



- (51) International Patent Classification: *C12Q 1/6869* (2018.01)
- (21) International Application Number: PCT/US2024/024437
- (22) International Filing Date: 12 April 2024 (12.04.2024)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 63/459,214 13 April 2023 (13.04.2023) US
- (71) Applicant: **ESBIOLAB LLC** [US/US]; 8380 Miramar Mall Ste. 102, San Diego, California 92121 (US).
- (72) Inventors: **LI, Wei**; 8380 Miramar Mall Ste. 102, San Diego, California 92121 (US). **QIANG, Liangliang**; 8380 Miramar Mall Ste. 102, San Diego, California 92121 (US).
- (74) Agent: **YIN, Bu** et al.; Morrison & Foerster LLP, 12531 High Bluff Drive, Suite 100, San Diego, California 92130-2040 (US).
- (81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH,

(54) Title: METHODS AND COMPOSITIONS FOR NUCLEIC ACID SEQUENCING USING LABELED NUCLEOTIDES



FIG. 2

(57) Abstract: The present disclosure relates in some aspects to methods and compositions for analyzing nucleic acid sequences, such as for determining sequences of nucleic acid molecules in a clonal cluster, including DNA sequencing methods and nucleotide compositions. In some embodiments, the nucleic acid sequences are determined by contacting the nucleic acid molecules with nucleotides of a first base labeled with a first label detectable at a first wavelength, nucleotides of a second base labeled with a second label detectable at a second wavelength, and nucleotides of a third base comprising: (i) a nucleotide labeled with a third label detectable at the first wavelength, and (ii) a nucleotide labeled with a fourth label detectable at the second wavelength, where the third label and the fourth label can be comprises in the same nucleotide molecule or separate nucleotide molecules of the third base. In some embodiments, three images are acquired for each clonal cluster (e.g., generated using bridge amplification or rolling circle amplification) at a) the



TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,  
ZA, ZM, ZW.

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

**Declarations under Rule 4.17:**

— *of inventorship (Rule 4.17(iv))*

**Published:**

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

---

first wavelength alone, b) the second wavelength alone, and c) both the first wavelength and the second wavelength. In some embodiments, signal codewords corresponding to the signals or absence thereof in a) through c) is generated to identify the bases bound or incorporated during sequencing.

## **METHODS AND COMPOSITIONS FOR NUCLEIC ACID SEQUENCING USING LABELED NUCLEOTIDES**

### **CROSS-REFERENCE TO RELATED APPLICATION**

**[0001]** This application claims priority to U.S. Provisional Application No. 63/459,214, filed on April 13, 2023, entitled “METHODS AND COMPOSITIONS FOR NUCLEIC ACID SEQUENCING USING LABELED NUCLEOTIDES,” which is herein incorporated by reference in its entirety for all purposes.

### **FIELD**

**[0002]** The present disclosure relates in some aspects to methods and compositions for analyzing nucleic acid sequences, such as for determining sequences of nucleic acid molecules in a clonal cluster, including DNA sequencing methods and nucleotide compositions.

### **BACKGROUND**

**[0003]** To detect the nucleotides which extended onto single strand DNA with primers, fluorescent labels can be used to present different signals at different wavelengths. To minimize the crosstalk and computation power to decode the four different bases, fewer than four different fluorescence emission channels are preferred, and improved methods for nucleic acid sequencing are needed. Provided herein are methods, compositions, and kits that meet such and other needs.

### **BRIEF SUMMARY**

**[0004]** In some embodiments, provided herein are methods and systems to code the four nucleotides (A, T/U, G, and C) with two excitation wavelengths and less than four emission bandwidths. In some embodiments, provided herein are methods to code the four bases in DNA sequencing with different detectable labels and two illumination light sources. Also provided herein are optimized label configurations, where three or two fluorescence images can be used to determine the base extended at individual clusters with improve accuracy and simplified optical design.

**[0005]** In some embodiments, provided herein is a method for determining a sequence of a polynucleotide template, comprising: a) contacting the polynucleotide template with a first pool of nucleotides comprising: i) a nucleotide of a first base labeled with a first label configured to be detected at a first wavelength, ii) a nucleotide of a second base which is different from the first base and which nucleotide is labeled with a second label configured to

be detected at a second wavelength which is different from the first wavelength, and iii) one or more nucleotides of a third base which is different from the first and second bases, wherein the one or more nucleotides comprise a nucleotide labeled with a third label configured to be detected at the first wavelength and a nucleotide labeled with a fourth label configured to be detected at the second wavelength. In some embodiments, the method comprises b) allowing binding and optional incorporation of a nucleotide of the first pool templated on the polynucleotide template, wherein the bound and optionally incorporated nucleotide is complementary to a nucleotide residue at a first nucleotide position in the polynucleotide template. In any of the preceding embodiments, the method can comprise c) imaging the polynucleotide template to detect a signal (or record an absence of the signal) associated with the bound and optionally incorporated nucleotide of the first pool. In any of the preceding embodiments, the method can comprise: i) using a first power to image the polynucleotide template at the first wavelength and not at the second wavelength to acquire a first image, wherein a signal associated with the first label and/or a signal associated with the third label is detected or an absence of the signal is recorded, ii) using a second power to image the polynucleotide template at the second wavelength and not at the first wavelength to acquire a second image, wherein a signal associated with the second label and/or a signal associated with the fourth label is detected or an absence of the signal is recorded, and iii) using a third power to image the polynucleotide template at the first and second wavelengths to acquire a third image, wherein the third power is less than the first power and/or the second power, wherein a signal associated with the first label and/or a signal associated with the third label is detected at a signal intensity lower than that detected in c)i) or an absence of the signal is recorded, and wherein a signal associated with the second label and/or a signal associated with the fourth label is detected at a signal intensity lower than that detected in c)ii) or an absence of the signal is recorded. In any of the preceding embodiments, the method can comprise generating a signal codeword comprising signal codes each corresponding to the intensity of the signal detected in c)i) through c)iii) (or recorded absence of the signal), wherein the signal codeword corresponds to the identity of the base in the bound and optionally incorporated nucleotide. In any of the preceding embodiments, the method can comprise identifying the nucleotide residue at the first nucleotide position in the polynucleotide template.

**[0006]** In any of the preceding embodiments, the first label and the third label can be different labels or the same label. In any of the preceding embodiments, the second label and the fourth label can be different labels or the same label. In any of the preceding

embodiments, the one or more nucleotides of the third base can comprise a nucleotide labeled with the third label and the fourth label. In any of the preceding embodiments, the one or more nucleotides of the third base can comprise a first nucleotide species labeled with the third label and a second nucleotide species labeled with the fourth label. In any of the preceding embodiments, the third label and the fourth label can be different labels or the same label. In any of the preceding embodiments, the first and second wavelengths can correspond to different channels of a fluorescence microscope.

**[0007]** In any of the preceding embodiments, in c)i), the signal associated with the first label or the signal associated with the third label can be detected at a first signal intensity. In any of the preceding embodiments, in c)ii), the signal associated with the second label or the signal associated with the fourth label can be detected at a second signal intensity. In any of the preceding embodiments, the first and second signal intensities can be different by no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, or no more than 1%. In any of the preceding embodiments, the first and second signal intensities can be the same. In any of the preceding embodiments, in c)iii), the signal associated with the first label or the signal associated with the third label can be detected at a signal intensity lower than the first signal intensity. In any of the preceding embodiments, in c)iii), the signal associated with the second label or the signal associated with the fourth label can be detected at a signal intensity lower than the second signal intensity.

**[0008]** In any of the preceding embodiments, the signal associated with the first label or the signal associated with the third label can be detected at a signal intensity between about 25% and about 75% of the first signal intensity. In any of the preceding embodiments, the signal associated with the second label or the signal associated with the fourth label can be detected at a signal intensity between about 25% and about 75% of the second signal intensity. In any of the preceding embodiments, the signal associated with the first label or the signal associated with the third label can be detected at a signal intensity about 50% of the first signal intensity. In any of the preceding embodiments, the signal associated with the second label or the signal associated with the fourth label can be detected at a signal intensity about 50% of the second signal intensity.

**[0009]** In any of the preceding embodiments, the first pool of nucleotides can comprise a nucleotide of a fourth base which is different from the first, second, and third bases. In any of the preceding embodiments, the nucleotide of the fourth base can be configured to be nondetectable at the first wavelength and the second wavelength. In any of the preceding embodiments, the nucleotide of the fourth base can be free of a detectable label. In any of the

preceding embodiments, the nucleotide of the fourth base can be configured to be detected at a third wavelength different from the first and second wavelengths.

**[0010]** In any of the preceding embodiments, the polynucleotide template can be one of a plurality of polynucleotide templates having different template sequences. In any of the preceding embodiments, the sequences of the plurality of polynucleotide templates are determined. In any of the preceding embodiments, the polynucleotide template can be immobilized on a substrate. In any of the preceding embodiments, the polynucleotide template can be in one or more polynucleotides immobilized on the substrate. In any of the preceding embodiments, the polynucleotide template can be in a rolling circle amplification product immobilized on the substrate, wherein the rolling circle amplification product comprises multiple copies of the template sequence in the polynucleotide template. In any of the preceding embodiments, the polynucleotide template can be in a cluster of multiple polynucleotides immobilized on the substrate, wherein each polynucleotide in the cluster comprises a copy of the template sequence in the polynucleotide template.

**[0011]** In any of the preceding embodiments, in c)i), the plurality of polynucleotide templates can be imaged, and the first image can comprise the signal associated with the first label at the location of a first polynucleotide template, a recorded absence of a signal at the location of a second polynucleotide template, and the signal associated with the third label at the location of a third polynucleotide template. In any of the preceding embodiments, in c)ii), the plurality of polynucleotide templates can be imaged and the second image can comprise a recorded absence of a signal at the location of the first polynucleotide template, the signal associated with the second label at the location of the second polynucleotide template, and the signal associated with the fourth label at the location of a third polynucleotide template. In any of the preceding embodiments, in c)iii), the plurality of polynucleotide templates can be imaged and the third image can comprise: the signal associated with the first label at the location of the first polynucleotide template; the signal associated with the second label at the location of the second polynucleotide template, and the signal associated with the third label and the signal associated with the fourth label at the location of a third polynucleotide template.

**[0012]** In any of the preceding embodiments, in c)i) and c)ii), the first and second images can be acquired using the same power or powers that differ by no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, or no more than 1%, and the c)iii), the third image can be acquired using a power that is between about 25% and about 75% (optionally 50%) of the power(s) used in c)i) and/or c)ii).

**[0013]** In some embodiments, provided herein is a method for determining sequences of a plurality of polynucleotide templates having different template sequences, comprising: a) contacting a substrate having clusters of polynucleotides immobilized thereon with a first pool of nucleotides and with oligonucleotide primers, in any order, whereby each cluster comprises: i) multiple copies of one of the different template sequences, and ii) an oligonucleotide primer of the oligonucleotide primers annealed to a primer binding sequence for extension of the oligonucleotide primer templated on a copy of the template sequence, wherein the first pool of nucleotides comprises: i) nucleotides of a first base which are labeled with a label configured to be detected at a first wavelength, ii) nucleotides of a second base which are labeled with a label configured to be detected at a second wavelength which is different from the first wavelength, and iii) nucleotides of a third base comprising nucleotides labeled with a label configured to be detected at the first wavelength and nucleotides labeled with a label configured to be detected at the second wavelength.

**[0014]** In some embodiments, the method further comprises: b) allowing binding and optional incorporation of the nucleotides of the first pool templated on the multiple copies of the template sequence at each cluster, wherein the bound and optionally incorporated nucleotides are complementary to nucleotides at a first nucleotide position in the template sequences.

**[0015]** In any of one or more of the preceding embodiments, the method can further comprise: c) imaging the substrate to detect signals or absence thereof at the clusters, wherein the signals are associated with the bound and optionally incorporated nucleotides of the first pool, and wherein the imaging comprises: i) imaging the substrate at the first wavelength and not at the second wavelength to acquire a first image, ii) imaging the substrate at the second wavelength and not at the first wavelength to acquire a second image, and iii) imaging the substrate at the first and second wavelengths simultaneously to acquire a third image, wherein in the third image, the signal intensity associated with the third base is no less than the signal intensity associated with the first base or the signal intensity associated with the second base.

**[0016]** In any of one or more of the preceding embodiments, the method can further comprise generating, for each cluster, a signal codeword comprising signal codes corresponding to the signals or absence thereof detected in c)i) through c)iii), wherein different signal codewords correspond to different bases, thereby determining the bases at the first nucleotide position in the plurality of polynucleotide templates.

**[0017]** In any of one or more of the preceding embodiments, the first pool of nucleotides can comprise nucleotides of a fourth base which are unlabeled and/or are undetected in the

imaging. Alternatively, in any of one or more of the preceding embodiments, the first pool of nucleotides may not comprise nucleotides of a fourth base.

**[0018]** In any of one or more of the preceding embodiments, the nucleotides of the third base can comprise a nucleotide labeled both with the label configured to be detected at the first wavelength and with the label configured to be detected at the second wavelength.

**[0019]** In any of one or more of the preceding embodiments, the nucleotides of the third base can comprise a first nucleotide labeled with only the label configured to be detected at the first wavelength and a second nucleotide labeled with only the label configured to be detected at the second wavelength.

**[0020]** In any of one or more of the preceding embodiments, in each of the nucleotides of the third base, the ratio of the labels configured to be detected at the first or second wavelength and the nucleotide can be independently 1:1, 2:1, 3:1, 4:1, 5:1, or greater.

**[0021]** In any of one or more of the preceding embodiments, the labels can be conjugated to the nucleotide via a binding pair, optionally wherein the binding pair comprises biotin, DNP, DIG, or desthiobiotin, and a corresponding binding partner thereof.

**[0022]** In any of one or more of the preceding embodiments, in c)i), the substrate can be illuminated at the first wavelength at a first full power to acquire the first image. In any of one or more of the preceding embodiments, in c)ii), the substrate can be illuminated at the second wavelength at a second full power to acquire the second image. In any of one or more of the preceding embodiments, in c)iii), the substrate can be illuminated at the first wavelength at less than the first full power, and illuminated at the second wavelength at less than the second full power, thereby acquiring the third image. In any of one or more of the preceding embodiments, in c)iii), the substrate can be illuminated at the first wavelength at about 50% of the first full power, and at the second wavelength at about 50% of the second full power to acquire the third image. In any of one or more of the preceding embodiments, the first and second full powers can be of the same constant current in LED illumination. In any of one or more of the preceding embodiments, the first and second full powers can be of different constant currents in LED illumination. In any of one or more of the preceding embodiments, in the third image, the signal intensity associated with the third base can be about twice the signal intensity associated with the first base or the signal intensity associated with the second base.

**[0023]** In any of one or more of the preceding embodiments, the signal codeword can comprise signal codes corresponding to signal intensities associated with the nucleotides of the first base, the nucleotides of the second base, or the nucleotides of the third base.



Alternatively, in any of one or more of the preceding embodiments, the signal codeword may not comprise signal codes corresponding to signal intensities associated with the nucleotides of the first base, the nucleotides of the second base, or the nucleotides of the third base.

**[0024]** In any of one or more of the preceding embodiments, the method can comprise: d) contacting the substrate with a second pool of nucleotides comprising unlabeled nucleotides of the first base, unlabeled nucleotides of the second base, unlabeled nucleotides of the third base, and/or unlabeled nucleotides of a fourth base. In any of one or more of the preceding embodiments, the method can comprise: e) allowing binding and optional incorporation of the nucleotides of the second pool templated on the multiple copies of the template sequence at each cluster, wherein the bound and optionally incorporated nucleotides are complementary to nucleotides at the first nucleotide position in the template sequences.

**[0025]** In any of one or more of the preceding embodiments, the method can comprise: d) contacting the substrate with a second pool of nucleotides comprising labeled nucleotides of a fourth base. In any of one or more of the preceding embodiments, the method can comprise: e) allowing binding and optional incorporation of the nucleotides of the second pool templated on the multiple copies of the template sequence at each cluster, wherein the bound and optionally incorporated nucleotides are complementary to nucleotides at the first nucleotide position in the template sequences. In any of one or more of the preceding embodiments, the method can comprise: f) imaging the substrate to detect signals or absence thereof at the clusters, wherein the signals are associated with the bound and optionally incorporated nucleotides of the second pool.

**[0026]** In any of one or more of the preceding embodiments, the clusters can be disposed at spatially discrete sites on the substrate. In any of one or more of the preceding embodiments, the average distance between adjacent clusters on the substrate can be between about 0.3  $\mu\text{m}$  and about 10  $\mu\text{m}$ . In any of one or more of the preceding embodiments, the density of clusters on the substrate can be between about 100  $\text{k}/\text{mm}^2$  and about 5000  $\text{k}/\text{mm}^2$ .

**[0027]** In any of one or more of the preceding embodiments, signals from any two adjacent clusters on the substrate can be optically resolvable. In any of one or more of the preceding embodiments, the clusters can comprise a random array of clusters on the substrate. In any of one or more of the preceding embodiments, the clusters can comprise an ordered array of clusters on the substrate.

**[0028]** In any of one or more of the preceding embodiments, any one or more of the clusters can be formed via bridge amplification. In any of one or more of the preceding embodiments, any one or more of the clusters can each comprise multiple molecules each

comprising: i) one or more adapter sequences and/or one or more primer binding sequences and ii) the same template sequence or complement thereof. In any of one or more of the preceding embodiments, any one or more of the clusters can be formed via rolling circle amplification (RCA). In any of one or more of the preceding embodiments, any one or more of the clusters can each comprise one or more RCA products (RCPs) each comprising: i) one or more adapter sequences and/or one or more primer binding sequences and ii) the same template sequence.

**[0029]** In any of one or more of the preceding embodiments, any one or more of the clusters can each comprise: i) one or more polynucleotides that are 5' immobilized on the substrate and 3' blocked; ii) one or more polynucleotides that are 3' immobilized on the substrate; and/or iii) one or more nucleic acid concatemers immobilized on the substrate.

**[0030]** In any of one or more of the preceding embodiments, the first pool of nucleotides can comprise nucleotides of any three of A, T/U, C, and G. In any of one or more of the preceding embodiments, the first pool of nucleotides can comprise nucleotides of A, T/U, and C, but no nucleotides of G.

**[0031]** In any of one or more of the preceding embodiments, the first and/or second pool of nucleotides can comprise one or more dNTP monomers and/or one or more dNTP multimers each comprising multiple dNTP molecules conjugated to a scaffold. In some embodiments, the scaffold is conjugated to one or more detectable labels such as fluorescent moieties. In some embodiments, the scaffold comprises a streptavidin or a variant or mutein thereof. In some embodiments, the scaffold comprises an oligomeric streptavidin or a variant or mutein thereof. In any of one or more of the preceding embodiments, the first and/or second pool of nucleotides can each comprise a reversible terminator. In any of one or more of the preceding embodiments, the first and/or second pool of nucleotides can each comprise a 3'-*O*-blocked reversible terminator or a 3'-unblocked reversible terminator. In any of one or more of the preceding embodiments, the method can comprise allowing nucleotide incorporation after each imaging of the substrate. In any of one or more of the preceding embodiments, the method can comprise removing unbound or unincorporated nucleotides from the substrate prior to imaging the substrate.

**[0032]** In any of one or more of the preceding embodiments, the method can comprise determining the bases at a second nucleotide position in the plurality of polynucleotide templates, wherein the second nucleotide position is 5' to the first nucleotide position in the plurality of polynucleotide templates.

**[0033]** Also provided herein in some embodiments include a kit, comprising: i) a first plurality of nucleotide molecules of a first base, comprising nucleotides labeled or configured to be labeled with a first label detectable at a first wavelength, ii) a second plurality of nucleotide molecules of a second base which is different from the first base, comprising nucleotides labeled or configured to be labeled with a second label detectable at a second wavelength which is different from the first wavelength, and iii) a third plurality of nucleotide molecules of a third base which is different from the first and second bases, comprising a nucleotide labeled or configured to be labeled with a third label detectable at the first wavelength and a nucleotide labeled or configured to be labeled with a fourth label detectable at the second wavelength.

**[0034]** In any of the preceding embodiments, the kit can comprise instructions for using the first plurality, the second plurality, and/or the third plurality of nucleotide molecules according to a method in the present disclosure.

**[0035]** In any of the preceding embodiments, the first plurality, the second plurality, and/or the third plurality can each independently comprise between about 5% and about 50% nucleotide molecules that are not detectably labeled. In any of the preceding embodiments, the first plurality, the second plurality, and/or the third plurality can each independently comprise about 10% nucleotide molecules that are not detectably labeled.

**[0036]** In any of the preceding embodiments, the kit can comprise a fourth plurality of nucleotide molecules of a fourth base which is different from the first, second, and third bases, optionally wherein the fourth plurality of nucleotide molecules are not detectably labeled.

**[0037]** In any of the preceding embodiments, the first, second, third, and/or fourth plurality can each independently comprise one or more reversibly terminated nucleotide molecules, and/or one or more complexes comprising nucleotide molecules conjugated to a scaffold such that the nucleotide molecules are not incorporable.

### **INCORPORATION BY REFERENCE**

**[0038]** All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference in their entirety to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference in its entirety. In the event of a conflict between a term herein and a term in an incorporated reference, the term herein controls.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0039] Various aspects of the disclosed methods, devices, and systems are set forth with particularity in the appended claims. A better understanding of the features and advantages of the disclosed methods, devices, and systems will be obtained by reference to the following detailed description of illustrative embodiments and the accompanying drawings, of which:

[0040] **FIG. 1** provides an exemplary flowchart for nucleic acid sequencing, according to some embodiments described herein.

[0041] **FIG. 2** provides an exemplary process for nucleic acid sequencing. In this example, sequencing-by-synthesis is performed with labeled nucleotides for three out of the four bases. A) shows a representative status of DNA strands with primer during sequencing with ATGC as next base; B) shows the extended base after contact with labeled dATP, labeled dCTP, and labeled dTTP, where G is omitted since dGTP is not added; C) shows an image truth table: Image 1 is green illumination at a constant current drive (e.g., 12 Amp), Image 2 is red illumination at the same current (e.g., 12 Amp), Image 3 is green and red illumination with a total current that is same as previous (e.g. 12 Amp). D) shows the 3' block can be cleaved after further contact with dye-labeled dGTP and/or unlabeled dGTP. All of the nucleotides in this illustration bear 3' blocker.

[0042] **FIG. 3** provides an exemplary ideal plot of intensities of DNA clusters in Image 1 and Image 2. A is red only, T is green only, and C is dual color with more efficient fluorescent dyes which intensities are equal to red only dye or green only dye at half or less surface dye concentration.

[0043] **FIG. 4** provides an exemplary plot of signal intensities of DNA clusters in Image 1 and Image 2 in reality. A is red only, T is green only, and C is dual color with which intensity at red or green channel is half of the red only dye or the green only dye, respectively.

[0044] **FIG. 5** provides an exemplary plot of signal intensities of DNA clusters in Image 1, Image 2 and Image 3. A is red only, T is green only, and C is dual color with which intensity at red or green channel is equal to the red only dye or the green only dye, respectively. The distance between bright clusters and dark clusters is improved due to the use of Image 3.

[0045] **FIG. 6** provides an exemplary plot of signal intensities of DNA clusters in Image 1, Image 2 and Image 3. A is red only, T is green only, and C is dual color with which intensity at red or green channel is half of the red only dye or the green only dye, respectively. The distance between bright clusters and dark clusters is improved due to the use of Image 3.

[0046] FIG. 7 provides an exemplary scheme showing streptavidin with dual dye labels which can be used to stain a biotin or desthiobiotin tagged nucleotide during sequencing to improve the brightness of dual coded nucleotide such as C in FIGS. 3-6.

[0047] FIG. 8 provides an exemplary scheme showing a streptavidin mixture with different dye labels which can be used to stain biotin or desthiobiotin tagged nucleotide during sequencing to improve the brightness of dual coded nucleotides such as C in FIGS. 3-6.

## DETAILED DESCRIPTION

### I. Definitions

[0048] Specific terminology is used throughout this disclosure to explain various aspects of the apparatus, systems, methods, and compositions that are described.

[0049] Having described some illustrative embodiments of the present disclosure, it should be apparent to those skilled in the art that the foregoing is merely illustrative and not limiting, having been presented by way of example only. Numerous modifications and other illustrative embodiments are within the scope of one of ordinary skill in the art and are contemplated as falling within the scope of the present disclosure. In particular, although many of the examples presented herein involve specific combinations of method acts or system elements, it should be understood that those acts and those elements may be combined in other ways to accomplish the same objectives.

[0050] As used herein, the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise. For example, “a” or “an” means “at least one” or “one or more.”

[0051] The term “about” as used herein refers to the usual error range for the respective value readily known to the skilled person in this technical field. Reference to “about” a value or parameter herein includes (and describes) embodiments that are directed to that value or parameter per se.

[0052] Throughout this disclosure, various aspects of the claimed subject matter are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the claimed subject matter. Accordingly, the description of a range should be considered to have specifically disclosed all the possible sub-ranges as well as individual numerical values within that range. For example, where a range of values is provided, it is understood that each intervening value, between the upper and lower limit of that range and

any other stated or intervening value in that stated range is encompassed within the claimed subject matter. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the claimed subject matter, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the claimed subject matter. This applies regardless of the breadth of the range.

**[0053]** Use of ordinal terms such as “first”, “second”, “third”, etc., in the claims to modify a claim element does not by itself connote any priority, precedence, or order of one claim element over another or the temporal order in which acts of a method are performed, but are used merely as labels to distinguish one claim element having a certain name from another element having a same name (but for use of the ordinal term) to distinguish the claim elements. Similarly, use of a), b), etc., or i), ii), etc. does not by itself connote any priority, precedence, or order of steps in the claims. Similarly, the use of these terms in the specification does not by itself connote any required priority, precedence, or order.

**[0054]** The terms “nucleic acid” and “nucleotide” are intended to be consistent with their use in the art and to include naturally-occurring species or functional analogs thereof. Particularly useful functional analogs of nucleic acids are capable of hybridizing to a nucleic acid in a sequence-specific fashion (e.g., capable of hybridizing to two nucleic acids such that ligation can occur between the two hybridized nucleic acids) or are capable of being used as a template for replication of a particular nucleotide sequence. Naturally-occurring nucleic acids generally have a backbone containing phosphodiester bonds. An analog structure can have an alternate backbone linkage including any of a variety of those known in the art. Naturally-occurring nucleic acids generally have a deoxyribose sugar (e.g., found in deoxyribonucleic acid (DNA)) or a ribose sugar (e.g. found in ribonucleic acid (RNA)).

**[0055]** A nucleic acid can contain nucleotides having any of a variety of analogs of these sugar moieties that are known in the art. A nucleic acid can include native or non-native nucleotides. In this regard, a native deoxyribonucleic acid can have one or more bases selected from the group consisting of adenine (A), thymine (T), cytosine (C), or guanine (G), and a ribonucleic acid can have one or more bases selected from the group consisting of uracil (U), adenine (A), cytosine (C), or guanine (G). Useful non-native bases that can be included in a nucleic acid or nucleotide are known in the art.

**[0056]** A “probe” or a “target,” when used in reference to a nucleic acid or sequence of a nucleic acids, is intended as a semantic identifier for the nucleic acid or sequence in the

context of a method or composition, and does not limit the structure or function of the nucleic acid or sequence beyond what is expressly indicated.

**[0057]** The terms “oligonucleotide” and “polynucleotide” are used interchangeably to refer to a single-stranded multimer of nucleotides from about 2 to about 500 nucleotides in length. Oligonucleotides can be synthetic, made enzymatically (e.g., via polymerization), or using a “split-pool” method. Oligonucleotides can include ribonucleotide monomers (e.g., can be oligoribonucleotides) and/or deoxyribonucleotide monomers (e.g., oligodeoxyribonucleotides). In some examples, oligonucleotides can include a combination of both deoxyribonucleotide monomers and ribonucleotide monomers in the oligonucleotide (e.g., random or ordered combination of deoxyribonucleotide monomers and ribonucleotide monomers). An oligonucleotide can be 4 to 10, 10 to 20, 21 to 30, 31 to 40, 41 to 50, 51 to 60, 61 to 70, 71 to 80, 80 to 100, 100 to 150, 150 to 200, 200 to 250, 250 to 300, 300 to 350, 350 to 400, or 400 to 500 nucleotides in length, for example. Oligonucleotides can include one or more functional moieties that are attached (e.g., covalently or non-covalently) to the multimer structure. For example, an oligonucleotide can include one or more detectable labels (e.g., a radioisotope or fluorophore).

**[0058]** The terms “detectable label,” “optical label,” and “label” are used interchangeably herein to refer to a directly or indirectly detectable moiety that is coupled to or may be coupled to another moiety, for example, a nucleotide or nucleotide analog. The detectable label can be directly detectable by itself (e.g., radioisotope labels or fluorescent labels) or, in the case of an enzymatic label, can be indirectly detectable, e.g., by catalyzing chemical alterations of a substrate compound or composition, which substrate compound or composition is directly detectable. The label can emit a signal or alter a signal delivered to the label so that the presence or absence of the label can be detected. In some cases, coupling may be via a linker, which may be cleavable, such as photo-cleavable (e.g., cleavable under ultra-violet light), chemically-cleavable (e.g., via a reducing agent, such as dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP)) or enzymatically cleavable (e.g., via an esterase, lipase, peptidase, or protease).

**[0059]** In some embodiments, a detectable label is or includes a fluorophore. Exemplary fluorophores include, but are not limited to, fluorescent nanocrystals; quantum dots; d-Rhodamine acceptor dyes including dichloro[R110], dichloro[R6G], dichloro[TAMRA], dichloro[ROX] or the like; fluorescein donor dye including fluorescein, 6-FAM, or the like; Cyanine dyes such as Cy3B; Alexa dyes, SETA dyes, Atto dyes such as atto 647N which forms a FRET pair with Cy3B and the like. Fluorophores include, but are not limited to,

MDCC (7-diethylamino-3-[[[(2-maleimidyl)ethyl]amino]carbonyl]coumarin), TET, HEX, Cy3, TMR, ROX, Texas Red, Cy5, LC red 705 and LC red 640.

**[0060]** In some embodiments, a detectable label is or includes a luminescent or chemiluminescent moiety. Common luminescent/chemiluminescent moieties include, but are not limited to, peroxidases such as horseradish peroxidase (HRP), soybean peroxidase (SP), alkaline phosphatase, and luciferase. These protein moieties can catalyze chemiluminescent reactions given the appropriate substrates (e.g., an oxidizing reagent plus a chemiluminescent compound). A number of compound families are known to provide chemiluminescence under a variety of conditions. Non-limiting examples of chemiluminescent compound families include 2,3-dihydro-1,4-phthalazinedione luminol, 5-amino-6,7,8-trimethoxy- and the dimethylamino[ca]benz analog. These compounds can luminesce in the presence of alkaline hydrogen peroxide or calcium hypochlorite and base. Other examples of chemiluminescent compound families include, e.g., 2,4,5-triphenylimidazoles, para-dimethylamino and -methoxy substituents, oxalates such as oxalyl active esters, p-nitrophenyl, N-alkyl acridinum esters, luciferins, lucigenins, or acridinium esters. In some embodiments, a detectable label is or includes a metal-based or mass-based label.

**[0061]** The terms “hybridizing,” “hybridize,” “annealing,” and “anneal” are used interchangeably in this disclosure, and refer to the pairing of substantially complementary or complementary nucleic acid sequences within two different molecules. Pairing can be achieved by any process in which a nucleic acid sequence joins with a substantially or fully complementary sequence through base pairing to form a hybridization complex. For purposes of hybridization, two nucleic acid sequences are “substantially complementary” if at least 60% (e.g., at least 70%, at least 80%, or at least 90%) of their individual bases are complementary to one another.

**[0062]** A “primer” is a single-stranded nucleic acid sequence having a 3' end that can be used as a substrate for a nucleic acid polymerase in a nucleic acid extension reaction. RNA primers are formed of RNA nucleotides, and are used in RNA synthesis, while DNA primers are formed of DNA nucleotides and used in DNA synthesis. Primers can also include both RNA nucleotides and DNA nucleotides (e.g., in a random or designed pattern). Primers can also include other natural or synthetic nucleotides described herein that can have additional functionality. In some examples, DNA primers can be used to prime RNA synthesis and vice versa (e.g., RNA primers can be used to prime DNA synthesis). Primers can vary in length. For example, primers can be about 6 bases to about 120 bases. For example, primers can



include up to about 25 bases. A primer, may in some cases, refer to a primer binding sequence.

**[0063]** A “nucleic acid extension” generally involves incorporation of one or more nucleic acids (e.g., A, G, C, T, U, nucleotide analogs, or derivatives thereof) into a molecule (such as, but not limited to, a nucleic acid sequence) in a template-dependent manner, such that consecutive nucleic acids are incorporated by an enzyme (such as a polymerase or reverse transcriptase), thereby generating a newly synthesized nucleic acid molecule. Enzymatic extension can be performed by an enzyme including, but not limited to, a polymerase and/or a reverse transcriptase. For example, a primer that hybridizes to a complementary nucleic acid sequence can be used to synthesize a new nucleic acid molecule by using the complementary nucleic acid sequence as a template for nucleic acid synthesis. Similarly, a 3' polyadenylated tail of an mRNA transcript that hybridizes to a poly (dT) sequence can be used as a template for single-strand synthesis of a corresponding cDNA molecule. Furthermore, a poly (dT) sequence may be used as a sequencing primer for sequencing RNA molecules comprising poly(A) tails.

**[0064]** A “non-terminating nucleotide” or “incorporating nucleotide” can include a nucleic acid moiety that can be attached to a 3' end of a polynucleotide using a polymerase or transcriptase, and that can have another non-terminating nucleic acid attached to it using a polymerase or transcriptase without the need to remove a protecting group or reversible terminator from the nucleotide. Naturally occurring nucleic acids are a type of non-terminating nucleic acid. Non-terminating nucleic acids may be labeled or unlabeled.

**[0065]** A “PCR amplification” refers to the use of a polymerase chain reaction (PCR) to generate copies of genetic material, including DNA and RNA sequences. Suitable reagents and conditions for implementing PCR are described, for example, in U.S. Patent Nos. 4,683,202, 4,683,195, 4,800,159, 4,965,188, and 5,512,462, the entire contents of each of which are incorporated herein by reference. In a typical PCR amplification, the reaction mixture includes the genetic material to be amplified, an enzyme, one or more primers that are employed in a primer extension reaction, and reagents for the reaction. The oligonucleotide primers are of sufficient length to provide for hybridization to complementary genetic material under annealing conditions. The length of the primers generally depends on the length of the amplification domains, but will typically be at least 4 bases, at least 5 bases, at least 6 bases, at least 8 bases, at least 9 bases, at least 10 base pairs (bp), at least 11 bp, at least 12 bp, at least 13 bp, at least 14 bp, at least 15 bp, at least 16 bp, at least 17 bp, at least 18 bp, at least 19 bp, at least 20 bp, at least 25 bp, at least 30 bp, at

least 35 bp, and can be as long as 40 bp or longer, where the length of the primers will generally range from 18 to 50 bp. The genetic material can be contacted with a single primer or a set of two primers (forward and reverse primers), depending upon whether primer extension, linear or exponential amplification of the genetic material is desired.

**[0066]** In some embodiments, the PCR amplification process uses a DNA polymerase enzyme. The DNA polymerase activity can be provided by one or more distinct DNA polymerase enzymes. In some embodiments, the DNA polymerase enzyme is from a bacterium, e.g., the DNA polymerase enzyme is a bacterial DNA polymerase enzyme. For instance, the DNA polymerase can be from a bacterium of the genus *Escherichia*, *Bacillus*, *Thermophilus*, or *Pyrococcus*.

**[0067]** In some embodiments, PCR amplification can include reactions such as, but not limited to, a strand-displacement amplification reaction, a rolling circle amplification reaction, a ligase chain reaction, a transcription-mediated amplification reaction, an isothermal amplification reaction, and/or a loop-mediated amplification reaction.

**[0068]** In some embodiments, PCR amplification uses a single primer that is complementary to the 3' tag of target DNA fragments. In some embodiments, PCR amplification uses a first and a second primer, where at least a 3' end portion of the first primer is complementary to at least a portion of the 3' tag of the target nucleic acid fragments, and where at least a 3' end portion of the second primer exhibits the sequence of at least a portion of the 5' tag of the target nucleic acid fragments. In some embodiments, a 5' end portion of the first primer is non-complementary to the 3' tag of the target nucleic acid fragments, and a 5' end portion of the second primer does not exhibit the sequence of at least a portion of the 5' tag of the target nucleic acid fragments. In some embodiments, the first primer includes a first universal sequence and/or the second primer includes a second universal sequence.

**[0069]** The term "DNA polymerase" includes not only naturally-occurring enzymes but also all modified derivatives thereof, including also derivatives of naturally-occurring DNA polymerase enzymes. For instance, in some embodiments, the DNA polymerase can have been modified to remove 5'-3' exonuclease activity. Sequence-modified derivatives or mutants of DNA polymerase enzymes that can be used include, but are not limited to, mutants that retain at least some of the functional, e.g., DNA polymerase activity of the wild-type sequence. Mutations can affect the activity profile of the enzymes, e.g., enhance or reduce the rate of polymerization, under different reaction conditions, e.g., temperature, template concentration, primer concentration, etc. Mutations or sequence-modifications can also affect the exonuclease activity and/or thermostability of the enzyme.

**[0070]** Suitable examples of DNA polymerases that can be used include, but are not limited to: *E.coli* DNA polymerase I, Bsu DNA polymerase, Bst DNA polymerase, Taq DNA polymerase, VENT™ DNA polymerase, DEEPVENT™ DNA polymerase, LongAmp® Taq DNA polymerase, LongAmp® Hot Start Taq DNA polymerase, Crimson LongAmp® Taq DNA polymerase, Crimson Taq DNA polymerase, OneTaq® DNA polymerase, OneTaq® Quick-Load® DNA polymerase, Hemo KlenTaq® DNA polymerase, REDTaq® DNA polymerase, Phusion® DNA polymerase, Phusion® High-Fidelity DNA polymerase, Platinum Pfx DNA polymerase, AccuPrime Pfx DNA polymerase, Phi29 DNA polymerase, Klenow fragment, Pwo DNA polymerase, Pfu DNA polymerase, T4 DNA polymerase and T7 DNA polymerase enzymes.

**[0071]** In some embodiments, genetic material is amplified by reverse transcription polymerase chain reaction (RT-PCR). The desired reverse transcriptase activity can be provided by one or more distinct reverse transcriptase enzymes, suitable examples of which include, but are not limited to: M-MLV, MuLV, AMV, HIV, ArrayScript™, MultiScribe™, ThermoScript™, and SuperScript® I, II, III, and IV enzymes. “Reverse transcriptase” includes not only naturally occurring enzymes, but all such modified derivatives thereof, including also derivatives of naturally-occurring reverse transcriptase enzymes.

**[0072]** In addition, reverse transcription can be performed using sequence-modified derivatives or mutants of M-MLV, MuLV, AMV, and HIV reverse transcriptase enzymes, including mutants that retain at least some of the functional, e.g., reverse transcriptase, activity of the wild-type sequence. The reverse transcriptase enzyme can be provided as part of a composition that includes other components, e.g., stabilizing components that enhance or improve the activity of the reverse transcriptase enzyme, such as RNase inhibitor(s), inhibitors of DNA-dependent DNA synthesis, e.g., actinomycin D. Many sequence-modified derivative or mutants of reverse transcriptase enzymes, e.g., M-MLV, and compositions including unmodified and modified enzymes are commercially available, e.g., ArrayScript™, MultiScribe™, ThermoScript™, and SuperScript® I, II, III, and IV enzymes.

**[0073]** Certain reverse transcriptase enzymes (e.g., Avian Myeloblastosis Virus (AMV) Reverse Transcriptase and Moloney Murine Leukemia Virus (M-MuLV, MMLV) Reverse Transcriptase) can synthesize a complementary DNA strand using both RNA (cDNA synthesis) and single-stranded DNA (ssDNA) as a template. Thus, in some embodiments, the reverse transcription reaction can use an enzyme (reverse transcriptase) that is capable of using both RNA and ssDNA as the template for an extension reaction, e.g., an AMV or MMLV reverse transcriptase.

## II. Overview

[0074] In some embodiments, provided herein are methods to detect different base combination of nucleic acids (e.g., ssDNA in clonal clusters) which are immobilized on surface (e.g., a surface in a flow cell for sequencing). The general base decoding processing are illustrated, by way of example, in **FIG. 2**, using DNA clusters as examples.

[0075] In the particular example, the DNA clusters which are amplified by surface PCR are immobilized on surface. Each DNA clusters contains hundreds to thousands of copies of the same DNA sequence which is to be sequenced. Any two different clusters can contain the same DNA sequence or different DNA sequences to be sequenced. There are four different bases shown in **FIG. 2** which represent the four possible initial configurations at each cluster at the beginning of DNA sequencing. The mixture of three dNTPs, e.g., dATP, dCTP and dTTP with different labels, and a polymerase can be flowed through (e.g., through a flow cell in which the clusters are immobilized on a substrate) and contacted with the ssDNA strands hybridized with sequencing primers. The matched dNTPs are incorporated onto the strands and the missing dNTP (e.g., dGTP) is not incorporated. The excess dNTPs can be washed off and fluorescent imaging can be performed under two different illumination sources, e.g., 530 nm and 630 nm. In **FIG. 2**, dATP labelled with fluorescent dye (L2) can be detected under 630 nm excitation wavelength (red). dTTP with L1 fluorescent label can be detected under green light (530 nm). dCTP which is labelled with two dyes (dCTP-L3+L4) or a mixture of nucleotide with two different dyes (e.g., dCTP-L3 and dCTP-L4) can be detected with both illumination conditions. Red and green colors are used as examples and fluorescent labels of other colors can be used. Wavelengths are often referred to using their associated color and include blue (400-470 nm), green (470-550 nm), red (630-700 nm) and NIR (700-1200) lights.

[0076] In the particular example, the image 1 is acquired with 530 nm LED illumination at full power (e.g., constant current at 12 Amps). The image 2 is acquired with 630 nm LED illumination at full power as well (e.g., constant current at 12 Amps). The base types can be called based on these two images. The base is called as "A" at a cluster if the cluster only lights up in Image 2. The base at a cluster is "T" if the cluster only lights up in Image 1. The base is "C" if the cluster is detectable in both Image 1 and Image 2. The base is "G" if the cluster is not detected in either image. However, the fluorescent dyes selections can be very limit for dual labelled or mixed dyes to have the same brightness as single labelled nucleotide in red or green channel, since the concentration is split into half by two dyes. The dual labelled dyes will have energy transfer between dyes or different quantum yield, which can

make it difficult to balance the brightness between these dyes. The ideal distribution of the intensity of DNA clusters in Image 1 and Image 2 is shown in **FIG. 3**. The average signal intensities of T and C in Image 1 is similar or equal and same for A and C in Image 2. However, in reality, in sequencing images the average signal intensity of C can be far weaker than T and A as demonstrated in **FIG. 4**, e.g., due to energy transfer between dyes.

**[0077]** To improve the distance between A, T, G, and C clouds of clusters, in some embodiments a third image can be acquired with both wavelengths, e.g., using 530 nm and 630 nm light sources (LED1 and LED2, respectively), where the clusters are illuminated but with 50% of power for LED1 and 50% of power of LED2 (12 Amps current in total). The dual labelled nucleotide can show 100% of the emission intensity since it can be excited by both light. The other two nucleotide with single labelled dye can only reach 50% of intensity in Image 1 and Image 2. These intensities can be used as third axis to improve the separation between the cluster intensity clouds. As shown in **FIG. 5**, the distance between dark cloud and the dual labeled “C” cloud can be increased for 1.414 to 1.732 which is about 20% improvement compared to an ideal case using a conventional method (e.g., using only Image 1 and Image 2). As shown in **FIG. 6**, the distance between dark cloud and the dual labeled “C” cloud can be improved by using Image 3, even when the dual color nucleotide (“C”) has intensity at red or green channel that is half of the red only dye or the green only dye. The distance between dark (“G”) and “C” clouds is increased from 0.707 to 0.866. The pass filter rate and quality score can be improved with increased distance between the cluster clouds.

**[0078]** To improve the brightness balance between nucleotides, the nucleotides can be labelled with non-fluorescent labels such as biotin, desthiobiotin, 2,4-Dinitrophenyl (DNP) and Digoxigenin etc., which can be stained with dye labelled streptavidin or antibody. The intensities of clusters can be easily tuned by varying streptavidin/dye ratio or antibody/ratio. In addition, the intensities can be also tune by mixing of streptavidin without any tag. The dual channel brightness can be simply tuned by fine tuning the ratio between streptavidin-dye1, streptavidin-dye2, and optionally unlabeled streptavidin, as show in **FIG. 7** and **FIG. 8**. In some embodiments, the protein/dye ratio can be achieved to reach more than 5 dye molecules per protein molecule.

**[0079]** The missing nucleotide such as G in **FIG. 2** can be incorporated by contacting the dGTP with or without labels and polymerase in a separate step or can be introduced together with other unlabeled or labelled nucleotides such as dATP, dCTP and dTTP to improve the incorporation yield and prevent grow lag. The advantage of introducing another dNTP mix after imaging is the following. In some embodiments, to improve the efficiency the dark

nucleotide can be mixed with other dye labelled nucleotides, e.g., 10% of total nucleotides can be unlabeled, which can improve the phasing too since dark dNTP is incorporated more efficiency than labelled nucleotides. In some embodiments, the missing nucleotide can be labeled with biotin or other hapten, a third or fourth image can be acquired to light up the fourth type of clusters such as G in **FIG. 4** to improve the accuracy of base calling after stain with streptavidin or antibody with dyes.

**[0080]** The internal fluorescence dye can be replaced with other biomolecules such as biotin, desthiobiotin, 2,4-Dinitrophenol and digoxigenin etc. The fluorescent dye can be introduced later with streptavidin for biotin and desthiobiotin or antibody for small hapten. The dye tag can also be released with biotin for desthiobiotin labeled oligo.

### **III. Samples and Nucleic Acid Molecules**

**[0081]** The nucleic acid molecules used in the methods described herein may be obtained from any suitable biological source, for example a tissue sample, a blood sample, a plasma sample, a saliva sample, a fecal sample, or a urine sample. The polynucleotides may be DNA or RNA molecules. In some embodiments, RNA molecules are reverse transcribed into DNA molecules prior to hybridizing the polynucleotide to a sequencing primer. In some embodiments, RNA molecules are not reverse transcribed and are hybridized to a sequencing primer for direct RNA sequencing. In some embodiments, the nucleic acid molecule is a cell-free DNA (cfDNA), such as a circulating tumor DNA (ctDNA) or a fetal cell-free DNA.

**[0082]** Examples of nucleic acid molecules include DNA molecules such as single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), genomic DNA, methylated DNA, specific methylated DNA sequences, fragmented DNA, mitochondrial DNA, in situ synthesized PCR products, and RNA/DNA hybrids. The DNA analyte can be a transcript of another nucleic acid molecule (e.g., DNA or RNA such as mRNA) present in a tissue sample.

**[0083]** Examples of nucleic acid molecules also include RNA molecules such as various types of coding and non-coding RNA, including viral RNAs. Examples of the different types of RNA molecules include messenger RNA (mRNA), including a nascent RNA, a pre-mRNA, a primary-transcript RNA, and a processed RNA, such as a capped mRNA (e.g., with a 5' 7-methyl guanosine cap), a polyadenylated mRNA (poly-A tail at the 3' end), and a spliced mRNA in which one or more introns have been removed. Also included in the nucleic acid molecules disclosed herein are non-capped mRNA, a non-polyadenylated

mRNA, and a non-spliced mRNA. The RNA analyte can be a transcript of another nucleic acid molecule (e.g., DNA or RNA such as viral RNA).

**[0084]** In some embodiments, a nucleic acid molecule may be a denatured nucleic acid, wherein the resulting denatured nucleic acid is single-stranded. The nucleic acid may be denatured, for example, optionally using formamide, heat, or both formamide and heat. In some embodiments, the nucleic acid is not denatured for use in a method disclosed herein.

**[0085]** In some embodiments, a nucleic acid molecule can be extracted from a cell, a virus, or a tissue sample comprising the cell or virus. Processing conditions can be adjusted to extract or release nucleic acid molecules (e.g., RNA) from a cell, a virus, or a tissue sample.

#### **IV. Sequencing Methods**

##### **A. Nucleotides and Nucleotide Analogs**

**[0086]** In some embodiments, a method disclosed herein comprises using one or more nucleotides or analogs thereof, including a native nucleotide or a nucleotide analog or modified nucleotide (e.g., labeled with one or more detectable labels). In some embodiments, a nucleotide analog comprises a nitrogenous base, five-carbon sugar, and phosphate group, wherein any component of the nucleotide may be modified and/or replaced. In some embodiments, a method disclosed herein may comprise but does not require using one or more non-incorporable nucleotides. Non-incorporable nucleotides may be modified to become incorporable at any point during the sequencing method.

**[0087]** Nucleotide analogs include, but are not limited to, alpha-phosphate modified nucleotides, alpha-beta nucleotide analogs, beta-phosphate modified nucleotides, beta-gamma nucleotide analogs, gamma-phosphate modified nucleotides, caged nucleotides, or ddNTPs. Examples of nucleotide analogs are described in U.S. Patent No. 8,071,755, which is incorporated by reference herein in its entirety.

**[0088]** In some embodiments, a method disclosed herein may comprise but does not require using terminators that reversibly prevent nucleotide incorporation at the 3'-end of the primer. One type of reversible terminator is a 3'-O-blocked reversible terminator. Here the terminator moiety is linked to the oxygen atom of the 3'-OH end of the 5-carbon sugar of a nucleotide. For example, U.S. Patent Nos. 7,544,794 and 8,034,923 (the disclosures of these patents are incorporated by reference) describe reversible terminator dNTPs having the 3'-OH group replaced by a 3'-ONH<sub>2</sub> group. Another type of reversible terminator is a 3'-unblocked reversible terminator, wherein the terminator moiety is linked to the nitrogenous base of a

nucleotide. For example, U.S. Patent No. 8,808,989 (the disclosure of which is incorporated by reference) discloses particular examples of base-modified reversible terminator nucleotides that may be used in connection with the methods described herein. Other reversible terminators that similarly can be used in connection with the methods described herein include those described in U.S. Patent Nos. 7,956,171, 8,071,755, and 9,399,798, herein incorporated by reference.

**[0089]** In some embodiments, a method disclosed herein may comprise but does not require using nucleotide analogs having terminator moieties that irreversibly prevent nucleotide incorporation at the 3'-end of the primer. Irreversible nucleotide analogs include 2', 3'-dideoxynucleotides, ddNTPs (ddGTP, ddATP, ddTTP, ddCTP). Dideoxynucleotides lack the 3'-OH group of dNTPs that is essential for polymerase-mediated synthesis.

**[0090]** In some embodiments, a method disclosed herein may comprise but does not require using non-incorporable nucleotides comprising a blocking moiety that inhibits or prevents the nucleotide from forming a covalent linkage to a second nucleotide (3'-OH of a primer) during the incorporation step of a nucleic acid polymerization reaction. The blocking moiety can be removed from the nucleotide, allowing for nucleotide incorporation.

**[0091]** In some embodiments, a method disclosed herein may comprise but does not require using 1, 2, 3, 4 or more nucleotide analogs present in the SBS reaction. In some embodiments, a nucleotide analog is replaced, diluted, or sequestered during an incorporation step. In some embodiments, a nucleotide analog is replaced with a native nucleotide. In some embodiments, a nucleotide analog is modified during an incorporation step. The modified nucleotide analog can be similar to or the same as a native nucleotide.

**[0092]** In some embodiments, a method disclosed herein may comprise but does not require using a nucleotide analog having a different binding affinity for a polymerase than a native nucleotide. In some embodiments, a nucleotide analog has a different interaction with a next base than a native nucleotide. Nucleotide analogs and/or non-incorporable nucleotides may base-pair with a complementary base of a template nucleic acid.

**[0093]** In some embodiments, one or more nucleotides can be labeled with distinguishing and/or detectable tags or labels. The tags may be distinguishable by means of their differences in fluorescence, Raman spectrum, charge, mass, refractive index, luminescence, length, or any other measurable property. The tag may be attached to one or more different positions on the nucleotide, so long as the fidelity of binding to the polymerase-nucleic acid complex is sufficiently maintained to enable identification of the complementary base on the template nucleic acid correctly. In some embodiments, the tag is attached to the nucleobase of



the nucleotide. Alternatively, a tag is attached to the gamma phosphate position of the nucleotide.

**[0094]** Detectable labels can be suitable for small scale detection and/or suitable for high-throughput screening. As such, suitable detectable labels include, but are not limited to, radioisotopes, fluorophores, chemiluminescent compounds, bioluminescent compounds, and dyes. The detectable label can be qualitatively detected (e.g., optically or spectrally), or it can be quantified. Qualitative detection generally includes a detection method in which the existence or presence of the detectable label is confirmed, whereas quantifiable detection generally includes a detection method having a quantifiable (e.g., numerically reportable) value such as an intensity, duration, polarization, and/or other properties. In some embodiments, the detectable label is bound to another moiety, for example, a nucleotide or nucleotide analog, and can include a fluorescent, a colorimetric, or a chemiluminescent label.

**[0095]** In some embodiments, a detectable label can be attached to another moiety, for example, a nucleotide or nucleotide analog. In some embodiments, the detectable label is a fluorophore. For example, the fluorophore can be from a group that includes: 7-AAD (7-Aminoactinomycin D), Acridine Orange (+DNA), Acridine Orange (+RNA), Alexa Fluor® 350, Alexa Fluor® 430, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 555, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 633, Alexa Fluor® 647, Alexa Fluor® 660, Alexa Fluor® 680, Alexa Fluor® 700, Alexa Fluor® 750, Allophycocyanin (APC), AMCA / AMCA-X, 7-Aminoactinomycin D (7-AAD), 7-Amino-4-methylcoumarin, 6-Aminoquinoline, Aniline Blue, ANS, APC-Cy7, ATTO-TAG™ CBQCA, ATTO-TAG™ FQ, Auramine O-Feulgen, BCECF (high pH), BFP (Blue Fluorescent Protein), BFP / GFP FRET, BOBO™-1 / BO-PRO™-1, BOBO™-3 / BO-PRO™-3, BODIPY® FL, BODIPY® TMR, BODIPY® TR-X, BODIPY® 530/550, BODIPY® 558/568, BODIPY® 564/570, BODIPY® 581/591, BODIPY® 630/650-X, BODIPY® 650-665-X, BTC, Calcein, Calcein Blue, Calcium Crimson™, Calcium Green-1™, Calcium Orange™, Calcofluor® White, 5-Carboxyfluorescein (5-FAM), 5-Carboxynaphthofluorescein, 6-Carboxyrhodamine 6G, 5-Carboxytetramethylrhodamine (5-TAMRA), Carboxy-X-rhodamine (5-ROX), Cascade Blue®, Cascade Yellow™, CCF2 (GeneBLAzer™), CFP (Cyan Fluorescent Protein), CFP / YFP FRET, Chromomycin A3, Cl-NERF (low pH), CPM, 6-CR 6G, CTC Formazan, Cy2®, Cy3®, Cy3.5®, Cy5®, Cy5.5®, Cy7®, Cychrome (PE-Cy5), Dansylamine, Dansyl cadaverine, Dansylchloride, DAPI, Dapoxyl, DCFH, DHR, DiA (4-Di-16-ASP), DiD (DiIC18(5)), DIDS, DiI (DiIC18(3)), DiO (DiOC18(3)), DiR (DiIC18(7)), Di-4 ANEPPS, Di-8 ANEPPS, DM-NERF (4.5-6.5 pH),

DsRed (Red Fluorescent Protein), EBFP, ECFP, EGFP, ELF® -97 alcohol, Eosin, Erythrosin, Ethidium bromide, Ethidium homodimer-1 (EthD-1), Europium (III) Chloride, 5-FAM (5-Carboxyfluorescein), Fast Blue, Fluorescein-dT phosphoramidite, FITC, Fluo-3, Fluo-4, FluorX®, Fluoro-Gold™ (high pH), Fluoro-Gold™ (low pH), Fluoro-Jade, FM® 1-43, Fura-2 (high calcium), Fura-2 / BCECF, Fura Red™ (high calcium), Fura Red™ / Fluo-3, GeneBLAzer™ (CCF2), GFP Red Shifted (rsGFP), GFP Wild Type, GFP / BFP FRET, GFP / DsRed FRET, Hoechst 33342 & 33258, 7-Hydroxy-4-methylcoumarin (pH 9), 1,5 IAEDANS, Indo-1 (high calcium), Indo-1 (low calcium), Indodicarbocyanine, Indotricarbocyanine, JC-1, 6-JOE, JOJO™-1 / JO-PRO™-1, LDS 751 (+DNA), LDS 751 (+RNA), LOLO™-1 / LO-PRO™-1, Lucifer Yellow, LysoSensor™ Blue (pH 5), LysoSensor™ Green (pH 5), LysoSensor™ Yellow/Blue (pH 4.2), LysoTracker® Green, LysoTracker® Red, LysoTracker® Yellow, Mag-Fura-2, Mag-Indo-1, Magnesium Green™, Marina Blue®, 4-Methylumbelliferone, Mithramycin, MitoTracker® Green, MitoTracker® Orange, MitoTracker® Red, NBD (amine), Nile Red, Oregon Green® 488, Oregon Green® 500, Oregon Green® 514, Pacific Blue, PBF1, PE (R-phycoerythrin), PE-Cy5, PE-Cy7, PE-Texas Red, PerCP (Peridinin chlorophyll protein), PerCP-Cy5.5 (TruRed), PharRed (APC-Cy7), C-phycocyanin, R-phycocyanin, R-phycoerythrin (PE), PI (Propidium Iodide), PKH26, PKH67, POPO™-1 / PO-PRO™-1, POPO™-3 / PO-PRO™-3, Propidium Iodide (PI), PyMPO, Pyrene, Pyronin Y, Quantam Red (PE-Cy5), Quinacrine Mustard, R670 (PE-Cy5), Red 613 (PE-Texas Red) , Red Fluorescent Protein (DsRed), Resorufin, RH 414, Rhod-2, Rhodamine B, Rhodamine Green™, Rhodamine Red™, Rhodamine Phalloidin, Rhodamine 110, Rhodamine 123, 5-ROX (carboxy-X-rhodamine), S65A, S65C, S65L, S65T, SBFI, SITS, SNAFL®-1 (high pH), SNAFL®-2, SNARF®-1 (high pH), SNARF®-1 (low pH), Sodium Green™, SpectrumAqua®, SpectrumGreen® #1, SpectrumGreen® #2, SpectrumOrange®, SpectrumRed®, SYTO® 11, SYTO® 13, SYTO® 17, SYTO® 45, SYTOX® Blue, SYTOX® Green, SYTOX® Orange, 5-TAMRA (5-Carboxytetramethylrhodamine), Tetramethylrhodamine (TRITC), Texas Red® / Texas Red®-X, Texas Red®-X (NHS Ester), Thiadcarbocyanine, Thiazole Orange, TOTO®-1 / TO-PRO®-1, TOTO®-3 / TO-PRO®-3, TO-PRO®-5, Tri-color (PE-Cy5), TRITC (Tetramethylrhodamine), TruRed (PerCP-Cy5.5), WW 781, X-Rhodamine (XRITC) , Y66F, Y66H, Y66W, YFP (Yellow Fluorescent Protein), YOYO®-1 / YO-PRO®-1, YOYO®-3 / YO-PRO®-3, 6-FAM (Fluorescein), 6-FAM (NHS Ester), 6-FAM (Azide), HEX, TAMRA (NHS Ester), Yakima Yellow, MAX, TET, TEX615, ATTO 488, ATTO 532, ATTO 542, ATTO 550, ATTO 565, ATTO Rho101, ATTO 590, ATTO 633, ATTO 647N, TYE 563,

TYE 665, TYE 705, 5' IRDye® 700, 5' IRDye® 800, 5' IRDye® 800CW (NHS Ester), WellRED D4 Dye, WellRED D3 Dye, WellRED D2 Dye, Lightcycler® 640 (NHS Ester), and Dy 750 (NHS Ester).

**[0096]** The detectable label can be directly detectable by itself (e.g., radioisotope labels or fluorescent labels) or, in the case of an enzymatic label, can be indirectly detectable, e.g., by catalyzing chemical alterations of a substrate compound or composition, which substrate compound or composition is directly detectable. The label can emit a signal or alter a signal delivered to the label so that the presence or absence of the label can be detected. In some cases, coupling may be via a linker, which may be cleavable, such as photo-cleavable (e.g., cleavable under ultra-violet light), chemically-cleavable (e.g., via a reducing agent, such as dithiothreitol (DTT), tris(2-carboxyethyl)phosphine (TCEP)) or enzymatically cleavable (e.g., via an esterase, lipase, peptidase, or protease).

## **B. Polymerases**

**[0097]** Polymerases that may be used to carry out the disclosed techniques include naturally-occurring polymerases and any modified variations thereof, including, but not limited to, mutants, recombinants, fusions, genetic modifications, chemical modifications, synthetics, and analogs. Naturally occurring polymerases and modified variations thereof are not limited to polymerases that retain the ability to catalyze a polymerization reaction. In some embodiments, the naturally occurring and/or modified variations thereof retain the ability to catalyze a polymerization reaction. In some embodiments, the naturally-occurring and/or modified variations have special properties that enhance their ability to sequence DNA, including enhanced binding affinity to nucleic acids, reduced binding affinity to nucleic acids, enhanced catalysis rates, reduced catalysis rates, etc. Mutant polymerases include polymerases wherein one or more amino acids are replaced with other amino acids (naturally or non-naturally occurring), and insertions or deletions of one or more amino acids.

**[0098]** In some embodiments, a method disclosed herein may comprise but does not require using modified polymerases containing an external tag (e.g., an exogenous detectable label), which can be used to monitor the presence and interactions of the polymerase. In some embodiments, intrinsic signals from the polymerase can be used to monitor their presence and interactions. Thus, the provided methods can include monitoring the interaction of the polymerase, nucleotide and template nucleic acid through detection of an intrinsic signal from the polymerase. In some embodiments, the intrinsic signal is a light scattering signal.

For example, intrinsic signals include native fluorescence of certain amino acids such as tryptophan.

**[0099]** In some embodiments, a method disclosed herein may comprise using an unlabeled polymerase, and monitoring is performed in the absence of an exogenous detectable label associated with the polymerase. Some modified polymerases or naturally occurring polymerases, under specific reaction conditions, may incorporate only single nucleotides and may remain bound to the primer-template after the incorporation of the single nucleotide.

**[0100]** In some embodiments, a method disclosed herein may comprise using a polymerase unlabeled with an exogenous detectable label (e.g., a fluorescent label). The label can be chemically linked to the structure of the polymerase by a covalent bond after the polymerase has been at least partially purified using protein isolation techniques. For example, the exogenous detectable label can be chemically linked to the polymerase using a free sulfhydryl or a free amine moiety of the polymerase. This can involve chemical linkage to the polymerase through the side chain of a cysteine residue, or through the free amino group of the N-terminus. In certain preferred embodiments, a fluorescent label attached to the polymerase is useful for locating the polymerase, as may be important for determining whether or not the polymerase has localized to a spot on an array corresponding to immobilized primed template nucleic acid. The fluorescent signal need not, and in some embodiments does not change absorption or emission characteristics as the result of binding any nucleotide. In some embodiments, the signal emitted by the labeled polymerase is maintained uniformly in the presence and absence of any nucleotide being investigated as a possible next correct nucleotide.

**[0101]** The term polymerase and its variants, as used herein, also refers to fusion proteins comprising at least two portions linked to each other, for example, where one portion comprises a peptide that can catalyze the polymerization of nucleotides into a nucleic acid strand is linked to another portion that comprises a second moiety, such as, a reporter enzyme or a processivity-modifying domain. For example, T7 DNA polymerase comprises a nucleic acid polymerizing domain and a thioredoxin binding domain, wherein thioredoxin binding enhances the processivity of the polymerase. Absent the thioredoxin binding, T7 DNA polymerase is a distributive polymerase with processivity of only one to a few bases. Although DNA polymerases differ in detail, they have a similar overall shape of a hand with specific regions referred to as the fingers, the palm, and the thumb; and a similar overall structural transition, comprising the movement of the thumb and/or finger domains, during the synthesis of nucleic acids.

[0102] DNA polymerases include, but are not limited to, bacterial DNA polymerases, eukaryotic DNA polymerases, archaeal DNA polymerases, viral DNA polymerases and phage DNA polymerases. Bacterial DNA polymerases include *E. coli* DNA polymerases I, II and III, IV and V, the Klenow fragment of *E. coli* DNA polymerase, *Clostridium stercorarium* (*Cst*) DNA polymerase, *Clostridium thermocellum* (*Cth*) DNA polymerase and *Sulfolobus solfataricus* (*Sso*) DNA polymerase. Eukaryotic DNA polymerases include DNA polymerases  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\eta$ ,  $\zeta$ ,  $\lambda$ ,  $\sigma$ ,  $\mu$ , and  $\kappa$ , as well as the Rev1 polymerase (terminal deoxycytidyl transferase) and terminal deoxynucleotidyl transferase (TdT). Viral DNA polymerases include T4 DNA polymerase, phi-29 DNA polymerase, GA-1, phi-29-like DNA polymerases, PZA DNA polymerase, phi-15 DNA polymerase, Cpl DNA polymerase, Cp7 DNA polymerase, T7 DNA polymerase, and T4 polymerase. Other DNA polymerases include thermostable and/or thermophilic DNA polymerases such as DNA polymerases isolated from *Thermus aquaticus* (*Taq*) DNA polymerase, *Thermus filiformis* (*Tfi*) DNA polymerase, *Thermococcus zilligi* (*Tzi*) DNA polymerase, *Thermus thermophilus* (*Tth*) DNA polymerase, *Thermus flavus* (*Tfl*) DNA polymerase, *Pyrococcus woesei* (*Pwo*) DNA polymerase, *Pyrococcus furiosus* (*Pfu*) DNA polymerase and Turbo Pfu DNA polymerase, *Thermococcus litoralis* (*Tli*) DNA polymerase, *Pyrococcus* sp. GB-D polymerase, *Thermotoga maritima* (*Tma*) DNA polymerase, *Bacillus stearothermophilus* (*Bst*) DNA polymerase, *Pyrococcus Kodakaraensis* (*KOD*) DNA polymerase, Pfx DNA polymerase, *Thermococcus* sp. JDF-3 (*JDF-3*) DNA polymerase, *Thermococcus gorgonarius* (*Tgo*) DNA polymerase, *Thermococcus acidophilium* DNA polymerase; *Sulfolobus acidocaldarius* DNA polymerase; *Thermococcus* sp. go N-7 DNA polymerase; *Pyrodictium occultum* DNA polymerase; *Methanococcus voltae* DNA polymerase; *Methanococcus thermoautotrophicum* DNA polymerase; *Methanococcus jannaschii* DNA polymerase; *Desulfurococcus* strain TOK DNA polymerase (*D. Tok Pol*); *Pyrococcus abyssi* DNA polymerase; *Pyrococcus horikoshii* DNA polymerase; *Pyrococcus islandicum* DNA polymerase; *Thermococcus fumicolans* DNA polymerase; *Aeropyrum pernix* DNA polymerase; and the heterodimeric DNA polymerase DP1/DP2. Engineered and modified polymerases also are useful in connection with the disclosed techniques. For example, modified versions of the extremely thermophilic marine archaea *Thermococcus* species 9° N (e.g., Terminator DNA polymerase from New England BioLabs Inc.; Ipswich, Mass.) can be used. Still other useful DNA polymerases, including the 3PDX polymerase are disclosed in U.S. Patent No. 8,703,461, the disclosure of which is incorporated by reference in its entirety.

**[0103]** RNA polymerases include, but are not limited to, viral RNA polymerases such as T7 RNA polymerase, T3 polymerase, SP6 polymerase, and K11 polymerase; Eukaryotic RNA polymerases such as RNA polymerase I, RNA polymerase II, RNA polymerase III, RNA polymerase IV, and RNA polymerase V; and Archaea RNA polymerase.

**[0104]** Reverse transcriptases include, but are not limited to, HIV-1 reverse transcriptase from human immunodeficiency virus type 1 (PDB 1HMY), HIV-2 reverse transcriptase from human immunodeficiency virus type 2, M-MLV reverse transcriptase from the Moloney murine leukemia virus, AMV reverse transcriptase from the avian myeloblastosis virus, and Telomerase reverse transcriptase that maintains the telomeres of eukaryotic chromosomes.

### **C. Sequencing Reactions**

**[0105]** In some embodiments of a sequencing-by-synthesis (SBS) method disclosed herein, a first labeled nucleotide that has been incorporated is not deactivated (e.g., by removal and/or photobleaching of the label) prior to the introduction and/or incorporation of the next, second labeled nucleotide. The first and second labeled nucleotides can comprise the same base or different bases. The first and second labeled nucleotides can be introduced into a sequencing reaction mix simultaneously or at different time points in any order. Further, the first and second labeled nucleotides can be introduced by itself (e.g., in a suitable solvent such as water) or in a mixture with another sequencing reagent, such as one or more other labeled nucleotides and/or one or more unlabeled nucleotides. The first and second labeled nucleotides can also comprise the same base or different bases. In some embodiments, nucleotides that have not been incorporated at a residue corresponding to a base in the template nucleic acid (e.g., because the first labeled nucleotide has been incorporated at that residue) are not removed from the sequencing reaction mix prior to the introduction and/or incorporation of the second labeled nucleotide. In some embodiments, the first and second labeled nucleotides (and optionally labeled nucleotides for interrogating subsequent bases in the template) are provided in the same sequencing reaction mix, and the first, second, and optionally any subsequent labeled nucleotide(s) are incorporated sequentially in a continuous manner.

**[0106]** Thus, unlike existing SBS methods, some embodiments of the method disclosed herein use continuous introduction and/or incorporation of nucleotides (e.g., fluorescently labeled A, T, C, and/or G nucleotides) without the need of label deactivation and/or wash steps in between sequential incorporation events for a given template nucleic acid molecule

to be sequenced. Rather, in some embodiments, label deactivation (e.g., by cleaving and/or photobleaching the label) of a first incorporated nucleotide may occur stochastically throughout the continuous nucleotide incorporation process, for instance, prior to, during, or after the incorporation of a second, third, fourth, or a subsequent labeled nucleotide.

**[0107]** Nucleic acid sequencing reaction mixtures, or simply “reaction mixtures,” typically include reagents that are commonly present in polymerase based nucleic acid synthesis reactions. The reaction mixture can include other molecules including, but not limited to, enzymes. In some embodiments, the reaction mixture comprises any reagents or biomolecules generally present in a nucleic acid polymerization reaction. Reaction components may include, but are not limited to, salts, buffers, small molecules, detergents, crowding agents, metals, and ions. In some embodiments, properties of the reaction mixture may be manipulated, for example, electrically, magnetically, and/or with vibration.

**[0108]** The provided methods herein may further comprise but do not require one or more wash steps; a temperature change; a mechanical vibration; a pH change; or an optical stimulation that is not dye illumination or photobleaching. In some embodiments, the wash step comprises contacting the substrate and the nucleic acid molecule, the primer, and/or the polymerase with one or more buffers, detergents, protein denaturants, proteases, oxidizing agents, reducing agents, or other agents capable of crosslinking or releasing crosslinks, e.g., crosslinks within a polymerase or crosslinks between a polymerase and nucleic acid. Methods and compositions for nucleic acid sequencing are known, for example, as described in U.S. Patent Nos. 10,246,744 and 10,844,428, incorporated herein by reference in their entireties for all purposes.

**[0109]** Reaction mixture reagents can include, but are not limited to, enzymes (e.g., polymerase), dNTPs, template nucleic acids, primer nucleic acids, salts, buffers, small molecules, co-factors, metals, and ions. The ions may be catalytic ions, divalent catalytic ions, non-catalytic ions, non-covalent metal ions, or a combination thereof. The reaction mixture can include salts, such as NaCl, KCl, potassium acetate, ammonium acetate, potassium glutamate, or NH<sub>4</sub>Cl or the like, that ionize in aqueous solution to yield monovalent cations. The reaction mixture can include a source of ions, such as Mg<sup>2+</sup>, Mn<sup>2+</sup>, Co<sup>2+</sup>, Cd<sup>2+</sup>, and/or Ba<sup>2+</sup> ions. The reaction mixture can include tin, Ca<sup>2+</sup>, Zn<sup>2+</sup>, Cu<sup>2+</sup>, Co<sup>2+</sup>, Fe<sup>2+</sup>, and/or Ni<sup>2+</sup>, or other divalent non-catalytic metal cations. In some embodiments, the reaction mixture can include metal cations that may inhibit formation of phosphodiester bonds between the primed template nucleic acid molecule and the cognate nucleotide. In some embodiments, the metal cations can be used (e.g., at a suitable

concentration) to slow down but not completely inhibit or prevent nucleotide incorporation, thereby reducing multiple nucleotide incorporation events in a single detection window.

**[0110]** In some embodiments, the sequencing reaction conditions comprise contacting the nucleic acid molecule and the primer with a buffer that regulates osmotic pressure. In some embodiments, the reaction mixture comprises a buffer that regulates osmotic pressure. In some embodiments, the buffer is a high salt buffer that includes a monovalent ion, such as a monovalent metal ion (e.g., potassium ion or sodium ion) at a concentration of from about 50 to about 1,500 mM. Salt concentrations in the range of from about 100 to about 1,500 mM, or from about 200 to 1,000 mM may also be used. In some embodiments, the buffer further comprises a source of glutamate ions (e.g., potassium glutamate). In some embodiments, the buffer comprises a stabilizing agent. In some embodiments, the stabilizing agent is a non-catalytic metal ion (e.g., a divalent non-catalytic metal ion). Non-catalytic metal ions useful in this context include, but are not limited to, calcium, strontium, scandium, titanium, vanadium, chromium, iron, cobalt, nickel, copper, zinc, gallium, germanium, arsenic, selenium, rhodium, europium, and/or terbium. In some embodiments, the non-catalytic metal ion is strontium, tin, or nickel. In some embodiments, the sequencing reaction mixture comprises strontium chloride or nickel chloride. In some embodiments, the stabilizing agent can be used (e.g., at a suitable concentration) to slow down but not completely inhibit or prevent nucleotide incorporation, thereby reducing multiple nucleotide incorporation events in a single detection window.

**[0111]** The buffer can include Tris, Tricine, HEPES, MOPS, ACES, MES, phosphate-based buffers, and acetate-based buffers. The reaction mixture can include chelating agents such as EDTA, EGTA, and the like. In some embodiments, the reaction mixture includes cross-linking reagents.

**[0112]** In some embodiments, the interaction between the polymerase and template nucleic acid may be manipulated by modulating sequencing reaction parameters such as ionic strength, pH, temperature, or any combination thereof, or by the addition of a destabilizing agent to the reaction. In some embodiments, the destabilizing agent can be used (e.g., at a suitable concentration) to slow down but not completely inhibit or prevent nucleotide incorporation, thereby reducing multiple nucleotide incorporation events in a single detection window.

**[0113]** In some embodiments, high salt (e.g., 50 to 1,500 mM) and/or pH changes are utilized to destabilize a complex between the polymerase and template nucleic acid. In some embodiments, the reaction conditions favor the stabilization of a complex among the



polymerase, the template nucleic acid, and a labeled nucleotide. By way of example, the pH of the reaction mixture can be adjusted from 4.0 to 10.0 to favor the stabilization of a complex among the polymerase, the template nucleic acid, and a labeled nucleotide. In some embodiments, the pH of the reaction mixture is from 4.0 to 6.0. In some embodiments, the pH of the reaction mixture is 6.0 to 10.0. In some embodiments, a suitable salt concentration and/or a suitable pH can be selected to slow down but not completely inhibit or prevent nucleotide incorporation, thereby reducing multiple nucleotide incorporation events in a single detection window.

**[0114]** In some embodiments, the reaction mixture comprises a competitive inhibitor, where the competitive inhibitor may reduce the occurrence of multiple incorporations events in a detection window. In one embodiment, the competitive inhibitor is a non-incorporable nucleotide. In an embodiment, the competitive inhibitor is an aminoglycoside. The competitive inhibitor is capable of replacing either the nucleotide or the catalytic metal ion in the active site, such that the competitive inhibitor occupies the active site preventing or slowing down a nucleotide incorporation. In some embodiments, both an incorporable nucleotide and a competitive inhibitor are introduced, such that the ratio of the incorporable nucleotide and the inhibitor can be adjusted to modulate the rate of incorporation of a single nucleotide at the 3'-end of the primer. In some embodiments, the competitive inhibitor can be used (e.g., at a low concentration) to slow down but not completely inhibit or prevent nucleotide incorporation, thereby reducing multiple nucleotide incorporation events in a single detection window.

**[0115]** In some embodiments, the reaction mixture comprises at least one nucleotide molecule that is a non-incorporable nucleotide. In some embodiments, the reaction mixture comprises one or more nucleotide molecules incapable of incorporation into the primer of the primed template nucleic acid molecule. Such nucleotides incapable of incorporation include, for example, monophosphate nucleotides. For example, the nucleotide may contain modifications to the triphosphate group that make the nucleotide non-incorporable. Examples of non-incorporable nucleotides may be found in U.S. Pat. No. 7,482,120, which is incorporated by reference herein in its entirety. In some embodiments, the primer may not contain a free hydroxyl group at its 3'-end, thereby rendering the primer incapable of incorporating any nucleotide, and, thus, making any nucleotide non-incorporable. In some embodiments, the primer may be processed such that it contains a free hydroxyl group at its 3'-end to allow nucleotide incorporation. In some embodiments, the non-incorporable nucleotide can be used (e.g., at a low concentration) to slow down but not completely inhibit

or prevent nucleotide incorporation, thereby reducing multiple nucleotide incorporation events in a single detection window.

**[0116]** In some embodiments, the reaction mixture comprises at least one nucleotide molecule that is incorporable but is incorporated at a slower rate compared to a corresponding naturally-occurring nucleoside triphosphate (e.g., NTP or dNTP). Such nucleotides incorporable at a slower rate may include, for example, diphosphate nucleotides. For example, the nucleotide may contain modifications to the triphosphate group that make the nucleotide incorporable at a slower rate. In some embodiments, the nucleotide incorporable at a slower rate can be used to slow down but not completely inhibit or prevent nucleotide incorporation, thereby reducing multiple nucleotide incorporation events in a single detection window.

**[0117]** In some embodiments, the reaction mixture comprises a polymerase inhibitor. In some embodiments, the polymerase inhibitor is a pyrophosphate analog. In some embodiments, the polymerase inhibitor is an allosteric inhibitor. In some embodiments, the polymerase inhibitor is a DNA or an RNA aptamer. In some embodiments, the polymerase inhibitor competes with a catalytic-ion binding site in the polymerase. In some embodiments, the polymerase inhibitor is a reverse transcriptase inhibitor. The polymerase inhibitor may be an HIV-1 reverse transcriptase inhibitor or an HIV-2 reverse transcriptase inhibitor. The HIV-1 reverse transcriptase inhibitor may be a (4/6-halogen/MeO/EtO-substituted benzo[d]thiazol-2-yl)thiazolidin-4-one. In some embodiments, the polymerase inhibitor can be used (e.g., at a low concentration) to slow down but not completely inhibit or prevent nucleotide incorporation, thereby reducing multiple nucleotide incorporation events in a single detection window.

**[0118]** In some embodiments, the contacting step is facilitated by the use of a chamber such as a flow cell. The methods and apparatus described herein may employ next generation sequencing technology (NGS), which allows massively parallel sequencing. In some embodiments, single DNA molecules are sequenced in a massively parallel fashion within a reaction chamber. A flow cell may be used but is not necessary. Flowing liquid reagents through the flow cell, which contains an interior solid support surface (e.g., a planar surface), conveniently permits reagent exchange. Immobilized to the interior surface of the flow cell is one or more primed template nucleic acids to be sequenced or interrogated using the procedures described herein. Typical flow cells will include microfluidic valving that permits delivery of liquid reagents (e.g., components of the “reaction mixtures” discussed herein) to

an entry port. Liquid reagents can be removed from the flow cell by exiting through an exit port.

**[0119]** In some embodiments, a reaction chamber disclosed herein can comprise a reagent wall, an imaging area, and optionally an outlet configured to remove molecules of one or more of the polymerase, the first detectably labeled nucleotide, the second detectably labeled nucleotide, and/or one or more other reagents from the imaging area. In some embodiments, the device may comprise one or more vents but no outlet or exit port for the reaction mixture. In some embodiments, a method disclosed herein does not comprise a step of removing liquid reagents through an outlet or exit port, e.g., from a reaction chamber such as a flow cell.

**[0120]** The methods disclosed herein may but do not need to be used in combination with any NGS sequencing methods. The sequencing technologies of NGS include but are not limited to pyrosequencing, sequencing-by-synthesis with reversible dye terminators, sequencing by oligonucleotide probe ligation, and ion semiconductor sequencing. Nucleic acids such as DNA or RNA from individual samples can be sequenced individually (singleplex sequencing) or nucleic acids such as DNA or RNA from multiple samples can be pooled and sequenced as indexed genomic molecules (multiplex sequencing) on a single sequencing run, to generate up to several hundred million reads of sequences. Examples of sequencing technologies that can be used to obtain the sequence information according to the present method are further described here.

**[0121]** Some sequencing technologies are available commercially, such as the sequencing-by-synthesis platforms from 454 Life Sciences (Bradford, Conn.), Illumina/Solexa (Hayward, Calif.) and Helicos Biosciences (Cambridge, Mass.).

**[0122]** While the automated Sanger method is considered as a 'first generation' technology, Sanger sequencing including the automated Sanger sequencing, can also be employed in the methods described herein. Additional suitable sequencing methods include, but are not limited to nucleic acid imaging technologies, e.g., atomic force microscopy (AFM) or transmission electron microscopy (TEM).

**[0123]** In some embodiments, the disclosed methods may be used in combination with massively parallel sequencing of nucleic acid molecules using Illumina's sequencing-by-synthesis and reversible terminator-based sequencing chemistry. In some implementation, a method disclosed herein can use a flow cell having a glass slide with lanes.

**[0124]** After sequencing of nucleic acid molecules, sequence reads of predetermined length, e.g., at least about 15 bp, are localized by mapping (alignment) to a known reference sequence or genome (e.g., viral sequences or genomes). A number of computer algorithms

are available for aligning sequences, including without limitation BLAST, BLITZ, FASTA, BOWTIE, or ELAND (Illumina, Inc., San Diego, Calif., USA).

**[0125]** In some embodiments, the provided sequencing methods disclosed herein may regulate polymerase interaction with the nucleotides and template nucleic acid (as well as rate of nucleotide incorporation) in a manner that reveals the identity of the next base while controlling the chemical addition of a nucleotide. In some embodiments, the SBS reaction condition comprises a plurality of primed template nucleic acids, polymerases, nucleotides, or any combination thereof. In some embodiments, the plurality of nucleotides comprises 1, 2, 3, 4, or more types of different nucleotides, for example dATP, dTTP (or dUTP), dGTP, and dCTP.

**[0126]** In some embodiments, the method can further comprise contacting the nucleic acid molecule with the substrate to immobilize the nucleic acid molecule. In some embodiments, the nucleic acid molecule can be immobilized at a density of one molecule per at least about 250 nm<sup>2</sup>, at least about 200 nm<sup>2</sup>, at least about 150 nm<sup>2</sup>, at least about 100 nm<sup>2</sup>, at least about 90 nm<sup>2</sup>, at least about 80 nm<sup>2</sup>, at least about 70 nm<sup>2</sup>, at least about 60 nm<sup>2</sup>, at least about 50 nm<sup>2</sup>, at least about 40 nm<sup>2</sup>, at least about 30 nm<sup>2</sup>, at least about 20 nm<sup>2</sup>, at least about 10 nm<sup>2</sup>, at least about 5 nm<sup>2</sup>, or in between any two of the aforementioned values. Methods and compositions for arraying biomolecules on a substrate, e.g., as described in US 2005/0042649 (incorporated herein by reference in its entirety for all purposes), may be used in methods disclosed herein.

**[0127]** In some embodiments, a subset of nucleic acid molecules (e.g., nucleic acid strands to be sequenced) on the substrate may be active at one or more time points. In some embodiments, at any one time, a first subset of nucleic acid molecules on the substrate is active (e.g., allowing nucleotide incorporation into a sequencing primer using a single-stranded sequence as template) while a second subset of nucleic acid molecules on the substrate is inactive (e.g., not allowing nucleotide incorporation into a sequencing primer using a single-stranded sequence as template). In some embodiments, at one or more time points, a first subset of nucleic acid molecules on the substrate is activated (e.g., by a first set of polymerase and/or primer molecules) for nucleotide incorporation, while a second subset of nucleic acid molecules on the substrate is not activated (e.g., by the first set of polymerase and/or primer molecules), thus only signals associated with the first subset of nucleic acid molecules are detected. At one or more other time points, the second subset of nucleic acid molecules on the substrate is activated (e.g., by a second set of polymerase and/or primer molecules) for nucleotide incorporation, while the first subset of nucleic acid molecules on

the substrate is not activated (e.g., by the second set of polymerase and/or primer molecules), thus only signals associated with the second subset of nucleic acid molecules are detected. In some embodiments, the first and second sets of polymerase and/or primer molecules can be introduced at different time points, e.g., in sequential cycles with optional washing steps between cycles (e.g., to remove a set of polymerase and/or primer molecules for SBS of a first subset of strands before introducing the next set of polymerase and/or primer molecules for SBS of a second subset of strands).

**[0128]** In some embodiments, the substrate can comprise a bead, a planar substrate, a solid surface, a flow cell, a semiconductor chip, a well, a pillar, a chamber, a channel, a through hole, a nanopore, or any combination thereof. In some embodiments, the substrate can comprise a microwell, a micropillar, a microchamber, a microchannel, or any combination thereof.

## **V. Compositions, Kits, and Applications**

**[0129]** Also provided herein are compositions and kits comprising one or more of the primers, nucleic acid molecules, substrates, nucleotides including detectably labeled nucleotides, polymerases, and reagents for performing the methods provided herein, for example reagents required for one or more steps comprising hybridization, ligation, amplification, detection, sequencing, and/or sample preparation as described herein, for example, in Section IV.

**[0130]** The various components of the kit may be present in separate containers or certain compatible components may be pre-combined into a single container. In some embodiments, the kits further contain instructions for using the components of the kit to practice the provided methods.

**[0131]** In some embodiments, the kits can contain reagents and/or consumables required for performing one or more steps of the provided methods. In some embodiments, the kits contain reagents for sample processing, such as nucleic acid extraction, isolation, and/or purification, e.g., RNA extraction, isolation, and/or purification. In some embodiments, the kits contain reagents, such as enzymes and buffers for ligation and/or amplification, such as ligases and/or polymerases. In some embodiments, the kits contain reagents, such as enzymes and buffers for primer extension and/or nucleic acid sequencing, such as polymerases and/or transcriptases. In some aspects, the kit can also comprise any of the reagents described herein, e.g., buffer components for tuning the rate of nucleotide

incorporation and/or for tuning the rate of signal deactivation (e.g., by photobleaching). In some embodiments, the kits contain reagents for signal detection during sequencing, such as detectable labels and detectably labeled molecules. In some embodiments, the kits optionally contain other components, for example nucleic acid primers, enzymes and reagents, buffers, nucleotides, modified nucleotides, and reagents for additional assays.

**[0132]** In some aspects, the provided embodiments can be applied in analyzing nucleic acid sequences, such as DNA and/or RNA sequencing. In some aspects, the embodiments can be applied in an imaging or detection method for multiplexed nucleic acid analysis. In some aspects, the provided embodiments can be used to identify or detect regions of interest in target nucleic acids, such as viral DNA or RNA. In some embodiments, the region of interest comprises one or more nucleotide residues, such as a single-nucleotide polymorphism (SNP), a single-nucleotide variant (SNV), substitutions such as a single-nucleotide substitution, mutations such as a point mutation, insertions such as a single-nucleotide insertion, deletions such as a single-nucleotide deletion, translocations, inversions, duplications, and/or other sequences of interest.

**[0133]** In some aspects, the embodiments can be applied in investigative and/or diagnostic applications, for example, for characterization or assessment of a sample from a subject. Applications of the provided method can comprise biomedical research and clinical diagnostics. For example, in biomedical research, applications comprise, but are not limited to, genetic and genomic analysis for biological investigation or drug screening. In clinical diagnostics, applications comprise, but are not limited to, detecting gene markers such as disease, immune responses, bacterial or viral DNA/RNA for patient samples, loss of genetic heterozygosity, the presence of gene alleles indicative of a predisposition towards disease or good health, likelihood of responsiveness to therapy, or in personalized medicine or ancestry.

**[0134]** The present disclosure is not intended to be limited in scope to the particular disclosed embodiments, which are provided, for example, to illustrate various aspects of the present disclosure. Various modifications to the compositions and methods described will become apparent from the description and teachings herein. Such variations may be practiced without departing from the true scope and spirit of the disclosure and are intended to fall within the scope of the present disclosure.

## CLAIMS

1. A method for determining a sequence of a polynucleotide template, comprising:
  - a) contacting the polynucleotide template with a first pool of nucleotides comprising:
    - i) a nucleotide of a first base labeled with a first label configured to be detected at a first wavelength,
    - ii) a nucleotide of a second base different from the first base, which nucleotide is labeled with a second label configured to be detected at a second wavelength which is different from the first wavelength, and
    - iii) one or more nucleotides of a third base which is different from the first and second bases, wherein the one or more nucleotides comprise a nucleotide labeled with a third label configured to be detected at the first wavelength and a nucleotide labeled with a fourth label configured to be detected at the second wavelength;
  - b) allowing binding and optional incorporation of a nucleotide of the first pool templated on the polynucleotide template, wherein the bound and optionally incorporated nucleotide is complementary to a nucleotide residue at a first nucleotide position in the polynucleotide template;
  - c) imaging the polynucleotide template to detect a signal (or record an absence of the signal) associated with the bound and optionally incorporated nucleotide of the first pool, and wherein the imaging comprises:
    - i) using a first power to image the polynucleotide template at the first wavelength and not at the second wavelength to acquire a first image, wherein a signal associated with the first label and/or a signal associated with the third label is detected or an absence of the signal is recorded,
    - ii) using a second power to image the polynucleotide template at the second wavelength and not at the first wavelength to acquire a second image, wherein a signal associated with the second label and/or a signal associated with the fourth label is detected or an absence of the signal is recorded, and
    - iii) using a third power to image the polynucleotide template at the first and second wavelengths to acquire a third image, wherein the third power is less than the first power and/or the second power, wherein a signal associated with the first label and/or a signal associated with the third label is detected at a signal intensity lower than that detected in c)i) or an absence of the signal is recorded, and wherein a signal associated with the second label and/or a signal associated with the fourth label is

detected at a signal intensity lower than that detected in c)ii) or an absence of the signal is recorded,

wherein a signal codeword comprising signal codes each corresponding to the intensity of the signal detected in c)i) through c)iii) (or recorded absence of the signal) is generated, wherein the signal codeword corresponds to the identity of the base in the bound and optionally incorporated nucleotide, thereby identifying the nucleotide residue at the first nucleotide position in the polynucleotide template.

2. The method of claim 1, wherein the first label and the third label are different labels.
3. The method of claim 1, wherein the first label and the third label are the same label.
4. The method of any one of claims 1-3, wherein the second label and the fourth label are different labels.
5. The method of any one of claims 1-3, wherein the second label and the fourth label are the same label.
6. The method of any one of claims 1-5, wherein the one or more nucleotides of the third base comprise a nucleotide labeled with the third label and the fourth label.
7. The method of any one of claims 1-6, wherein the one or more nucleotides of the third base comprise a first nucleotide species labeled with the third label and a second nucleotide species labeled with the fourth label.
8. The method of any one of claims 1-7, wherein the third label and the fourth label are different labels.
9. The method of any one of claims 1-7, wherein the third label and the fourth label are the same label.
10. The method of any one of claims 1-9, wherein the first and second wavelengths correspond to different channels of a fluorescence microscope.
11. The method of any one of claims 1-10, wherein in c)i), the signal associated with the first label or the signal associated with the third label is detected at a first signal intensity, and in c)ii), the signal associated with the second label or the signal associated with the fourth label is detected at a second signal intensity.



12. The method of claim 11, wherein the first and second signal intensities differ by no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, or no more than 1%.
13. The method of claim 11, wherein the first and second signal intensities are the same.
14. The method of any one of claims 11-13, wherein in c)iii):
  - the signal associated with the first label or the signal associated with the third label is detected at a signal intensity lower than the first signal intensity; and
  - the signal associated with the second label or the signal associated with the fourth label is detected at a signal intensity lower than the second signal intensity.
15. The method of claim 14, wherein:
  - the signal associated with the first label or the signal associated with the third label is detected at a signal intensity between about 25% and about 75% of the first signal intensity; and
  - the signal associated with the second label or the signal associated with the fourth label is detected at a signal intensity between about 25% and about 75% of the second signal intensity.
16. The method of claim 14, wherein:
  - the signal associated with the first label or the signal associated with the third label is detected at a signal intensity about 50% of the first signal intensity; and
  - the signal associated with the second label or the signal associated with the fourth label is detected at a signal intensity about 50% of the second signal intensity.
17. The method of any one of claims 1-16, wherein the first pool of nucleotides comprises a nucleotide of a fourth base which is different from the first, second, and third bases.
18. The method of claim 17, wherein the nucleotide of the fourth base is not configured to be detected at the first wavelength and/or the second wavelength.
19. The method of claim 17 or claim 18, wherein the nucleotide of the fourth base is not detectably labeled.
20. The method of claim 17 or claim 18, wherein the nucleotide of the fourth base is configured to be detected at a third wavelength different from the first and second wavelengths.

21. The method of any one of claims 1-20, wherein the polynucleotide template is one of a plurality of polynucleotide templates having different template sequences.
22. The method of claim 21, wherein the sequences of the plurality of polynucleotide templates are determined.
23. The method of claim 21 or claim 22, wherein the polynucleotide template is immobilized on a substrate.
24. The method of claim 23, wherein the polynucleotide template is in one or more polynucleotides immobilized on the substrate.
25. The method of claim 24, wherein the polynucleotide template is in a rolling circle amplification product immobilized on the substrate, wherein the rolling circle amplification product comprises multiple copies of the template sequence in the polynucleotide template.
26. The method of claim 24, wherein the polynucleotide template is in a cluster of multiple polynucleotides immobilized on the substrate, wherein each polynucleotide in the cluster comprises a copy of the template sequence in the polynucleotide template.
27. The method of any one of claims 21-26, wherein in c)i), the plurality of polynucleotide templates are imaged, and the first image comprises the signal associated with the first label at the location of a first polynucleotide template, a recorded absence of a signal at the location of a second polynucleotide template, and the signal associated with the third label at the location of a third polynucleotide template.
28. The method of any one of claims 21-27, wherein in c)ii), the plurality of polynucleotide templates are imaged and the second image comprises a recorded absence of a signal at the location of the first polynucleotide template, the signal associated with the second label at the location of the second polynucleotide template, and the signal associated with the fourth label at the location of a third polynucleotide template.
29. The method of any one of claims 21-28, wherein in c)iii), the plurality of polynucleotide templates are imaged and the third image comprises:
  - the signal associated with the first label at the location of the first polynucleotide template,
  - the signal associated with the second label at the location of the second polynucleotide template, and

the signal associated with the third label and the signal associated with the fourth label at the location of a third polynucleotide template.

30. The method of any one of claims 1-29, wherein in c)i) and c)ii), the first and second images are acquired using the same power or powers that differ by no more than 25%, no more than 20%, no more than 15%, no more than 10%, no more than 5%, or no more than 1%, and the c)iii), the third image is acquired using a power that is between about 25% and about 75% of the power(s) used in c)i) and/or c)ii).

31. A method for determining sequences of a plurality of polynucleotide templates having different template sequences, comprising:

a) contacting a substrate having clusters of polynucleotides immobilized thereon with a first pool of nucleotides and with oligonucleotide primers, in any order,

whereby each cluster comprises:

i) multiple copies of one of the different template sequences, and

ii) an oligonucleotide primer of the oligonucleotide primers annealed to a primer binding sequence for extension of the oligonucleotide primer templated on a copy of the template sequence,

wherein the first pool of nucleotides comprises:

i) nucleotides of a first base which are labeled with a label configured to be detected at a first wavelength,

ii) nucleotides of a second base which are labeled with a label configured to be detected at a second wavelength which is different from the first wavelength, and

iii) nucleotides of a third base comprising nucleotides labeled with a label configured to be detected at the first wavelength and nucleotides labeled with a label configured to be detected at the second wavelength;

b) allowing binding and optional incorporation of the nucleotides of the first pool templated on the multiple copies of the template sequence at each cluster, wherein the bound and optionally incorporated nucleotides are complementary to nucleotides at a first nucleotide position in the template sequences;

c) imaging the substrate to detect signals or absence thereof at the clusters, wherein the signals are associated with the bound and optionally incorporated nucleotides of the first pool, and wherein the imaging comprises:

i) imaging the substrate at the first wavelength and not at the second wavelength to acquire a first image,

ii) imaging the substrate at the second wavelength and not at the first wavelength to acquire a second image, and

iii) imaging the substrate at the first and second wavelengths simultaneously to acquire a third image, wherein in the third image, the signal intensity associated with the third base is no less than the signal intensity associated with the first base or the signal intensity associated with the second base;

wherein for each cluster, a signal codeword comprising signal codes corresponding to the signals or absence thereof detected in c)i) through c)iii) is generated, wherein different signal codewords correspond to different bases, thereby determining the bases at the first nucleotide position in the plurality of polynucleotide templates.

32. The method of claim 31, wherein the first pool of nucleotides comprises nucleotides of a fourth base which are unlabeled and/or are undetected in the imaging.

33. The method of claim 31, wherein the first pool of nucleotides does not comprise nucleotides of a fourth base.

34. The method of any one of claims 31-33, wherein the nucleotides of the third base comprise a nucleotide labeled with both the label configured to be detected at the first wavelength and the label configured to be detected at the second wavelength.

35. The method of any one of claims 31-34, wherein the nucleotides of the third base comprise a first nucleotide labeled with only the label configured to be detected at the first wavelength and a second nucleotide labeled with only the label configured to be detected at the second wavelength.

36. The method of any one of claims 31-35, wherein in each of the nucleotides of the third base, the ratio of the labels configured to be detected at the first or second wavelength and the nucleotide is independently 1:1, 2:1, 3:1, 4:1, 5:1, or greater.

37. The method of claim 36, wherein the labels are conjugated to the nucleotide via a binding pair, optionally wherein the binding pair comprises biotin, DNP, DIG, or desthiobiotin, and a corresponding binding partner thereof.

38. The method of any one of claims 31-37, wherein in c)i), the substrate is illuminated at the first wavelength at a first full power to acquire the first image.

39. The method of any one of claims 31-38, wherein in c)ii), the substrate is illuminated at the second wavelength at a second full power to acquire the second image.
40. The method of any one of claims 31-39, wherein in c)iii), the substrate is illuminated: at the first wavelength at less than the first full power, and at the second wavelength at less than the second full power, thereby acquiring the third image.
41. The method of claim 40, wherein in c)iii), the substrate is illuminated at the first wavelength at about 50% of the first full power, and at the second wavelength at about 50% of the second full power to acquire the third image.
42. The method of any one of claims 39-41, wherein the first and second full powers are of the same constant current in LED illumination.
43. The method of any one of claims 39-41, wherein the first and second full powers are of different constant currents in LED illumination.
44. The method of any one of claims 31-43, wherein in the third image, the signal intensity associated with the third base is about twice the signal intensity associated with the first base or the signal intensity associated with the second base.
45. The method of any one of claims 31-44, wherein the signal codeword comprises signal codes corresponding to signal intensities associated with the nucleotides of the first base, the nucleotides of the second base, or the nucleotides of the third base.
46. The method of any one of claims 31-44, wherein the signal codeword does not comprise signal codes corresponding to signal intensities associated with the nucleotides of the first base, the nucleotides of the second base, or the nucleotides of the third base.
47. The method of any one of claims 31-46, comprising:
- d) contacting the substrate with a second pool of nucleotides comprising unlabeled nucleotides of the first base, unlabeled nucleotides of the second base, unlabeled nucleotides of the third base, and/or unlabeled nucleotides of a fourth base; and
  - e) allowing binding and optional incorporation of the nucleotides of the second pool templated on the multiple copies of the template sequence at each cluster, wherein the bound and optionally incorporated nucleotides are complementary to nucleotides at the first nucleotide position in the template sequences.

48. The method of any one of claims 31-46, comprising:
- d) contacting the substrate with a second pool of nucleotides comprising labeled nucleotides of a fourth base;
  - e) allowing binding and optional incorporation of the nucleotides of the second pool templated on the multiple copies of the template sequence at each cluster, wherein the bound and optionally incorporated nucleotides are complementary to nucleotides at the first nucleotide position in the template sequences; and
  - f) imaging the substrate to detect signals or absence thereof at the clusters, wherein the signals are associated with the bound and optionally incorporated nucleotides of the second pool.
49. The method of any one of claims 31-48, wherein the clusters are disposed at spatially discrete sites on the substrate.
50. The method of any one of claims 31-49, wherein the average distance between adjacent clusters on the substrate is between about 0.3  $\mu\text{m}$  and about 10  $\mu\text{m}$ .
51. The method of any one of claims 31-50, wherein the density of clusters on the substrate is between about 100  $\text{k}/\text{mm}^2$  and about 5000  $\text{k}/\text{mm}^2$ .
52. The method of any one of claims 31-51, wherein signals from adjacent clusters on the substrate are optically resolvable.
53. The method of any one of claims 31-52, wherein the clusters comprise a random array of clusters on the substrate.
54. The method of any one of claims 31-53, wherein the clusters comprise an ordered array of clusters on the substrate.
55. The method of any one of claims 31-54, wherein one or more of the clusters are formed via bridge amplification.
56. The method of claim 55, wherein the one or more clusters each comprises multiple molecules each comprising: i) one or more adapter sequences and/or one or more primer binding sequences and ii) the same template sequence or complement thereof.
57. The method of any one of claims 31-54, wherein one or more of the clusters are formed via rolling circle amplification (RCA).

58. The method of claim 57, wherein the one or more clusters each comprises one or more RCA products (RCPs) each comprising: i) one or more adapter sequences and/or one or more primer binding sequences and ii) the same template sequence.
59. The method of any one of claims 31-58, wherein each cluster comprises:
- i) one or more polynucleotides that are 5' immobilized on the substrate and 3' blocked;
  - ii) one or more polynucleotides that are 3' immobilized on the substrate; and/or
  - iii) one or more nucleic acid concatemers immobilized on the substrate.
60. The method of any one of claims 31-59, wherein the first pool of nucleotides comprises nucleotides of any three of A, T/U, C, and G.
61. The method of any one of claims 31-60, wherein the first pool of nucleotides comprises nucleotides of A, T/U, and C but no nucleotides of G.
62. The method of any one of claims 31-61, wherein the first and/or second pool of nucleotides comprise one or more dNTP monomers and/or one or more dNTP multimers each comprising multiple dNTP molecules conjugated to a scaffold, optionally wherein the scaffold is conjugated to one or more fluorescent moieties and optionally wherein the scaffold comprises a streptavidin.
63. The method of any one of claims 31-62, wherein the first and/or second pool of nucleotides each comprises a reversible terminator, optionally wherein the reversible terminator is a 3'-O-blocked reversible terminator or a 3'-unblocked reversible terminator, and the method comprises allowing nucleotide incorporation after imaging the substrate.
64. The method of any one of claims 31-63, comprising removing unbound or unincorporated nucleotides from the substrate prior to imaging the substrate in c).
65. The method of any one of claims 31-64, comprising determining the bases at a second nucleotide position in the plurality of polynucleotide templates, wherein the second nucleotide position is 5' to the first nucleotide position in the plurality of polynucleotide templates.
66. A kit, comprising:

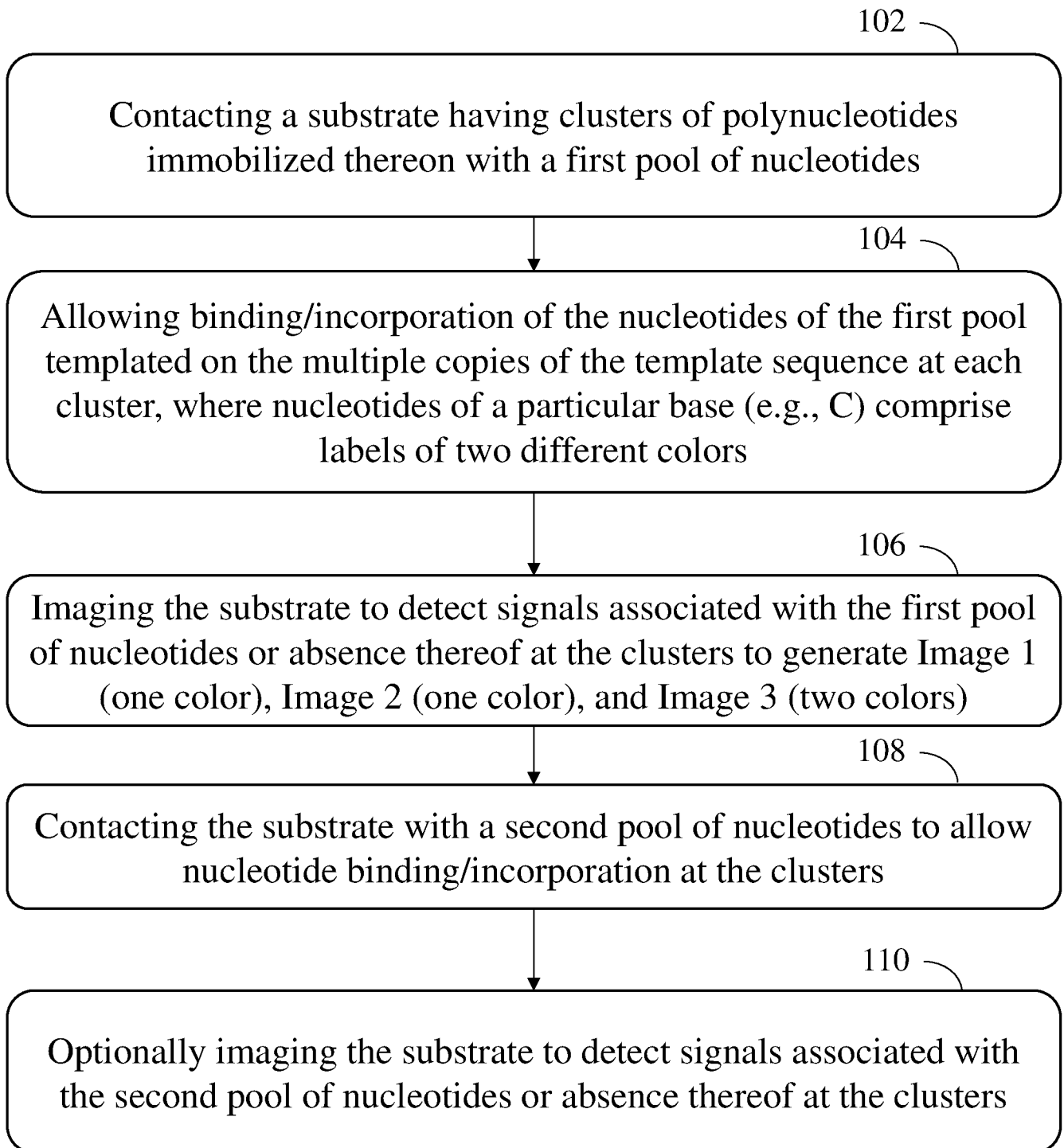
- i) a first plurality of nucleotide molecules of a first base, comprising nucleotides labeled or configured to be labeled with a first label detectable at a first wavelength,
  - ii) a second plurality of nucleotide molecules of a second base which is different from the first base, comprising nucleotides labeled or configured to be labeled with a second label detectable at a second wavelength which is different from the first wavelength, and
  - iii) a third plurality of nucleotide molecules of a third base which is different from the first and second bases, comprising a nucleotide labeled or configured to be labeled with a third label detectable at the first wavelength and a nucleotide labeled or configured to be labeled with a fourth label detectable at the second wavelength,
- wherein the first plurality, the second plurality, and/or the third plurality each independently comprises between about 5% and about 50% nucleotide molecules that are not detectably labeled; and
- iv) optionally instructions for using the first plurality, the second plurality, and/or the third plurality of nucleotide molecules according to the method in any one of claims 1-65.

67. The kit of claim 66, wherein the first plurality, the second plurality, and/or the third plurality each independently comprises about 10% nucleotide molecules that are not detectably labeled.

68. The kit of claim 66 or claim 67, comprising: iv) a fourth plurality of nucleotide molecules of a fourth base which is different from the first, second, and third bases, optionally wherein the fourth plurality of nucleotide molecules are not detectably labeled.

69. The kit of any one of claims 66-68, wherein the first, second, third, and/or fourth plurality each independently comprises one or more reversibly terminated nucleotide molecules, and/or one or more complexes comprising nucleotide molecules conjugated to a scaffold such that the nucleotide molecules are not incorporable and the one or more complexes are labeled with detectable labels.



100**FIG. 1**

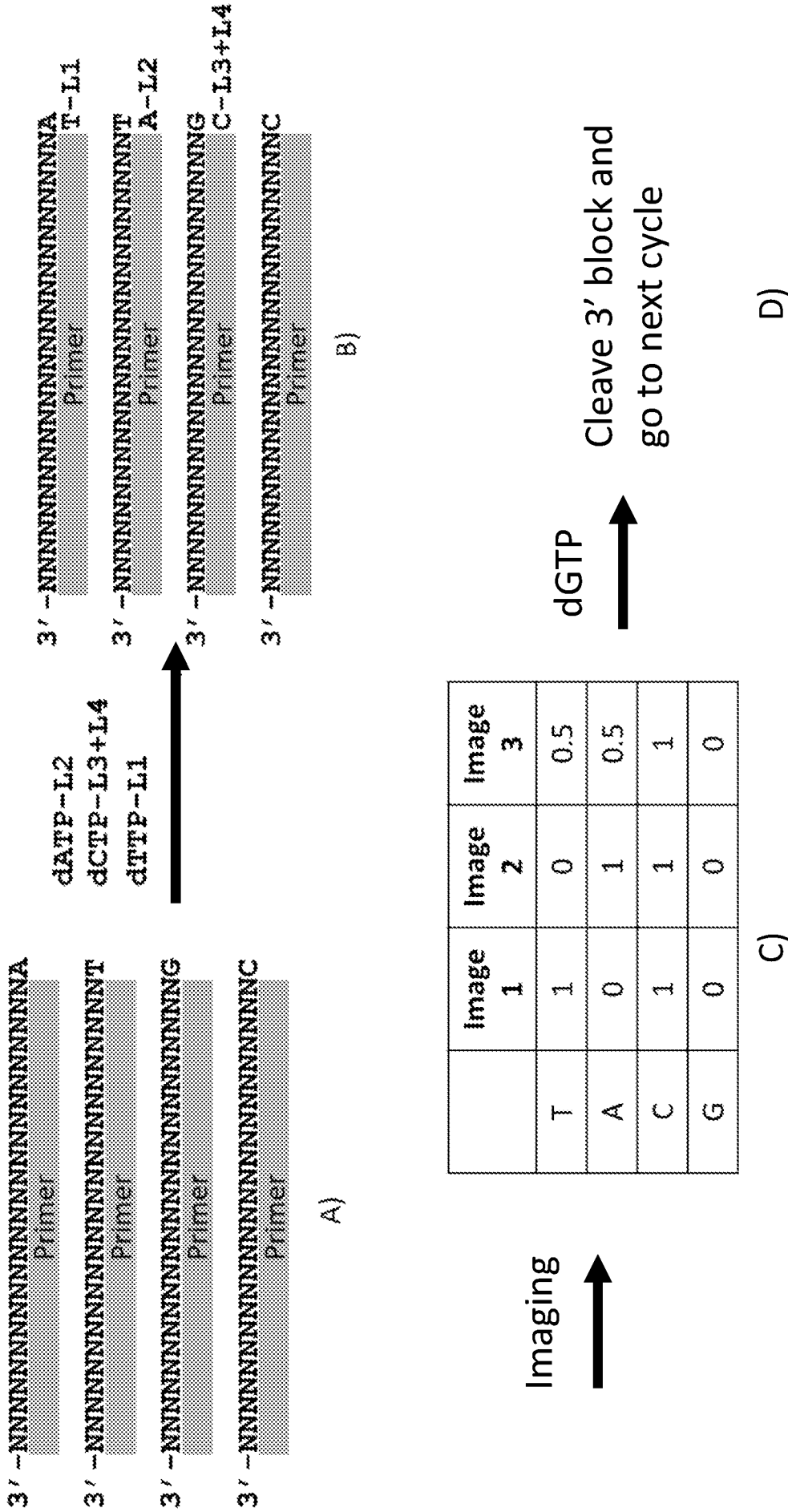
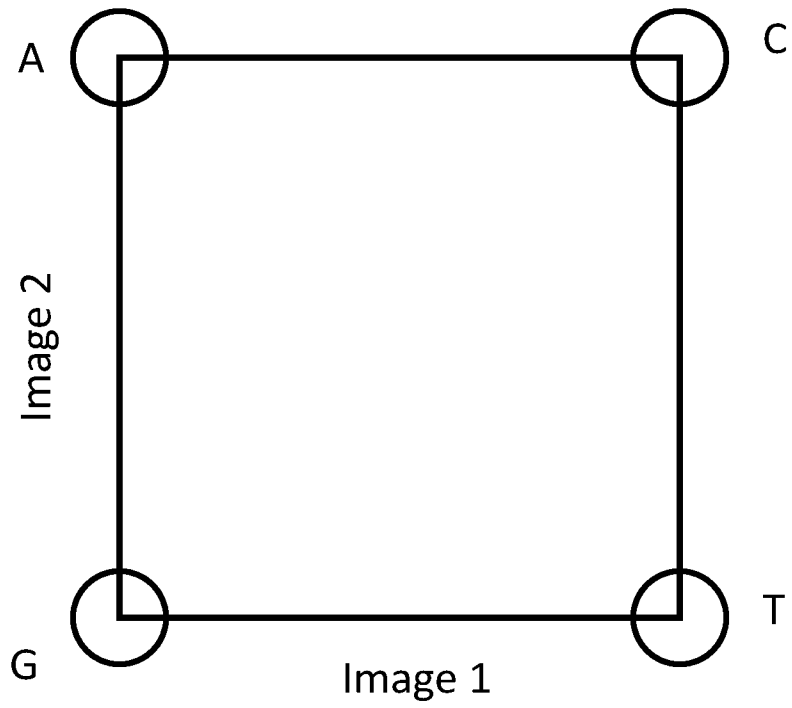
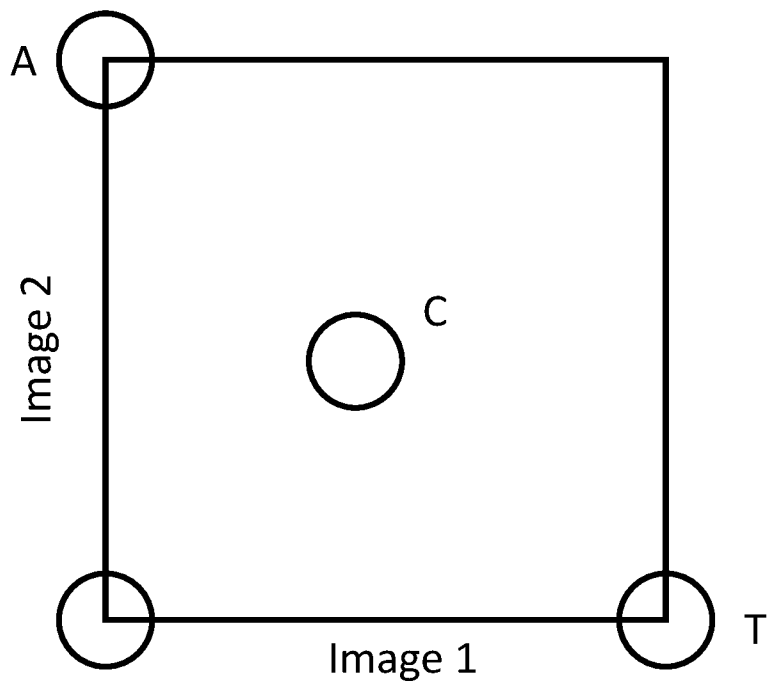


FIG. 2



**FIG. 3**



**FIG. 4**

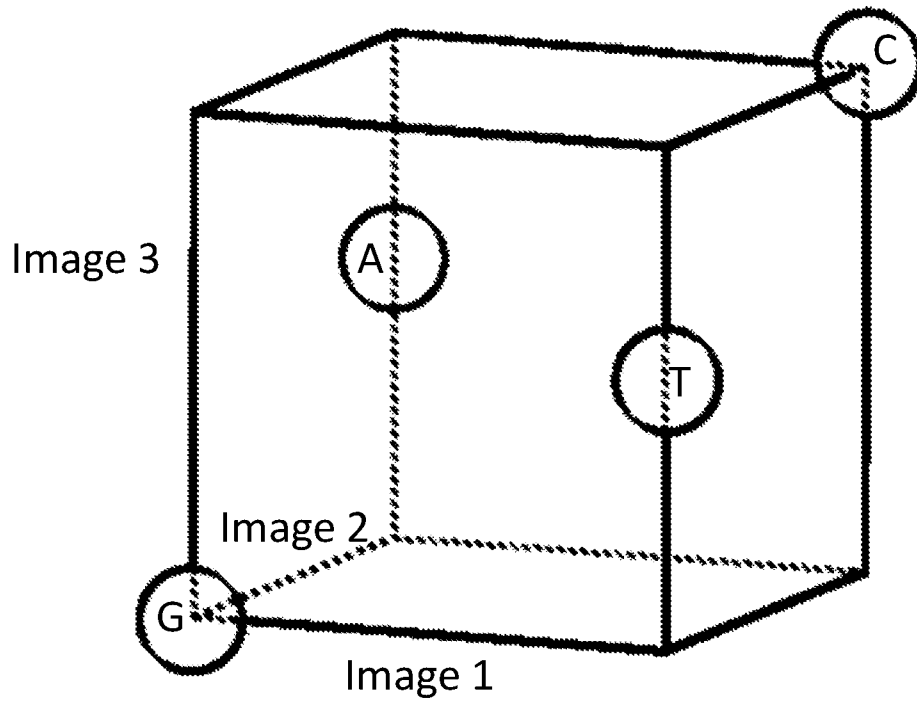


FIG. 5

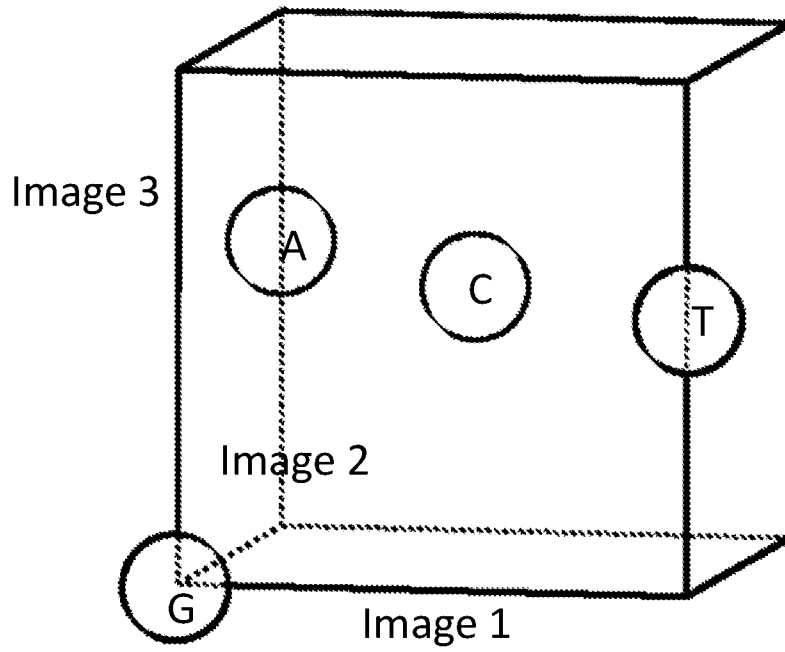
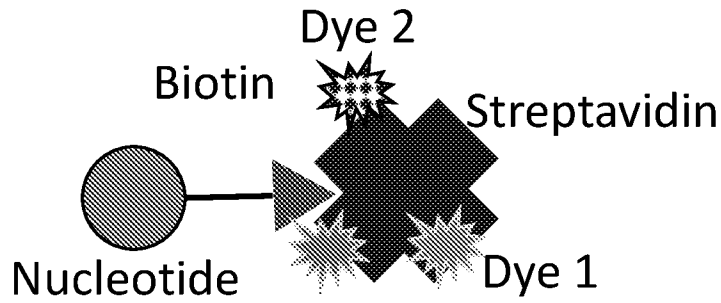
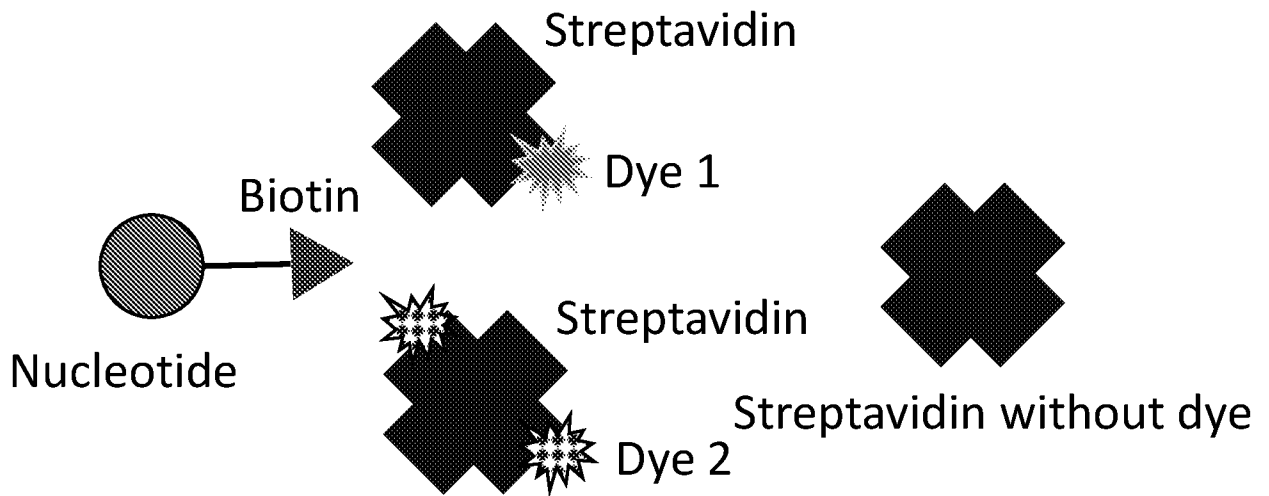


FIG. 6



**FIG. 7**



**FIG. 8**

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2024/024437

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. C12Q1/6869 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) <b>C12Q</b>		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) <b>EPO-Internal, BIOSIS, EMBASE, WPI Data</b>		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2013/044018 A1 (ILLUMINA INC [US]; KAIN ROBERT C [US] ET AL.) 28 March 2013 (2013-03-28) the whole document -----	1 - 69
X	WO 2018/165099 A1 (ILLUMINA INC [US]; ILLUMINA CAMBRIDGE LTD [GB]) 13 September 2018 (2018-09-13) the whole document -----	1 - 69
X	WO 2022/136402 A1 (ILLUMINA CAMBRIDGE LTD [GB]) 30 June 2022 (2022-06-30) the whole document -----	1 - 69
X	WO 2017/087823 A1 (MIR KALIM U [US]) 26 May 2017 (2017-05-26) the whole document ----- - / - -	1 - 69
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <span style="margin-left: 200px;"><input checked="" type="checkbox"/> See patent family annex.</span>		
* Special categories of cited documents :		
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search	Date of mailing of the international search report	
<b>19 July 2024</b>	<b>30/07/2024</b>	
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  <b>Bradbrook, Derek</b>	

# INTERNATIONAL SEARCH REPORT

International application No PCT/US2024/024437
---

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2022/129439 A1 (ILLUMINA CAMBRIDGE LTD [GB]) 23 June 2022 (2022-06-23) the whole document -----	1 - 69
X	WO 2020/193765 A1 (ILLUMINA CAMBRIDGE LTD [GB]) 1 October 2020 (2020-10-01) the whole document -----	1 - 69

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2024/024437

Patent document cited in search report	Publication date	Patent family member(s)	Publication date			
WO 2013044018	A1	28-03-2013	CA 2859660 A1 28-03-2013			
			CA 3104322 A1 28-03-2013			
			CY 1125261 T1 24-03-2023			
			DK 2768972 T3 25-09-2017			
			DK 3290528 T3 04-11-2019			
			DK 3623481 T3 15-11-2021			
			EP 2768972 A1 27-08-2014			
			EP 3290528 A1 07-03-2018			
			EP 3623481 A1 18-03-2020			
			EP 3981886 A1 13-04-2022			
			ES 2639938 T3 30-10-2017			
			ES 2752653 T3 06-04-2020			
			ES 2895184 T3 17-02-2022			
			HK 1201079 A1 21-08-2015			
			HK 1247254 A1 21-09-2018			
			HR P20211523 T1 24-12-2021			
			HU E056246 T2 28-02-2022			
			LT 3623481 T 27-09-2021			
			PL 3290528 T3 31-03-2020			
			PL 3623481 T3 17-01-2022			
			PT 3290528 T 14-10-2019			
			PT 3623481 T 15-10-2021			
			SI 3623481 T1 31-01-2022			
			US 2013079232 A1 28-03-2013			
			US 2017022559 A1 26-01-2017			
			US 2019264279 A1 29-08-2019			
			US 2021139978 A1 13-05-2021			
			US 2024141428 A1 02-05-2024			
			WO 2013044018 A1 28-03-2013			
			-----			
			WO 2018165099	A1	13-09-2018	AU 2018231017 A1 20-06-2019
						AU 2021269291 A1 09-12-2021
BR 112019014683 A2 18-02-2020						
CA 3046015 A1 13-09-2018						
CN 110234771 A 13-09-2019						
EP 3592865 A1 15-01-2020						
IL 267051 A 29-08-2019						
JP 7041684 B2 24-03-2022						
JP 2020507758 A 12-03-2020						
JP 2022088444 A 14-06-2022						
JP 2024028778 A 05-03-2024						
KR 20190103267 A 04-09-2019						
KR 20210047980 A 30-04-2021						
NZ 754255 A 28-05-2021						
RU 2736384 C1 16-11-2020						
US 2020080142 A1 12-03-2020						
WO 2018165099 A1 13-09-2018						
ZA 202106656 B 27-07-2022						
-----						
WO 2022136402	A1	30-06-2022				AU 2021406375 A1 05-01-2023
			CA 3182311 A1 30-06-2022			
			CN 115803458 A 14-03-2023			
			EP 4267765 A1 01-11-2023			
			JP 2023552936 A 20-12-2023			
			KR 20230123871 A 24-08-2023			
			US 2022195518 A1 23-06-2022			
			WO 2022136402 A1 30-06-2022			
-----						



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No PCT/US2024/024437
---

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2017087823 A1	26-05-2017	CN 108603227 A	28-09-2018
		EP 3411494 A1	12-12-2018
		US 2018327829 A1	15-11-2018
		US 2022002799 A1	06-01-2022
		WO 2017087823 A1	26-05-2017
-----			
WO 2022129439 A1	23-06-2022	AU 2021399113 A1	05-01-2023
		BR 112022025475 A2	07-03-2023
		CA 3183764 A1	23-06-2022
		CN 115803457 A	14-03-2023
		EP 4263865 A1	25-10-2023
		IL 299523 A	01-02-2023
		JP 2023552935 A	20-12-2023
		KR 20230120985 A	17-08-2023
		US 2022195516 A1	23-06-2022
		WO 2022129439 A1	23-06-2022
		-----	
WO 2020193765 A1	01-10-2020	AU 2020247472 A1	07-01-2021
		CA 3103909 A1	01-10-2020
		CN 112639128 A	09-04-2021
		EP 3947729 A1	09-02-2022
		JP 2022525821 A	20-05-2022
		KR 20210144550 A	30-11-2021
		SG 11202012806U A	28-01-2021
		US 2020308639 A1	01-10-2020
		US 2023055284 A1	23-02-2023
		WO 2020193765 A1	01-10-2020
-----			