer used in addition to this bead) and sequencing**

(tagmentation and inlets not shown)

- after oligo release, Tn5 stripping gap filling and PCR (only top strand shown)

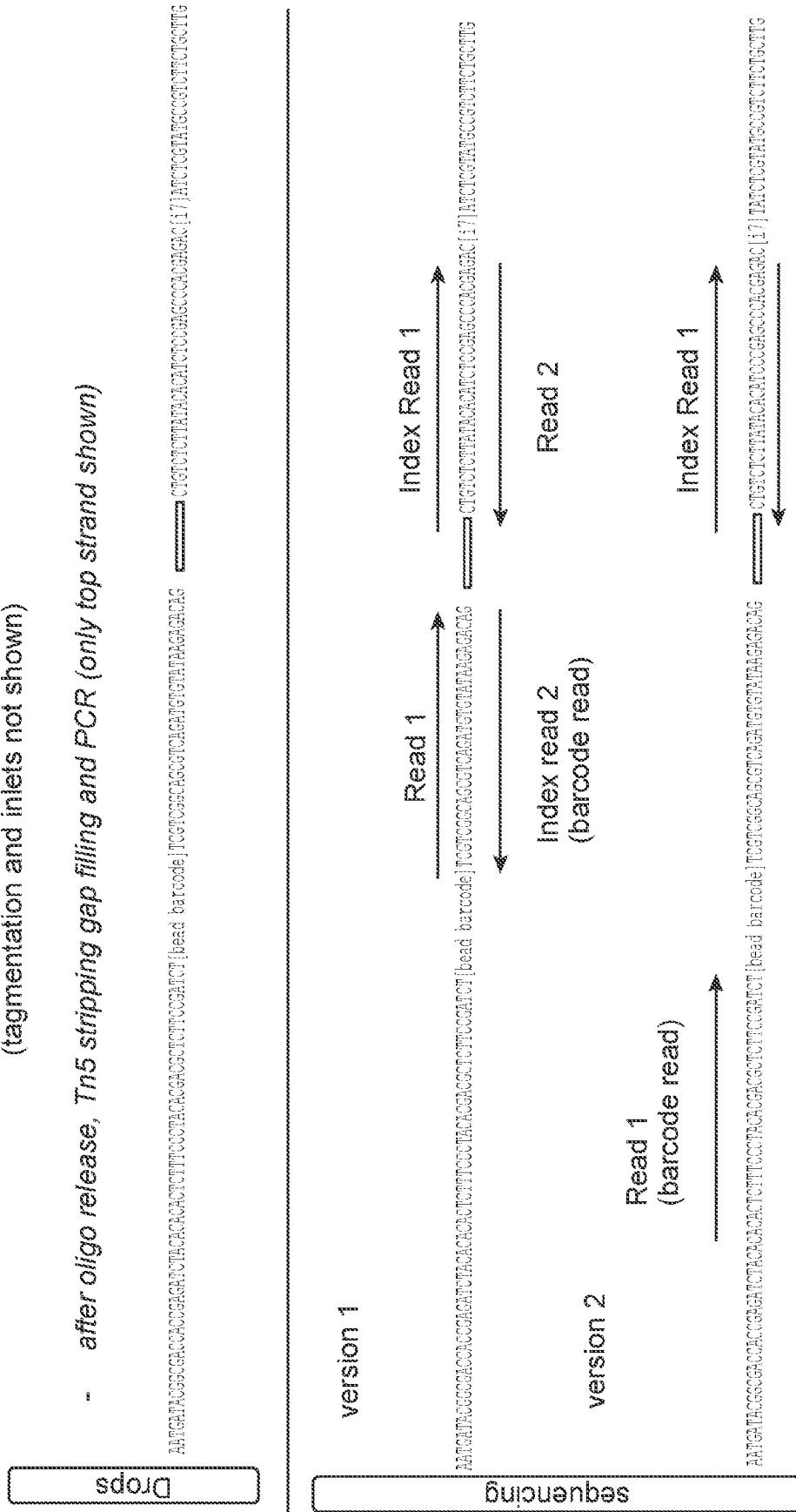


FIG. 10

Heteroadapter Tn5 biochemistry, bead option 3 (N50X solution primer used in addition to this bead) and sequencing

(tagmentation and inlets not shown)

- after oligo release, Tn5 stripping gap filling and PCR (only top strand shown)

Drops

AAATGATACGGGACCCGAGATCTACAC [15] TCGTCGGCAGCGTCGATGTCTTAAGGACAGC [barcode] ATCTCTTATACACATCTCCGAGCCCGCAGC [bead barcode] ATCTCTTATACACATCTCCGAGCCCGCAGC [bead barcode] ATCTCTTATACACATCTCCGAGCCCGCAGC [bead barcode]

sequencing

Read 1

Index Read 1 (barcode read)

Index read 2

Read 2

AAATGATACGGGACCCGAGATCTACAC [15] TCGTCGGCAGCGTCGATGTCTTAAGGACAGC [barcode] ATCTCTTATACACATCTCCGAGCCCGCAGC [bead barcode] ATCTCTTATACACATCTCCGAGCCCGCAGC [bead barcode] ATCTCTTATACACATCTCCGAGCCCGCAGC [bead barcode]

FIG. 11



Additional Primer definitions: for homoadapter Tn5 chemistry

Canonical

N70x primer  
(PCR)

P7 grafting - INDEX-NEXTERA ADAPTER specific

CAAGCAGAAGACGGCATACGAGAT [17] GTCTCGTGGGCTCGG

N50x primer  
(PCR)

P5 grafting - INDEX NEXTERA ADAPTER specific

AATGATACGGGACCACCGAGATCTACAC [15] TCGTCGGCAGCGTC

N70x primer-  
Truseq RD2  
(PCR)

P7 grafting - INDEX- Universal seq (here *TruSeqRD2*) NEXTERA ADAPTER specific

CAAGCAGAAGACGGCATACGAGAT [17] GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCGGCAGCGTC

Modified

N50x primer  
Truseq RD1  
(PCR)

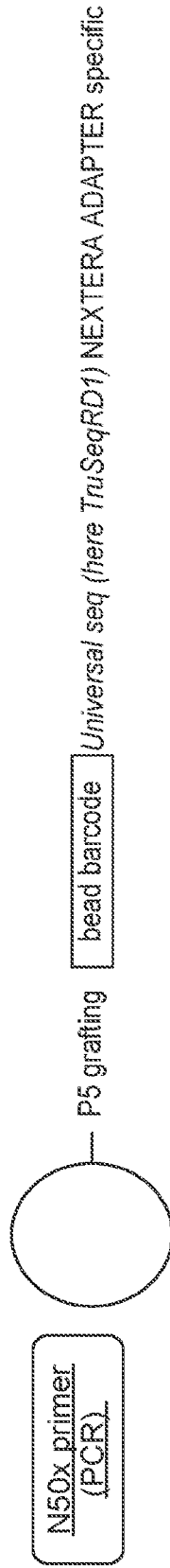
P5 grafting - INDEX- Universal seq (here *TruSeqRD1*) NEXTERA ADAPTER specific

AATGATACGGGACCACCGAGATCTACAC [15] ACACHTCTTCCCTACACGACGCTCTTCCGATCTCTCGTGGGCTCGG

FIG. 13

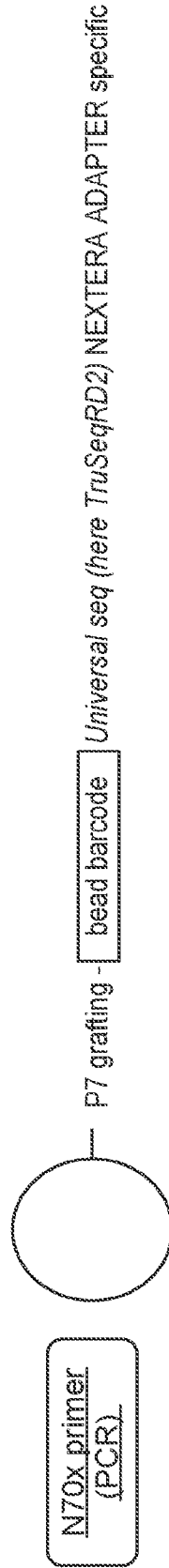
**Bead related primer definitions for homoadapter chemistry**

*Bead option 5 (compared to bead option 2, the universal sequence is moved 3' of the bead barcode:*



AATGATACGGGACCACCGAGATCTACAC [bead barcode] ACACTCCTTCCCTACACGACGCTCTTCCGATCTTCGGCAGCGTC

*Bead option 6 (compared to bead option 4, the universal sequence is moved 3' of the bead barcode:*



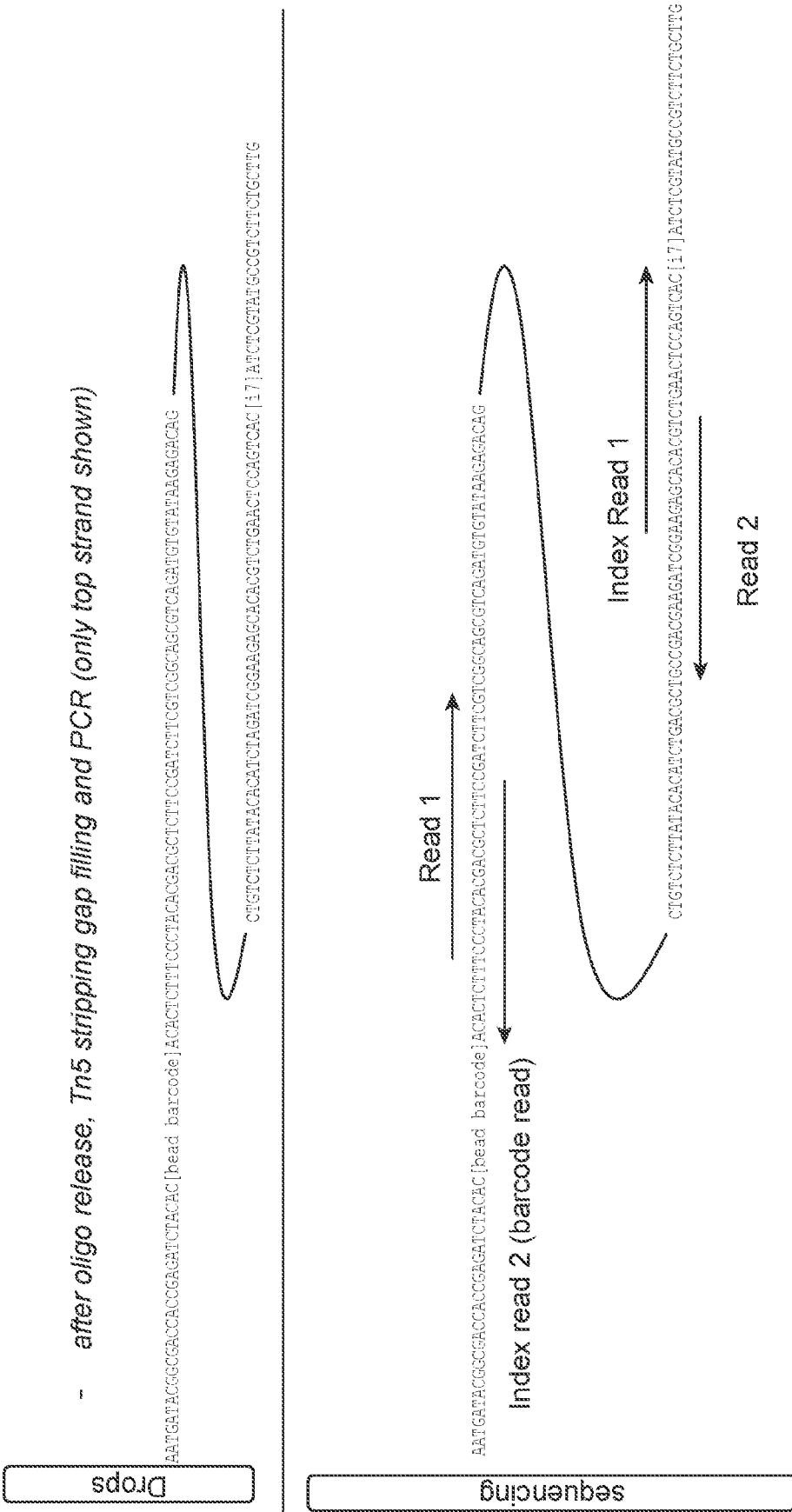
CAAGCAGAAGACGGCATACGAGAT [bead barcode] GTGACTGGAGTTCAGACGGTGTGCTCTTCCGATCTTCGGCAGCGTC

**FIG. 14**



Homoadapter Tn5 biochemistry, bead option 5 and sequencing

- after oligo release, Tn5 stripping gap filling and PCR (only top strand shown)



NB: Depending on where the sequencing primer can sit and still be specific to an end, there could be up to 19 bp of dark cycle sequencing (loss) to be able to get to the genomic DNA – this is the red common nextera part

FIG. 16





Homoadapter Tn5 biochemistry, bead option 6 and sequencing

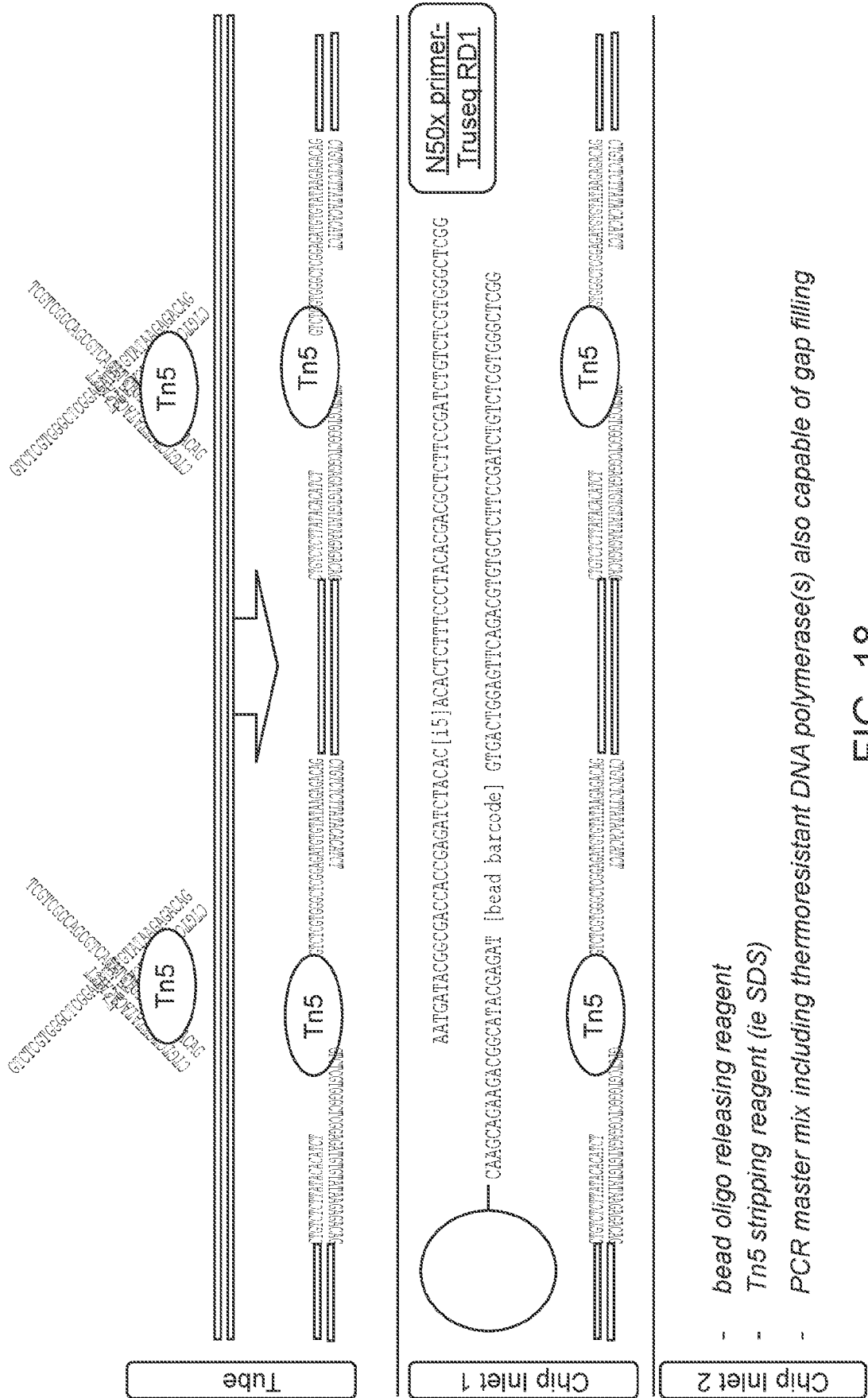


FIG. 18

**Homoadapter Tn5 biochemistry, bead option 6 and sequencing**

Drops

- after oligo release, Tn5 stripping gap filling and PCR (only top strand shown)

AATGATAGGGGACACCGGAGATCTACAC [15] ACACCTCTTCCCTACACGAGGCTCTCCGATCTCTCGTGGGCTGGAGATGCTATAAGACACAG

CTGTCTTTATACACATCTCCGAGCCGAGACAGATCGGAGAGCACACCTCTGAACTCCAGTCAC [bead barcode] ATCTCGTATCCCGTCTTCTGCTTG

Read 1

AATGATAGGGGACACCGGAGATCTACAC [15] ACACCTCTTCCCTACACGAGGCTCTCCGATCTCTCGTGGGCTGGAGATGCTATAAGACACAG

Index read 2

Index Read 1 (barcode read)

CTGTCTTTATACACATCTCCGAGCCGAGACAGATCGGAGAGCACACCTCTGAACTCCAGTCAC [bead barcode] ATCTCGTATCCCGTCTTCTGCTTG

Read 2

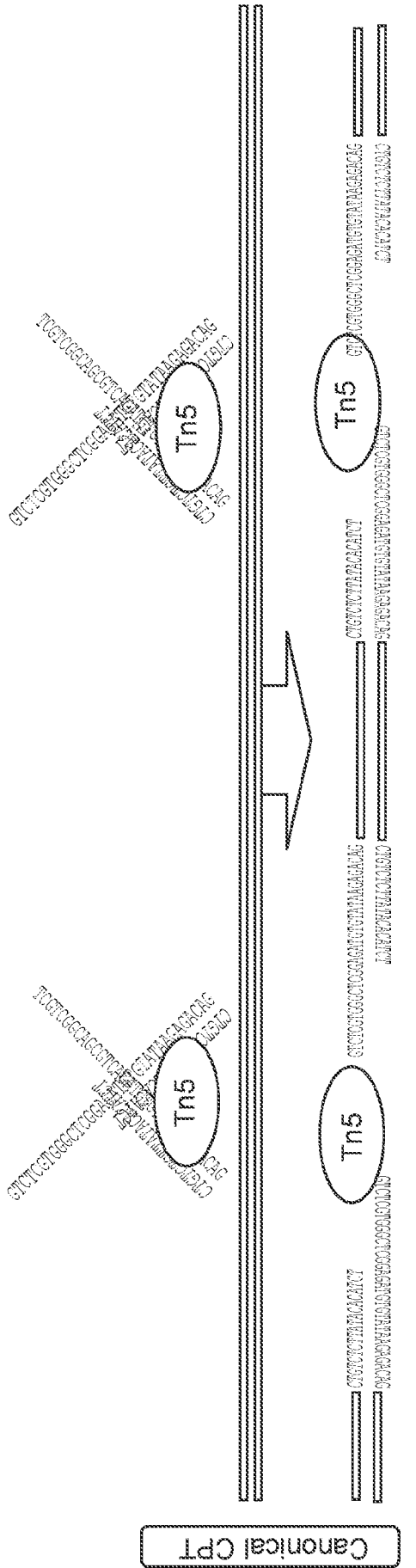
Sequencing

NB: Depending on where the sequencing primer can sit and still be specific to an end, there could be up to 19 bp of dark cycle sequencing (loss) to be able to get to the genomic DNA –this is the red common nextera part

FIG. 19



Option to preserve contiguity (1)



What Tn5 does in nature

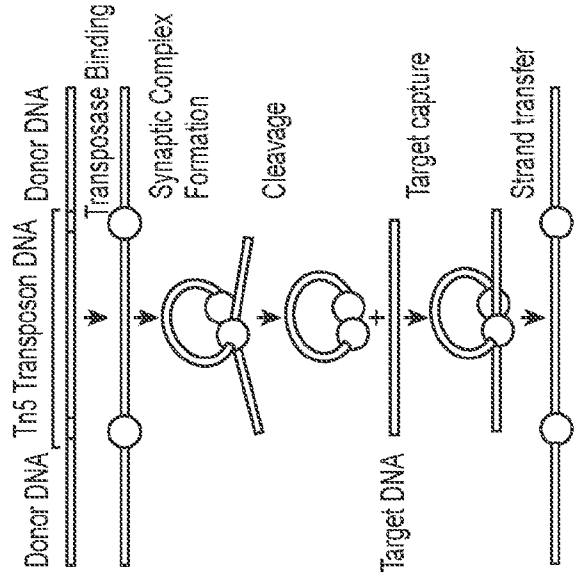


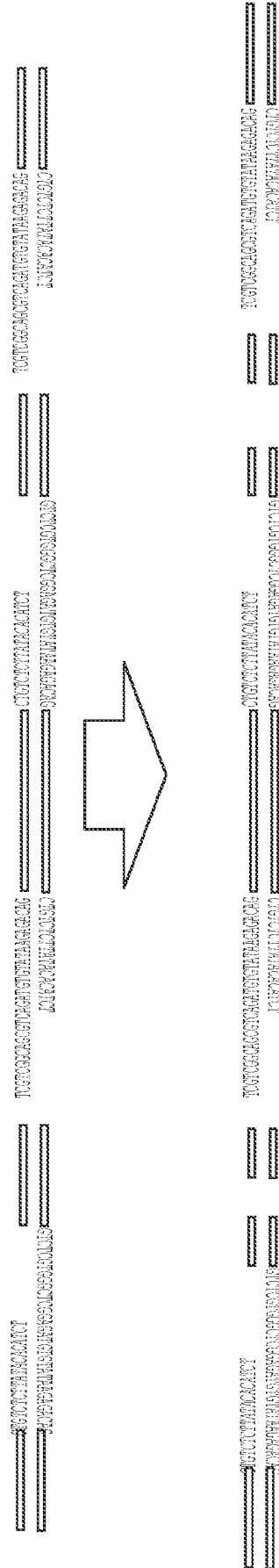
FIG. 21



Option to preserve contiguity (3): Preservation through DNA linked adapters

In drops, the first step is to cut the Tn5 adapter linkers

*In drops*



Options:

- Restriction enzyme cutting
- dUTP cleavage (provided that original linker had Us)
- RNA cleavage (provided that original linker had RNA)

NB: Sequence and type of nt for linker can be chosen

FIG. 23

Shortened mosaic end only Nextera homoadapter

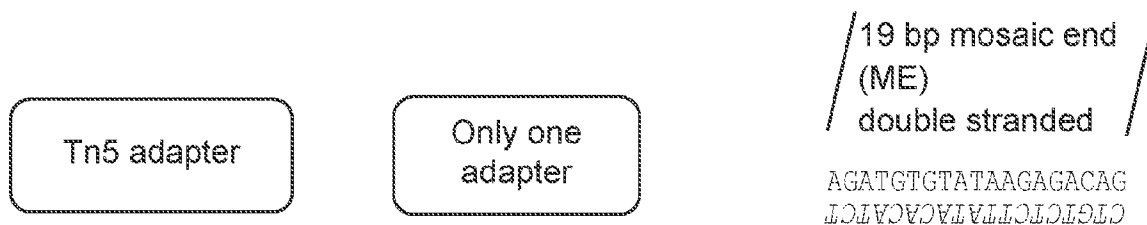


FIG. 24

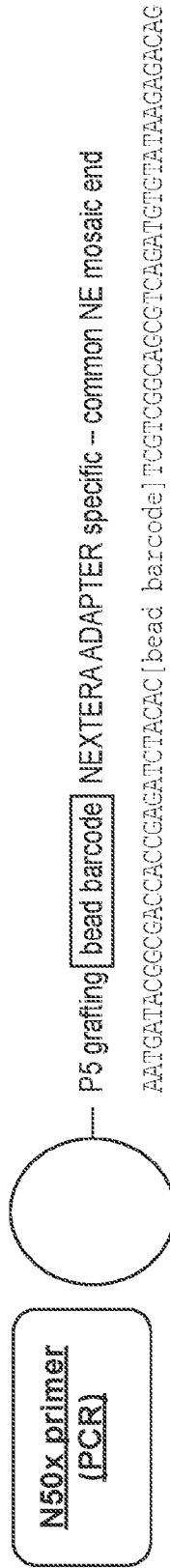
**N50x primer (PCR)**

**P5 grafting - INDEX- NEXTERA ADAPTER specific (A14) - common NE mosaic end**  
 AATGATACGGCGACCACCGAGATCTACAC [15] TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

**N70x primer (PCR)**

**P7 grafting - INDEX- NEXTERA ADAPTER specific (B15) - common NE mosaic end**  
 CAAGCAGAAGACGGCATACGAGAT [17] GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Option 7:



Option 8:

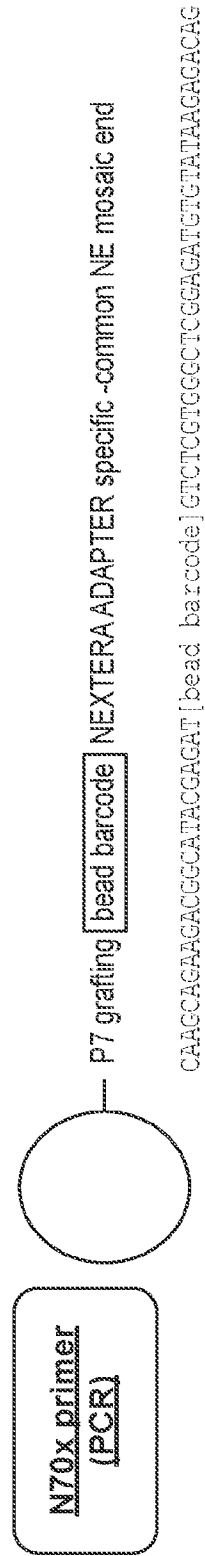
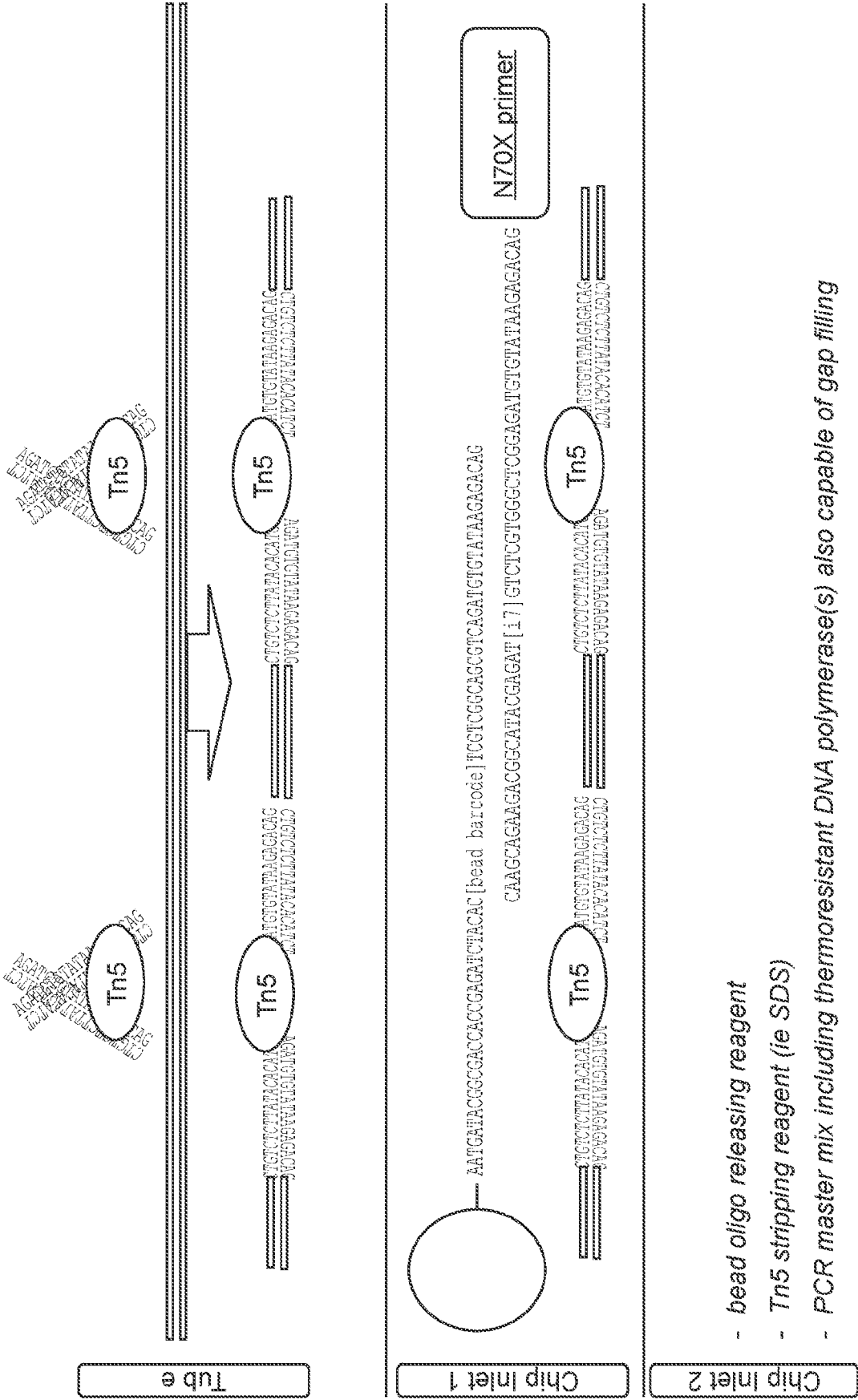


FIG. 25



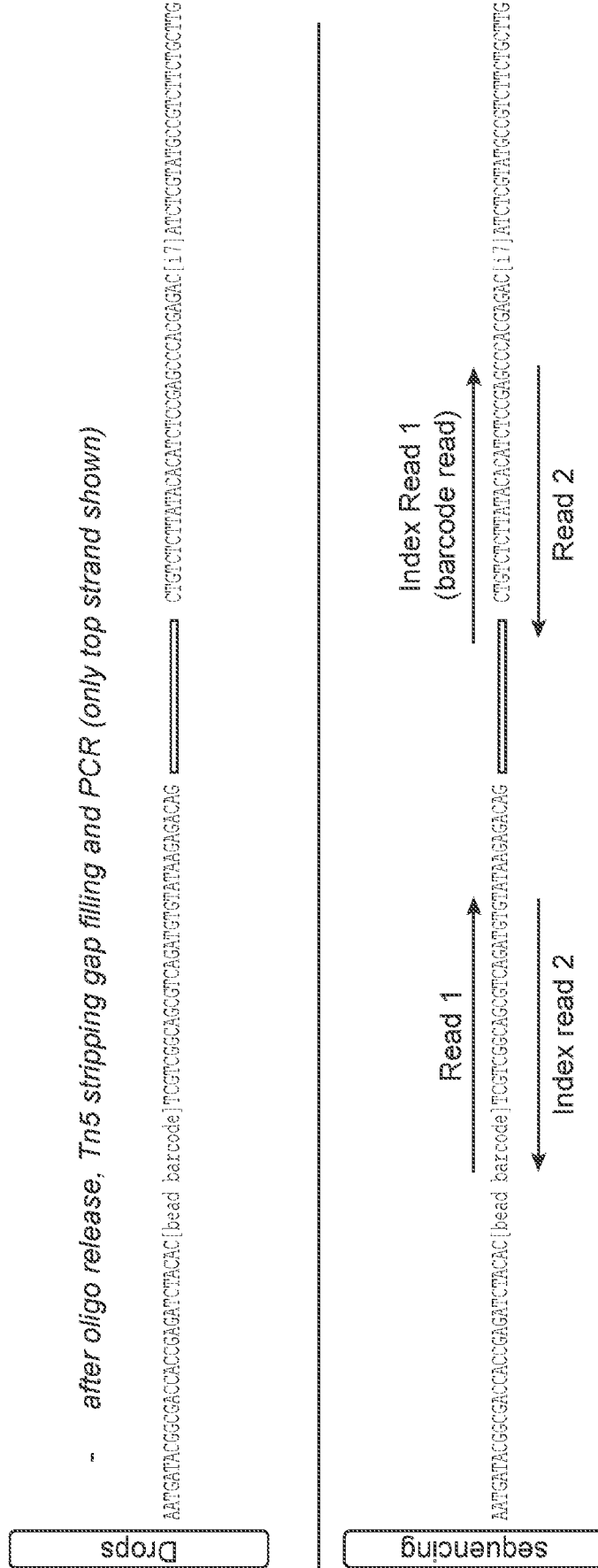
**Homoadapter Tn5 biochemistry, bead option 7 and sequencing**



**FIG. 26**

Homoadapter Tn5 biochemistry, bead option 7 and sequencing

- after oligo release, Tn5 stripping gap filling and PCR (only top strand shown)



NB: No dark cycles, since sequencing primer sits at end of mosaic end.

NBS: PCR suppression for this scheme not shown although applies

FIG. 27

**Homoadapter Tn5 biochemistry, bead option 8 and sequencing**

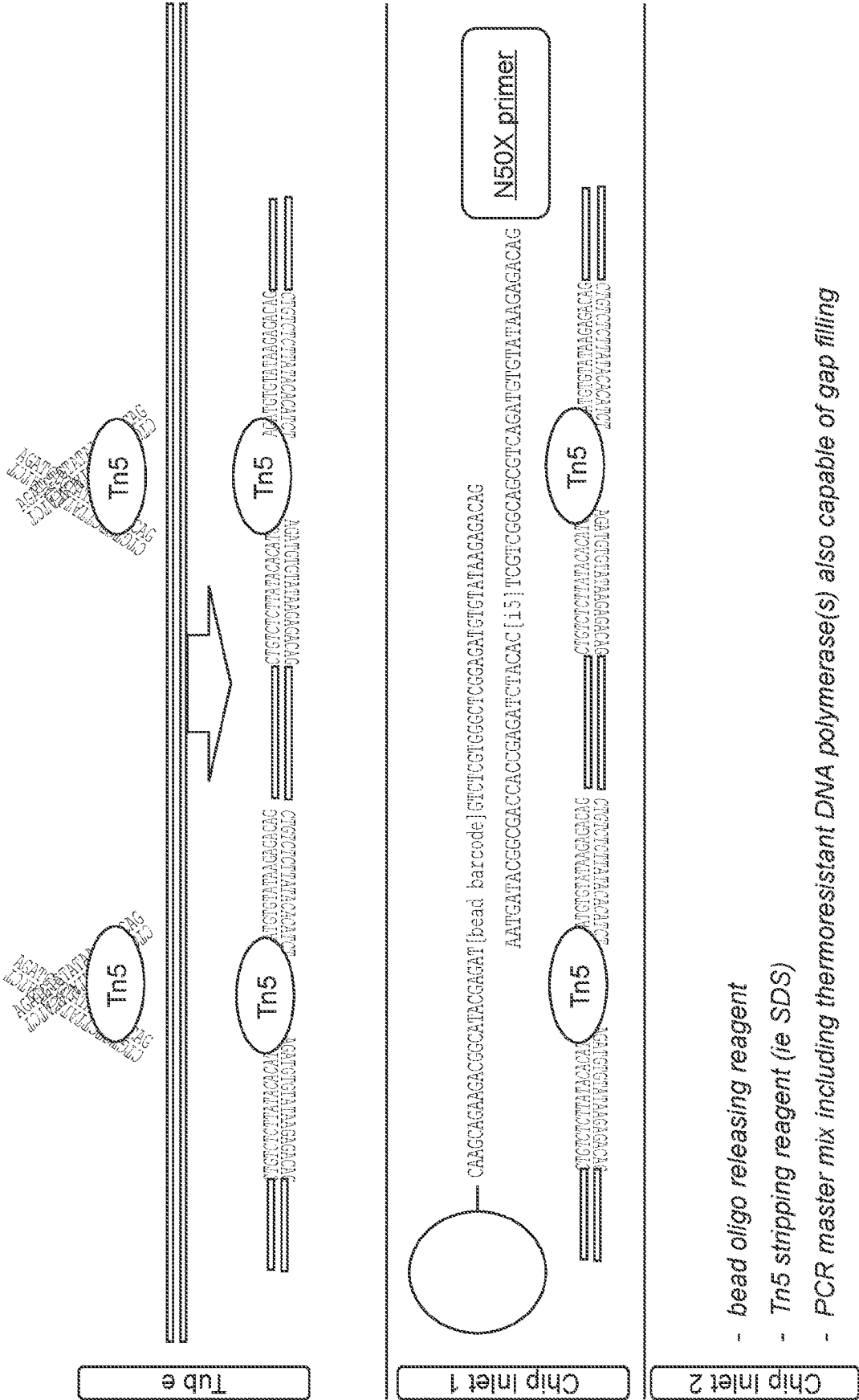
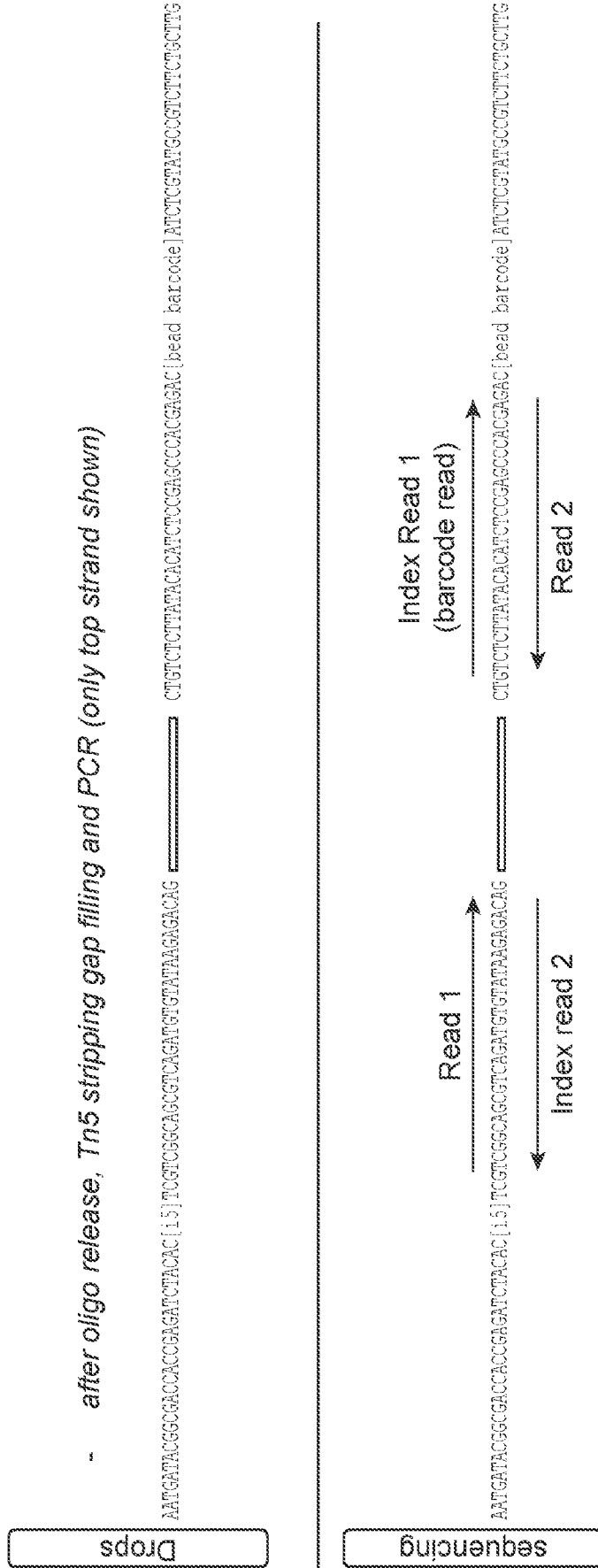


FIG. 28

**Homoadapter Tn5 biochemistry, bead option 8 and sequencing**

- after oligo release, Tn5 stripping gap filling and PCR (only top strand shown)



NB: No dark cycles, since sequencing primer sits at end of mosaic end.

NBS: PCR suppression for this scheme not shown although applies

FIG. 29

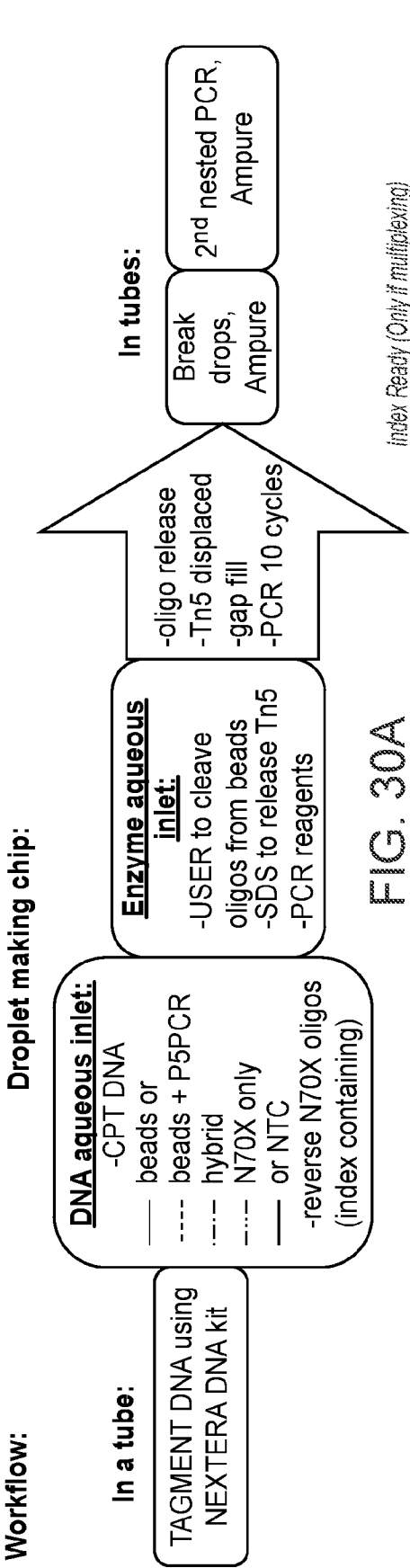


FIG. 30A

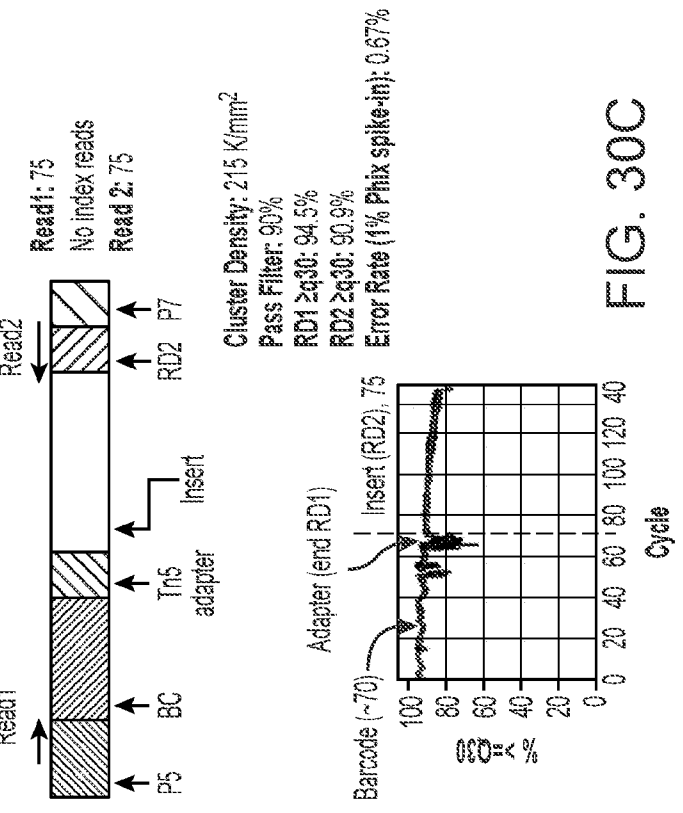


FIG. 30C

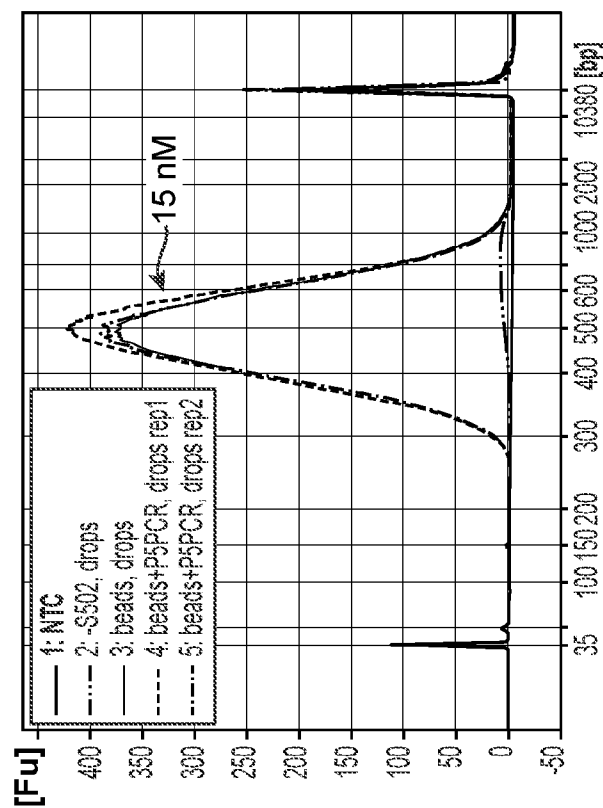


FIG. 30B

*CPT ddSEQ projected performance*

WGS	Beads in bulk (Zhang et al 2017 Nature Biotech)	10X Genomics (Zheng et al 2016 Nature Biotech)	CPTseq in droplets expected results
Fold coverage	19X	31-37X	>30X
% aligned	75%	95-97%	>80%
% duplication	21%	0.5-6%	<5%
Genome equivalents per drop	0.26%	0.11-0.44%	<0.4%
Barcode >10K reads	147K	111-150K	>100K
Average island length	58kb	38-75kb	>20kb
SNPs phased	98%	95-99%	>95%
N50 phase block	1.1Mb	0.89-2.8 Mb	>1Mb
Short switch SNP error	0.13%	0.01-0.93%	<1%
Long switch error	0.0085%	0.01-0.09%	<0.1%

FIG. 31

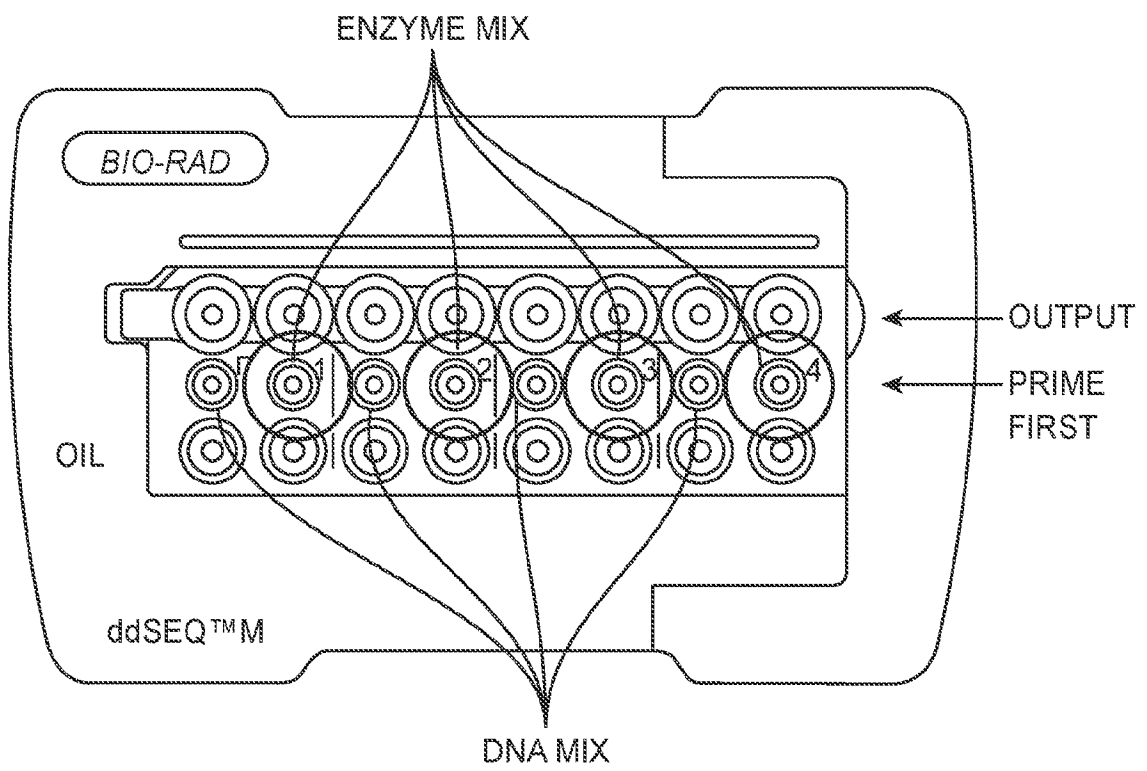


FIG. 32

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/067387

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 15/10; C12P 19/34; C40B 40/06; C40B 50/18; C40B 70/00 (2018.01)

CPC - C12N 15/10; C12N 15/1065; C12N 15/1093; C12P 19/34; C12Q 1/6869; C12Q 2525/191; C12Q 2535/122; C12Q 2563/179; C12Q 2565/514; C40B 40/06; C40B 50/18; C40B 70/00 (2018.02)

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)

See Search History document

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

USPC - 435/91.5; 506/2; 506/4; 506/16; 506/26; 536/23.1 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2016/061517 A2 (ILLUMINA CAMBRIDGE LIMITED et al) 21 April 2016 (21.04.2016) entire document	1-18, 21-37, 40-51, 54-56 ----- 19, 20, 38, 39, 52, 53
Y	US 2015/0045257 A1 (KAVANAGH et al) 12 February 2015 (12.02.2015) entire document	19, 20, 38, 39, 52, 53
A	US 2015/0363550 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 17 December 2015 (17.12.2015) entire document	1-56
A	WO 2012/106546 A2 (UNIVERSITY OF WASHINGTON THROUGH ITS CENTER FOR COMMERCIALIZATION et al) 09 August 2012 (09.08.2012) entire document	1-56
A	US 2010/0120098 A1 (GRUNENWALD et al) 13 May 2010 (13.05.2010) entire document	1-56

 Further documents are listed in the continuation of Box C.

 See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"I" 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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search

28 March 2018

Date of mailing of the international search report

26 APR 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
P.O. Box 1450, Alexandria, VA 22313-1450  
Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300  
PCT OSP: 571-272-7774



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/067387

**Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a.  forming part of the international application as filed:

in the form of an Annex C/ST.25 text file.

on paper or in the form of an image file.

b.  furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c.  furnished subsequent to the international filing date for the purposes of international search only:

in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2.  In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

ISA/225 mailed on 26 January 2018. No required late furnishing fee as set forth in 37 CFR 1.445(a)(5) was provided. Accordingly, ISA/US cannot consider the sequence listing submitted on 26 February 2018.