



- (51) International Patent Classification:  
C12Q 1/6806 (2018.01) C12Q 1/6869 (2018.01)
- (21) International Application Number:  
PCT/US2021/060328
- (22) International Filing Date:  
22 November 2021 (22.11.2021)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
63/116,552 20 November 2020 (20.11.2020) US
- (71) Applicants: **CAMENA BIOSCIENCE LIMITED** [GB/GB]; The Science Village Chesterford Research, Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB). **ELRIFI, Ivor** [US/US]; Cooley LLP, 1299 Pennsylvania Avenue, Suite 700, Washington, District of Columbia 20004 (US).

- (72) Inventor; and
- (71) Applicant: **BELL, Neil** [GB/GB]; The Science Village Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL (GB).
- (74) Agent: **ELRIFI, Ivor** et al.; Cooley LLP, 1299 Pennsylvania Avenue, Suite 700, Washington, District of Columbia 20004 (US).
- (81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.
- (84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH,

(54) Title: GEOMETRIC SYNTHESIS METHODS AND COMPOSITIONS FOR DOUBLE-STRANDED NUCLEIC ACID SEQUENCING

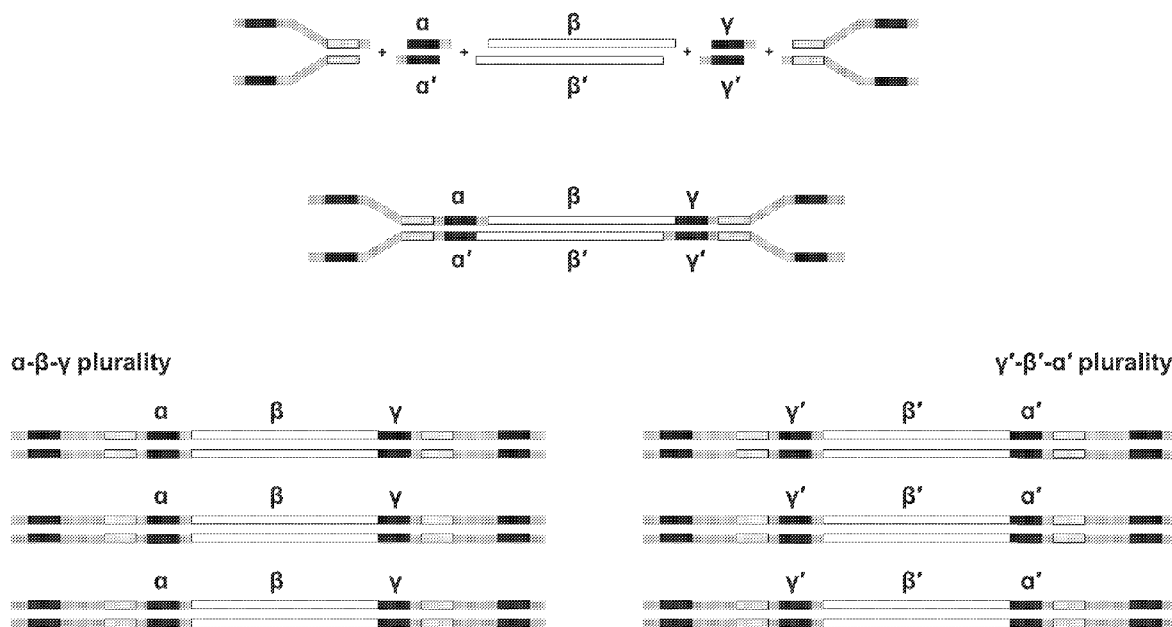


FIG. 1

(57) Abstract: The present disclosure provides compositions, kits and methods for sequencing double-stranded nucleic acids. The compositions, kits and methods comprise partially double-stranded identifier molecules and partially double-stranded adapter molecules. The compositions, kits and methods can be used to determine the abundance and/or identity of specific transcripts in a plurality of double-stranded nucleic acids, as well as identifying the frequency of mutations within certain transcripts in the plurality of double-stranded nucleic acids.



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

**Declarations under Rule 4.17:**

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

**Published:**

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
- *with sequence listing part of description (Rule 5.2(a))*

## **GEOMETRIC SYNTHESIS METHODS AND COMPOSITIONS FOR DOUBLE-STRANDED NUCLEIC ACID SEQUENCING**

### **CROSS-REFERENCE TO RELATED APPLICATIONS**

**[0001]** This application claims priority to, and the benefit of, U.S. Provisional Application No. 63/116,552, filed November 20, 2020, the contents of which are incorporated herein by reference in their entireties for all purposes.

### **SEQUENCE LISTING**

**[0002]** The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on November 21, 2021, is named “DNWR-008\_SeqList.txt” and is about 46,527 bytes in size.

### **BACKGROUND**

**[0003]** There is a need in the art for sensitive and specific methods and materials to detect mutations (variations) in circulating tumor DNA or for the identification of variants in complex mixed metagenomic samples as well as many other applications. Next generation sequencing (NGS) platforms provide enormous depth of sequence data across many genes simultaneously. Often, however, due to amplification artifact combined with intrinsic errors, the fidelity of variant calling is compromised. The present disclosure provides methods, compositions and kits for the sensitive and specific identification of variant alleles in complex DNA samples. The present disclosure provides methods, compositions and kits that can be used to identify variants at the sensitivity appropriate to the expected degree of variation by controlling the number of potential bar codes and the depth of sequence read coverage.

### **BACKGROUND**

**[0004]** The present disclosure provides pluralities of partially double-stranded identifier molecules, wherein the partially double-stranded identifier molecules comprise: a double-stranded region comprising an identifier sequence; and a first overhang; wherein the plurality comprises at least about 12 species of the partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter

molecules in the plurality, wherein the identifier sequence of one species of partially double-stranded identifier molecules will have a hamming distance of at least about two to any other identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0005]** In some aspects, the partially double-stranded identifier molecules further comprise a second overhang.

**[0006]** In some aspects the first and second overhangs are a) 5' overhangs; or b) 3' overhangs.

**[0007]** In some aspects, the identifier sequence spans the entire double-stranded region. In some aspects, the identifier sequence spans a portion of the double-stranded region.

**[0008]** In some aspects, the identifier sequence is: a) about 9 nucleotides in length; b) about 10 nucleotides in length; c) about 11 nucleotides in length; d) about 12 nucleotides in length; e) about 19 nucleotides in length; f) about 20 nucleotides in length; g) about 21 nucleotides in length; or h) about 22 nucleotides in length.

**[0009]** In some aspects, the first overhang and/or the second overhang is about 1 nucleotide in length. In some aspects, the first overhang and/or the second overhang is about 1 nucleotide in length, and the first overhang and/or the second overhang is: a) an adenine or a thymine; or b) a guanosine or a cytosine.

**[0010]** In some aspects, the first overhang and/or the second overhang is: a) about 2 nucleotides in length; b) about 3 nucleotides in length; c) about 4 nucleotides in length; or d) about 5 nucleotides in length.

**[0011]** In some aspects, the partially double-stranded identifier molecules comprise DNA.

**[0012]** In some aspects, a plurality comprises: a) at least about 24 species of the partially double-stranded identifier molecules; b) at least about 48 species of the partially double-stranded identifier molecules; or c) at least about 96 species of the partially double-stranded identifier molecules.

**[0013]** The present disclosure provides pluralities of partially double-stranded adapter molecules, wherein the partially double-stranded adapter molecules comprise: a double-stranded region; an overhang; a single-stranded 5' arm; and a single-stranded 3' arm; wherein the single-stranded 5' arm comprises at least one amplification primer binding site and the single-stranded 3' arm comprises at least one amplification primer binding site. In some aspects, the double-stranded region comprises an identifier sequence. In some aspects, a plurality comprises at least about 12

species of the partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter molecules in the plurality.

**[0014]** In some aspects, the overhang is: a) a 5' overhang; or b) a 3' overhang.

**[0015]** In some aspects, an overhang is about 1 nucleotide in length. In some aspects, an overhang is about 1 nucleotide in length, and wherein the overhang is: a) an adenine or a thymine; or b) a guanosine or cytosine.

**[0016]** In some aspects, an overhang is: a) about 2 nucleotides in length; b) about 3 nucleotides in length; c) about 4 nucleotides in length; d) about 5 nucleotides in length.

**[0017]** In some aspects, an identifier sequence is: a) about 9 nucleotides in length; b) about 10 nucleotides in length; c) about 11 nucleotides in length; d) about 12 nucleotides in length; e) about 19 nucleotides in length; f) about 20 nucleotides in length; g) about 21 nucleotides in length; or h) about 22 nucleotides in length.

**[0018]** In some aspects, partially double-stranded adapter molecules comprise DNA.

**[0019]** The present disclosure provides methods of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with the plurality of partially double-stranded identifier molecules of any one of claims 1-9 and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids; b) contacting the products of step (a) with the plurality of partially double-stranded adapter molecules of any one of claims 10-16 and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[0020]** In some aspects, the ligation products in step (a) comprise: a) at least 10% of the combinations of two species of partially double-stranded identifier molecules; b) at least 20% of the combinations of two species of partially double-stranded identifier molecules; c) at least 30% of the combinations of two species of partially double-stranded identifier molecules; d) at least 40% of the combinations of two species of partially double-stranded identifier molecules; e) at least 50% of the combinations of two species of partially double-stranded identifier molecules; f) at least 60% of the combinations of two species of partially double-stranded identifier molecules;

g) at least 70% of the combinations of two species of partially double-stranded identifier molecules; h) at least 80% of the combinations of two species of partially double-stranded identifier molecules; i) at least 90% of the combinations of two species of partially double-stranded identifier molecules; or j) each of the combinations of two species of partially double-stranded identifier molecules.

**[0021]** In some aspects, the methods further comprise after step (b) and prior to step (c), constructing a sequencing library using the products of step (b).

**[0022]** In some aspects, step (a) and step (b) are performed sequentially or are performed concurrently.

**[0023]** In some aspects, the methods further comprise after step (b) and prior to step (c), amplifying the products of step (b). In some aspects, amplifying the products of step (b) comprises contacting the products of step (b) with amplification primers that bind to amplification primer binding sites in the partially double-stranded adapter molecules and at least one polymerase.

**[0024]** In some aspects, the methods further comprise determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c). In some aspects, determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises correcting for errors using the identifier sequences of the ligated partially double-stranded identifier molecules. In some aspects, the errors comprise amplification errors, sample preparation errors, sequencing errors or any combination thereof; In some aspects, determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises creating consensus sequences using identifier sequences of the ligated partially double-stranded identifier molecules.

**[0025]** In some aspects, determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises grouping the sequencing reads obtained in step (c) by the ligated identifier sequences in the sequencing reads, grouping the sequencing reads obtained in step (c) by the specific genomic sequence that the sequencing reads most likely correspond to, or any combination thereof.

**[0026]** In some aspects, determining the abundance and/or identify of specific transcripts in the plurality of double-stranded target nucleic acids can comprise determining the frequency of one or more mutations in a specific transcript in the plurality of double-stranded target nucleic acid, In some aspects, the one or more mutations comprise one or more insertions, one or more deletion-insertions, one or more duplications, one or more inversions, one or more repeat expansions or any combination thereof.

**[0027]** The present disclosure provides kits comprising at least one plurality of partially double-stranded identifier molecules of the present disclosure. In some aspects, the kits can further comprise at least one plurality of partially double-stranded adapter molecules of the present disclosure.

**[0028]** Any of the above aspects, or any other aspect described herein, can be combined with any other aspect.

**[0029]** Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. In the Specification, the singular forms also include the plural unless the context clearly dictates otherwise; as examples, the terms “a,” “an,” and “the” are understood to be singular or plural and the term “or” is understood to be inclusive. By way of example, “an element” means one or more element. Throughout the specification the word “comprising,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from the context, all numerical values provided herein are modified by the term “about.”

**[0030]** Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. The references cited herein are not admitted to be prior art to the claimed invention. In the case of conflict, the present Specification, including definitions, will control. In addition, the materials, methods, and

examples are illustrative only and are not intended to be limiting. Other features and advantages of the disclosure will be apparent from the following detailed description and claim.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

**[0031]** The above and further features will be more clearly appreciated from the following detailed description when taken in conjunction with the accompanying drawings.

**[0032]** FIG. 1 is a schematic overview of the methods and compositions of the present disclosure.

**[0033]** FIG. 2 is a schematic overview of the methods and compositions of the present disclosure.

**[0034]** FIG. 3 is a schematic overview of partially double-stranded adapter molecules of the present disclosure.

**[0035]** FIG. 4 is a schematic of an exemplary sequencing data analysis workflow of the present disclosure.

**[0036]** FIG. 5 shows the results of an experiment using the methods and compositions of the present disclosure, specifically identifier molecules and adapter molecules comprising multiple base overhangs. The nucleic acid sequences shown in this figure correspond to SEQ ID NOs: 197-208.

**[0037]** FIG. 6 shows the results of an experiment using the methods and compositions of the present disclosure, specifically identifier molecules and adapter molecules comprising single base overhangs with varying sizes of double-stranded regions. The nucleic acid sequences shown in this figure correspond to SEQ ID NOs: 211-229.

**[0038]** FIG. 7 is a schematic comparison between existing next generation sequencing barcode compositions and methods that rely on the use of pre-pooled, degenerate barcodes and the compositions and the methods of the present disclosure.

**[0039]** FIG. 8 shows heatmaps generated for the coverage of each UMI created using the sequencing compositions and methods prior (left) and post (middle) error correction; the difference for the coverage between the UMIs prior and post error-correction is also shown (right), showing regions where UMI coverage decreased and increased. CorrectUmis (fgbio tools) was used for the UMI error-correction.

**[0040]** FIG. 9 shows an example Bioanalyzer trace from sequencing libraries assembled using amplicons prepared from gDNA (Quantitative Multiplex Reference Standard, Horizon Discovery)



using AmpliSeq primers and partially double stranded identifier molecules and adapter molecules of the present disclosure.

**[0041]** FIG. 10 shows the sequencing results for the EGFR4 gene and the measured mutant frequencies for a DNA base change of GGC→AGC obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 226.

**[0042]** FIG. 11 shows the sequencing results for the PI3KCA10 gene and the measured mutant frequencies for a DNA base change of CAT→CGT obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 227.

**[0043]** FIG. 12 shows the sequencing results for the KRAS1 gene and the measured mutant frequencies for a DNA base change of GGC→GAC obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 228.

**[0044]** FIG. 13 shows the sequencing results for the NRAS gene and the measured mutant frequencies for a DNA base change of CAA→AAA obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 229.

**[0045]** FIG. 14 shows the sequencing results for the BRAF gene and the measured mutant frequencies for a DNA base change of CTG→CAG obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 230.

**[0046]** FIG. 15 shows the sequencing results for the KIT gene and the measured mutant frequencies for a DNA base change of GAC→GTC obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 231.

**[0047]** FIG. 16 shows the sequencing results for the PI3KCA7 gene and the measured mutant frequencies for a DNA base change of GAG→AAG obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 232.

[0048] FIG. 17 shows the sequencing results for the KRAS1 gene and the measured mutant frequencies for a DNA base change of GGT→GAT obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 233.

[0049] FIG. 18 shows the sequencing results for the EGFR8 gene and the measured mutant frequencies for a DNA base change of CTG→CGG obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 234.

[0050] FIG. 19 shows the sequencing results for the EGFR5 gene and the measured mutant frequencies for a DNA base change of AAGGAATTAAGAGAAGCA→AA obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 235.

[0051] FIG. 20 shows the sequencing results for the EGFR6 gene and the measured mutant frequencies for a DNA base change of ACG→ATG obtained using existing NGS methods and the sequencing methods of the present disclosure. The nucleic acid sequence shown in this figure corresponds to SEQ ID NO: 236.

### DETAILED DESCRIPTION

[0052] Existing methods for generating Next Generation Sequencing (NGS) libraries provide unique molecular identifiers (UMI) to individual fragments of DNA. When the DNA is amplified, the sequence of the original molecule may be unambiguously identified, because each of the multiple copies will have the same UMI as the original DNA fragment. Not only does this allow normalization after amplification for quantitative use, but also reduces sequencing error by providing a consensus.

[0053] The present disclosure provides improved double-stranded nucleic acid sequencing methods by adding an adjustable number of available barcodes. In this approach, modular adapter and identifier molecules are simultaneously ligated to complex mixtures of individual target DNA fragments to generate an NGS library. Individual identifier molecules are added to DNA fragments through single base overhangs (*e.g.* A/T). Simultaneously, partially double-stranded Y-shaped adapter molecules are ligated to the ends of identifier molecules already attached to the target DNA molecules using an overhanging sequence, which is reverse complementary on the identifier and adapter molecules.

**[0054]** In some aspects, the identifier molecules are a small subset of all possible 11- or 20-mer base pair identifier sequences and are selected to be unambiguous when sequenced. In a non-limiting example, an identifier set may comprise only 12 individual sequences, but because they will be randomly affixed to either end of a target DNA molecule, there will be 144 barcode possibilities ( $12 \times 12 = 144$ ). Without wishing to be bound by theory, because of the randomness of the target DNA fragment length and genomic location, and the unique encoding of each strand from the duplex, the resulting barcodes allow the unique identification of the original DNA molecules. For rare events, such as mutated bases in a circulating free DNA that may be associated with cancer, the number of barcodes employed can be adjusted, along with the depth of sequencing, to provide the appropriate sensitivity for the specific application. Without wishing to be bound by theory, higher sensitivity in deep sequencing will require a larger number of possible barcodes. For example, one may want to use a set of identifier molecules that has 96 different identifier sequences allowing for a total of 9216 ( $96 \times 96 = 9216$ ) distinct barcodes.

**[0055]** In some aspects, individual libraries are uniquely identified from a mix of libraries by an index identifier that is added during amplification carried by amplification primers. Additionally, platform specific adapter molecules are incorporated, allowing the user to employ any existing NGS systems, including, but not limited to, Illumina, Oxford Nanopore or Pacific Bioscience.

**[0056]** *Partially double-stranded identifier molecules*

**[0057]** The present disclosure provides partially double-stranded identifier molecules. Partially double-stranded identifier molecules are nucleic acid molecules comprising at least one double-stranded region and at least one single stranded region. In some aspects, a partially double-stranded identifier molecule is a nucleic acid molecule comprising one double-stranded region and one single-stranded region. In some aspects, a partially double-stranded identifier molecule is a nucleic acid molecule comprising one double-stranded region and two single-stranded regions.

**[0058]** In some aspects a partially double-stranded identifier molecule comprises DNA. In some aspects, a partially double-stranded identifier molecule comprises RNA. In some aspects, a partially double-stranded identifier molecule can comprise XNA. In some aspects, a partially double-stranded identifier molecule comprises any combination of DNA, RNA and XNA.

**[0059]** As used herein, the term “XNA” is used to refer to xeno nucleic acids. As would be appreciated by the skilled artisan, xeno nucleic acids are synthetic nucleic acid analogues

comprising a different sugar backbone than the natural nucleic acids DNA and RNA. XNAs can include, but are not limited to, 1,5-anhydrohexitol nucleic acid (HNA), Cyclohexene nucleic acid (CeNA), Threose nucleic acid (TNA), Glycol nucleic acid (GNA), Locked nucleic acid (LNA), Peptide nucleic acid (PNA) and FANA (Fluoro Arabino nucleic acid).

**[0060]** In some aspects, a partially double-stranded identifier molecule can comprise an identifier sequence, also referred to herein as an identifier nucleic acid sequence, a barcode sequence or a hemi-barcode sequence. As would be appreciated by the skilled artisan, an identifier sequence is a nucleic acid sequence that can be used as part of a sequencing method to identify individual molecules within a sample. An identifier sequence can comprise a degenerate, a semi-degenerate or discrete (non-degenerate) nucleic acid sequence.

**[0061]** In some aspects, an identifier sequence can be a nucleic acid sequence that is known not to occur or that occurs infrequently in the genome of an organism from which a sample is derived. In a non-limiting example, an identifier sequence can be a nucleic acid sequence that is known not to occur or that occurs infrequently in the human genome.

**[0062]** In some aspects, a partially double-stranded identifier molecule can comprise one overhang. The overhang can be a 3' overhang or a 5' overhang. In some aspects, a partially double-stranded identifier molecule can comprise two overhangs. The overhangs can be 3' overhangs or 5' overhangs.

**[0063]** As would be appreciated by the skilled artisan, an "overhang", in the context of a partially double-stranded nucleic acid molecule refers to a single-stranded region of a partially-double stranded nucleic acid molecule located at a terminus of the partially double-stranded nucleic acid molecule for which there is no single-stranded region located on the opposite strand. FIG. 3 shows 3' and 5' overhangs in exemplary partially double-stranded nucleic acid molecules, namely partially double-stranded adapter molecules of the present disclosure, which are described in further detailed herein.

**[0064]** As used herein, the term 5' overhang is used to refer to a single-stranded region of a partially double-stranded nucleic acid molecule that is located at the 5' terminus of one of the strands.

**[0065]** As used herein, the term 3' overhang is used to refer to a single-stranded region of a partially double-stranded nucleic acid molecule that is located at the 3' terminus of one of the strands.

**[0066]** In some aspects, a partially double-stranded identifier molecule can comprise an identifier sequence and one overhang. In some aspects, the overhang can be a 3' overhang. In some aspects, the overhang can be a 5' overhang.

**[0067]** In some aspects, a partially double-stranded identifier molecule can comprise an identifier sequence and two overhangs. In some aspects, the overhangs can be 3' overhangs. In some aspects, the overhang can be a 5' overhangs.

**[0068]** In some aspects of the compositions of the present disclosure a 5' overhang of a partially double-stranded identifier molecule can be about 1 nucleotide in length, or about 2 nucleotides in length, or about 3 nucleotides in length, or about 4 nucleotides in length, or about 5 nucleotides in length. In some aspects of the compositions of the present disclosure a 5' overhang can be at least about 1 nucleotide, or at least about 2 nucleotides, or at least about 3 nucleotides, or at least about 4 nucleotides, or at least about 5 nucleotides, or at least about 6 nucleotides, or at least about 7 nucleotides, or at least about 8 nucleotides, or at least about 9 nucleotides, or at least about 10 nucleotides in length.

**[0069]** In some aspects, a 5' overhang of a partially double-stranded identifier molecule is no more than 1, or no more than 2, or no more than 3, or no more than 4, or no more than 5, or no more than 6, or no more than 7, or no more than 8, or no more than 9, or no more than 10 nucleotides, or no more than 11 nucleotides, or no more than 12 nucleotides, or no more than 13 nucleotide, or no more than 14 nucleotides, or no more than 15 nucleotides, or no more than 16 nucleotides, or no more than 17 nucleotides, or no more than 18 nucleotides, or no more than 19 nucleotides, or no more than 20 nucleotides in length.

**[0070]** In some aspects of the compositions of the present disclosure a 3' overhang of a partially double-stranded identifier molecule can be about 1 nucleotide in length, or about 2 nucleotides in length, or about 3 nucleotides in length, or about 4 nucleotides in length, or about 5 nucleotides in length. In some aspects of the compositions of the present disclosure a 3' overhang can be at least about 1 nucleotide, or at least about 2 nucleotides, or at least about 3 nucleotides, or at least about 4 nucleotides, or at least about 5 nucleotides, or at least about 6 nucleotides, or at least about 7 nucleotides, or at least about 8 nucleotides, or at least about 9 nucleotides, or at least about 10 nucleotides in length.

**[0071]** In some aspects, a 3' overhang of a partially double-stranded identifier molecule is no more than 1, or no more than 2, or no more than 3, or no more than 4, or no more than 5, or no

more than 6, or no more than 7, or no more than 8, or no more than 9, or no more than 10 nucleotides, or no more than 11 nucleotides, or no more than 12 nucleotides, or no more than 13 nucleotide, or no more than 14 nucleotides, or no more than 15 nucleotides, or no more than 16 nucleotides, or no more than 17 nucleotides, or no more than 18 nucleotides, or no more than 19 nucleotides, or no more than 20 nucleotides in length.

**[0072]** In some aspects, a 3' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length. In some aspects, a 3' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is an adenine or a thymine. In some aspects, a 3' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is an adenine. In some aspects, a 3' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a thymine. In some aspects, a 3' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a guanine or a cytosine. In some aspects, a 3' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a guanine. In some aspects, a 3' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a cytosine.

**[0073]** In some aspects, a 5' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length. In some aspects, a 5' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is an adenine or a thymine. In some aspects, a 5' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is an adenine. In some aspects, a 5' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a thymine. In some aspects, a 5' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a guanine or a thymine. In some aspects, a 5' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a guanine. In some aspects, a 5' overhang of a partially double-stranded identifier molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a cytosine.

**[0074]** In some aspects, the double-stranded region of a partially double-stranded identifier molecule can be at least about 1 nucleotide in length, at least about 2 nucleotides in length, or at least about 3 nucleotides in length, or at least about 4 nucleotides in length, or at least about 5 nucleotides in length, or at least about 6 nucleotides in length, or at least about 7 nucleotides in

length, or at least about 8 nucleotides in length, or at least about 9 nucleotides in length, or at least about 10 nucleotides in length, or at least about 11 nucleotides in length, or at least about 12 nucleotides in length, or at least about 13 nucleotides in length, or at least about 14 nucleotides in length, or at least about 15 nucleotides in length, or at least about 16 nucleotides in length, or at least about 17 nucleotides in length, or at least about 18 nucleotides in length, or at least about 19 nucleotides in length, or at least about 20 nucleotides in length in length, or at least about 21 nucleotides in length, or at least about 22 nucleotides in length, or at least about 23 nucleotides in length, or at least about 24 nucleotides in length, or at least about 25 nucleotides in length, or at least about 26 nucleotides in length, or at least about 27 nucleotides in length, or at least about 28 nucleotides in length, or at least about 29 nucleotides in length, or at least about 30 nucleotides in length.

**[0075]** In some aspects, the double-stranded region of a partially double-stranded identifier molecule can be about 1 nucleotide in length, about 2 nucleotides in length, or about 3 nucleotides in length, or about 4 nucleotides in length, or about 5 nucleotides in length, or about 6 nucleotides in length, or about 7 nucleotides in length, or about 8 nucleotides in length, or about 9 nucleotides in length, or about 10 nucleotides in length, or about 11 nucleotides in length, or about 12 nucleotides in length, or about 13 nucleotides in length, or about 14 nucleotides in length, or about 15 nucleotides in length, or about 16 nucleotides in length, or about 17 nucleotides in length, or about 18 nucleotides in length, or about 19 nucleotides in length, or about 20 nucleotides in length in length, or about 21 nucleotides in length, or about 22 nucleotides in length, or about 23 nucleotides in length, or about 24 nucleotides in length, or about 25 nucleotides in length, or about 26 nucleotides in length, or about 27 nucleotides in length, or about 28 nucleotides in length, or about 29 nucleotides in length, or about 30 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded identifier molecule is about 9 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded identifier molecule is about 10 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded identifier molecule is about 11 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded identifier molecule is about 12 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded identifier molecule is about 19 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded identifier molecule is about 20 nucleotides in length. In some aspects, the double-stranded region

of a partially double-stranded identifier molecule is about 21 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded identifier molecule is about 22 nucleotides in length.

**[0076]** In some aspects, an identifier sequence of a partially double-stranded identifier molecule can span the entire double-stranded region of a partially-double stranded identifier molecule. In some aspects, an identifier sequence of a partially double-stranded identifier molecule can span a portion of the double-stranded region of a partially-double stranded identifier molecule.

**[0077]** Accordingly, an identifier sequence can be at least about 1 nucleotide in length, at least about 2 nucleotides in length, or at least about 3 nucleotides in length, or at least about 4 nucleotides in length, or at least about 5 nucleotides in length, or at least about 6 nucleotides in length, or at least about 7 nucleotides in length, or at least about 8 nucleotides in length, or at least about 9 nucleotides in length, or at least about 10 nucleotides in length, or at least about 11 nucleotides in length, or at least about 12 nucleotides in length, or at least about 13 nucleotides in length, or at least about 14 nucleotides in length, or at least about 15 nucleotides in length, or at least about 16 nucleotides in length, or at least about 17 nucleotides in length, or at least about 18 nucleotides in length, or at least about 19 nucleotides in length, or at least about 20 nucleotides in length in length, or at least about 21 nucleotides in length, or at least about 22 nucleotides in length, or at least about 23 nucleotides in length, or at least about 24 nucleotides in length, or at least about 25 nucleotides in length, or at least about 26 nucleotides in length, or at least about 27 nucleotides in length, or at least about 28 nucleotides in length, or at least about 29 nucleotides in length, or at least about 30 nucleotides in length.

**[0078]** Accordingly, an identifier sequence be about 1 nucleotide in length, about 2 nucleotides in length, or about 3 nucleotides in length, or about 4 nucleotides in length, or about 5 nucleotides in length, or about 6 nucleotides in length, or about 7 nucleotides in length, or about 8 nucleotides in length, or about 9 nucleotides in length, or about 10 nucleotides in length, or about 11 nucleotides in length, or about 12 nucleotides in length, or about 13 nucleotides in length, or about 14 nucleotides in length, or about 15 nucleotides in length, or about 16 nucleotides in length, or about 17 nucleotides in length, or about 18 nucleotides in length, or about 19 nucleotides in length, or about 20 nucleotides in length in length, or about 21 nucleotides in length, or about 22 nucleotides in length, or about 23 nucleotides in length, or about 24 nucleotides in length, or about 25 nucleotides in length, or about 26 nucleotides in length, or about 27 nucleotides in length, or about



28 nucleotides in length, or about 29 nucleotides in length, or about 30 nucleotides in length. In some aspects, an identifier sequence is about 9 nucleotides in length. In some aspects, an identifier sequence is about 10 nucleotides in length. In some aspects, an identifier sequence is about 11 nucleotides in length. In some aspects, an identifier sequence is about 12 nucleotides in length. In some aspects, an identifier sequence is about 19 nucleotides in length. In some aspects, an identifier sequence is about 20 nucleotides in length. In some aspects, an identifier sequence is about 21 nucleotides in length. In some aspects, an identifier sequence is about 22 nucleotides in length.

[0079] Exemplary identifier sequences are shown in Table 1. Accordingly, an identifier sequence can comprise any of the sequences in Table 1, or a reverse complement thereof.

[0080] Table 1. Exemplary Identifier Sequences

<b>Strand #1</b>	<b>SEQ ID NO</b>	<b>Strand #2</b>	<b>SEQ ID NO</b>
AACCGCGGCT	1	GCCGCGGTTAGA	13
GGTTATAACT	2	GTTATAACCAGA	14
CCAAGTCCGT	3	CGGACTTGGAGA	15
TTGGACTTCT	4	GAAGTCCAAAGA	16
CAGTGGATCT	5	GATCCACTGAGA	17
TGACAAGCGT	6	CGCTTGTC AAGA	18
CGAGATATGT	7	CATATCTCGAGA	19
TAGAGCGCGT	8	CGCGCTCTAAGA	20
AACCTGTTGT	9	CAACAGGTTAGA	21
GGTTCACCGT	10	CGGTGAACCAGA	22
CATTGTTGCT	11	GCAACAATGAGA	23
TGCCACCAGT	12	CTGGTGGCAAGA	24
<b>Strand #1</b>	<b>SEQ ID NO</b>	<b>Strand #2</b>	<b>SEQ ID NO</b>
CTAGCTTGCT	25	GCAAGCTAGAGA	37
TCGATCCAGT	26	CTGGATCGAAGA	38
CCTGAACTGT	27	CAGTTCAGGAGA	39
TTCAGGTCGT	28	CGACCTGAAAGA	40
AGTAGAGAGT	29	CTCTCTACTAGA	41
GACGAGAGCT	30	GCTCTCGTCAGA	42
CTCTGCCTGT	31	CAGGCAGAGAGA	43
TCTCATTCGT	32	CGAATGAGAAGA	44
ACGCCGCAGT	33	CTGCGGCGTAGA	45
GTATTATGCT	34	GCATAATACAGA	46
GATAGATCGT	35	CGATCTATCAGA	47
AGCGAGCTGT	36	CAGCTCGCTAGA	48
<b>Strand #1</b>	<b>SEQ ID NO</b>	<b>Strand #2</b>	<b>SEQ ID NO</b>
AGACTTGGCT	49	GCCAAGTCTAGA	61
GAGTCCAAGT	50	CTTGGACTCAGA	62
CTTAAGCCGT	51	CGGCTTAAGAGA	63

TCCGGATTCT	52	GAATCCGGAAGA	64
CTGTATTACT	53	GTAATACAGAGA	65
TCACGCCGCT	54	GCGGCGTGAAGA	66
CAGTTCGCT	55	GCGGAACTGAGA	67
TGACCTTAGT	56	CTAAGGTCAAGA	68
CTAGGCAAGT	57	CTTGCCTAGAGA	69
TCGAATGGCT	58	GCCATTCGAAGA	70
CTTAGTGTCT	59	GACACTAAGAGA	71
TCCGACACGT	60	CGTGTCCGGAAGA	72
<b>Strand #1</b>	<b>SEQ ID NO</b>	<b>Strand #2</b>	<b>SEQ ID NO</b>
ACTTACATGT	73	CATGTAAGTAGA	85
GTCCGTGCGT	74	CGCACGGACAGA	86
AAGGTACCGT	75	CGGTACCTTAGA	87
GGAACGTTCT	76	GAACGTTCCAGA	88
AATTCTGCGT	77	CGCAGAATTAGA	89
GGCCTCATGT	78	CATGAGGCCAGA	90
AACAGGAAGT	79	CTTCCTGTTAGA	91
GGTGAAGGCT	80	GCCTTCACCAGA	92
CCTGTGGCGT	81	CGCCACAGGAGA	93
TTCACAATGT	82	CATTGTGAAAGA	94
ACACGAGTGT	83	CACTCGTGTAGA	95
GTGTAGACGT	84	CGTCTACACAGA	96
<b>Strand #1</b>	<b>SEQ ID NO</b>	<b>Strand #2</b>	<b>SEQ ID NO</b>
AGCGCTAGCGT	97	CGCTAGCGCTAGA	109
GATATCGATCT	98	GATCGATATCAGA	110
CGCAGACGCGT	99	CGCGTCTGCGAGA	111
TATGAGTATCT	100	GATACTCATAAGA	112
AGGTGCGTCGT	101	CGACGCACCTAGA	113
GAACATACGCT	102	GCGTATGTTTCCAGA	114
ATCTTAGTCGT	103	CGACTAAGATAGA	115
GCTCCGACTGT	104	CAGTCGGAGCAGA	116
ATACCAAGCGT	105	CGCTTGGTATAGA	117
GCGTTGGATCT	106	GATCCAACGCAGA	118
CTTCACGGCGT	107	CGCCGTGAAGAGA	119
TCCTGTAATCT	108	GATTACAGGAAGA	120
<b>Strand #1</b>	<b>SEQ ID NO</b>	<b>Strand #2</b>	<b>SEQ ID NO</b>
ACATAGCGCGT	121	CGCGCTATGTAGA	133
GTGCGATATCT	122	GATATCGCACAGA	134
CCAACAGATGT	123	CATCTGTTGGAGA	135
TTGGTGAGCGT	124	CGCTCACCAAAGA	136
CGCGGTTCTCT	125	GAGAACCGCGAGA	137
TATAACCTCGT	126	CGAGGTTATAAGA	138
AGAATGCCTGT	127	CAGGCATTCTAGA	139
GAGGCATTCGT	128	CGAATGCCTCAGA	140
CCTCGGTATCT	129	GATACCGAGGAGA	141

TTCTAACGCGT	130	CGCGTTAGAAAGA	142
ATGAGGCTCGT	131	CGAGCCTCATAGA	143
GCAGAATCTGT	132	CAGATTCTGCAGA	144
<b>Strand #1</b>	<b>SEQ ID NO</b>	<b>Strand #2</b>	<b>SEQ ID NO</b>
AAGGATGATCT	145	GATCATCCTTAGA	157
GGAAGCAGCGT	146	CGCTGCTTCCAGA	158
TCGTGACCGCT	147	GCGGTCACGAAGA	159
CTACAGTTCGT	148	CGAACTGTAGAGA	160
ATATTCACGCT	149	GCGTGAATATAGA	161
GCGCCTGTACT	150	GTACAGGCGCAGA	162
CACTACGATGT	151	CATCGTAGTGAGA	163
TGTCGTAGCGT	152	CGCTACGACAAGA	164
ACCACTTATGT	153	CATAAGTGGTAGA	165
GTTGTCCGCGT	154	CGCGGACAACAGA	166
ATCCATATCGT	155	CGATATGGATAGA	167
GCTTGCGCTCT	156	GAGCGCAAGCAGA	168
<b>Strand #1</b>	<b>SEQ ID NO</b>	<b>Strand #2</b>	<b>SEQ ID NO</b>
ACTCTATGCGT	169	CGCATAGAGTAGA	181
GTCTCGCATCT	170	GATGCGAGACAGA	182
AAGACGTCTGT	171	CAGACGTCTTAGA	183
GGAGTACTCGT	172	CGAGTACTCCAGA	184
ACCGGCCATGT	173	CATGGCCGGTAGA	185
GTTAATTGCGT	174	CGCAATTAACAGA	186
AGTATCTTCGT	175	CGAAGATACTAGA	187
GACGCTCCTGT	176	CAGGAGCGTCAGA	188
CATGCCATAGT	177	CTATGGCATGAGA	189
TGCATTGCGCT	178	GCGCAATGCAAGA	190
ATTGGAACGCT	179	GCGTTCCAATAGA	191
GCCAAGGTCGT	180	CGACCTTGGCAGA	192

**[0081]** The present disclosure provides pluralities of partially double-stranded identifier molecules.

**[0082]** The present disclosure provides pluralities of partially double-stranded identifier molecules, wherein the plurality comprises at least about one, or at least about two, or at least about three, or at least about four, or at least about five, or at least about six, or at least about seven, or at least about eight, or at least about nine, or at least about 10, or at least about 11, or at least about 12, or at least about 13, or at least about 14, or at least about 15, or at least about 16, or at least about 17, or at least about 18, or at least about 19, or at least about 20, or at least about 21, or at least about 22, or at least about 23, or at least about 24, or at least about 25, or at least about 26, or at least about 27, or at least about 28, or at least about 29, or at least about 30, or at

least about 31, or at least about 32, or at least about 33, or at least about 34, or at least about 35, or at least about 36, or at least about 37, or at least about 38, or at least about 39, or at least about 40, or at least about 41, or at least about 42, or at least about 43, or at least about 44, or at least about 45, or at least about 46, or at least about 47, or at least about 48, or at least about 49, or at least about 50, or at least about 51, or at least about 52, or at least about 53, or at least about 54, or at least about 55, or at least about 56, or at least about 57, or at least about 58, or at least about 59, or at least about 60, or at least about 61, or at least about 62, or at least about 63, or at least about 64, or at least about 65, or at least about 66, or at least about 67, or at least about 68, or at least about 69, or at least about 70, or at least about 71, or at least about 72, or at least about 73, or at least about 74, or at least about 75, or at least about 76, or at least about 77, or at least about 78, or at least about 79, or at least about 80, or at least about 81, or at least about 82, or at least about 83, or at least about 84, or at least about 85, or at least about 86, or at least about 87, or at least about 88, or at least about 89, or at least about 90, or at least about 91, or at least about 92, or at least about 93, or at least about 94, or at least about 95, or at least about 96, or at least about 97, or at least about 98, or at least about 99, or at least about 100 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0083]** The present disclosure provides pluralities of partially double-stranded identifier molecules, wherein the plurality comprises about one, or about two, or about three, or about four, or about five, or about six, or about seven, or about eight, or about nine, or about 10, or about 11, or about 12, or about 13, or about 14, or about 15, or about 16, or about 17, or about 18, or about 19, or about 20, or about 21, or about 22, or about 23, or about 24, or about 25, or about 26, or about 27, or about 28, or about 29, or about 30, or about 31, or about 32, or about 33, or about 34, or about 35, or about 36, or about 37, or about 38, or about 39, or about 40, or about 41, or about 42, or about 43, or about 44, or about 45, or about 46, or about 47, or about 48, or about 49, or about 50, or about 51, or about 52, or about 53, or about 54, or about 55, or about 56, or about 57, or about 58, or about 59, or about 60, or about 61, or about 62, or about 63, or about 64, or about 65, or about 66, or about 67, or about 68, or about 69, or about 70, or about 71, or about 72, or about 73, or about 74, or about 75, or about 76, or about 77, or about 78, or about 79, or about 80, or about 81, or about 82, or about 83, or about 84, or about 85, or about 86, or

about 87, or about 88, or about 89, or about 90, or about 91, or about 92, or about 93, or about 94, or about 95, or about 96, or about 97, or about 98, or about 99, or about 100 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0084]** In some aspects of the pluralities of partially double-stranded identifier molecules, each of the species of partially double-stranded identifier molecules can be present in the same amount, or different species of partially double-stranded identifier molecules can be present in different amounts.

**[0085]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about 12 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0086]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about 24 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0087]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about 48 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0088]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about 96 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0089]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about 12 species of partially double-stranded

identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0090]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about 24 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0091]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about 48 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0092]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about 96 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0093]** In some aspects of the pluralities of partially double-stranded identifier molecules of the present disclosure, the identifier sequence of one species of partially double-stranded identifier molecules have a hamming distance of at least about two, or at least about three, or at least about four, or at least about five, or at least about six, or at least about seven, or at least about eight, or at least about nine, or at least about ten to any other identifier sequence of any other species of partially double-stranded identifier molecules in the plurality. In some aspects of the pluralities of partially double-stranded identifier molecules of the present disclosure, the identifier sequence of one species of partially double-stranded identifier molecules have a hamming distance of about two, or about three, or about four, or about five, or about six, or about seven, or about eight, or about nine, or about ten to any other identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[0094]** As would be appreciated by the skilled artisan, the "hamming distance" between two identifier sequences, *identifier sequence x* and *identifier sequence y*, corresponds to the number

of changes that would need to be made in *identifier sequence x* to transform *identifier sequence x* into *identifier sequence y*, or vice versa.

**[0095]** *Partially double-stranded adapter molecules*

**[0096]** The present disclosure provides partially double-stranded adapter molecules. Partially double-stranded adapter molecules are nucleic acid molecules comprising at least one double-stranded region, at least three single stranded regions. In some aspects, a partially double-stranded adapter molecule is a nucleic acid molecule comprising one double-stranded region and three single stranded regions.

**[0097]** In some aspects a partially double-stranded adapter molecule comprises DNA. In some aspects, a partially double-stranded adapter molecule comprises RNA. In some aspects, a partially double-stranded adapter molecule can comprise XNA. In some aspects, a partially double-stranded adapter molecule comprises any combination of DNA, RNA and XNA.

**[0098]** In some aspects, a partially double-stranded adapter molecule can comprise one overhang. The overhang can be a 3' overhang or a 5' overhang.

**[0099]** In some aspects of the compositions of the present disclosure a 5' overhang of a partially double-stranded adapter molecule can be about 1 nucleotide in length, or about 2 nucleotides in length, or about 3 nucleotides in length, or about 4 nucleotides in length, or about 5 nucleotides in length. In some aspects of the compositions of the present disclosure a 5' overhang can be at least about 1 nucleotide, or at least about 2 nucleotides, or at least about 3 nucleotides, or at least about 4 nucleotides, or at least about 5 nucleotides, or at least about 6 nucleotides, or at least about 7 nucleotides, or at least about 8 nucleotides, or at least about 9 nucleotides, or at least about 10 nucleotides in length.

**[00100]** In some aspects, a 5' overhang of a partially double-stranded adapter molecule is no more than 1, or no more than 2, or no more than 3, or no more than 4, or no more than 5, or no more than 6, or no more than 7, or no more than 8, or no more than 9, or no more than 10 nucleotides, or no more than 11 nucleotides, or no more than 12 nucleotides, or no more than 13 nucleotide, or no more than 14 nucleotides, or no more than 15 nucleotides, or no more than 16 nucleotides, or no more than 17 nucleotides, or no more than 18 nucleotides, or no more than 19 nucleotides, or no more than 20 nucleotides in length.

**[00101]** In some aspects of the compositions of the present disclosure a 3' overhang of a partially double-stranded adapter molecule can be about 1 nucleotide in length, or about 2 nucleotides in

length, or about 3 nucleotides in length, or about 4 nucleotides in length, or about 5 nucleotides in length. In some aspects of the compositions of the present disclosure a 3' overhang can be at least about 1 nucleotide, or at least about 2 nucleotides, or at least about 3 nucleotides, or at least about 4 nucleotides, or at least about 5 nucleotides, or at least about 6 nucleotides, or at least about 7 nucleotides, or at least about 8 nucleotides, or at least about 9 nucleotides, or at least about 10 nucleotides in length.

**[00102]** In some aspects, a 3' overhang of a partially double-stranded adapter molecule is no more than 1, or no more than 2, or no more than 3, or no more than 4, or no more than 5, or no more than 6, or no more than 7, or no more than 8, or no more than 9, or no more than 10 nucleotides, or no more than 11 nucleotides, or no more than 12 nucleotides, or no more than 13 nucleotide, or no more than 14 nucleotides, or no more than 15 nucleotides, or no more than 16 nucleotides, or no more than 17 nucleotides, or no more than 18 nucleotides, or no more than 19 nucleotides, or no more than 20 nucleotides in length.

**[00103]** In some aspects, a 3' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length. In some aspects, a 3' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is an adenine or a thymine. In some aspects, a 3' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is an adenine. In some aspects, a 3' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a thymine. In some aspects, a 3' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a guanosine or a cytosine. In some aspects, a 3' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a guanosine. In some aspects, a 3' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a cytidine.

**[00104]** In some aspects, a 5' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length. In some aspects, a 5' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is an adenine or a thymine. In some aspects, a 5' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is an adenine. In some aspects, a 5' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a thymine. In some aspects, a 5' overhang of a partially double-stranded adapter molecule can be 1



nucleotide in length, wherein the 1 nucleotide is a guanosine or a cytosine. In some aspects, a 5' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a guanosine. In some aspects, a 5' overhang of a partially double-stranded adapter molecule can be 1 nucleotide in length, wherein the 1 nucleotide is a cytidine.

**[00105]** In some aspects, a partially double-stranded adapter molecule can comprise a single-stranded arm. As would be appreciated by the skilled artisan, an "arm", in the context of a partially double-stranded nucleic acid molecule refers to a single-stranded region of a partially double-stranded nucleic acid molecule located at a terminus of the partially double-stranded nucleic acid for which there is a corresponding single-stranded region located directly on the opposite strand. In some aspects, a single-stranded arm can be a single-stranded 5' arm. In some aspects, a single-stranded arm can be a single-stranded 3' arm. FIG. 3 shows both single-stranded 5' arms and single-stranded 3' arms in exemplary partially double-stranded adapter molecules of the present disclosure.

**[00106]** In some aspects, a single-stranded 5' arm and/or single-stranded 3' arm can be at least about 1 nucleotide in length, at least about 2 nucleotides in length, or at least about 3 nucleotides in length, or at least about 4 nucleotides in length, or at least about 5 nucleotides in length, or at least about 6 nucleotides in length, or at least about 7 nucleotides in length, or at least about 8 nucleotides in length, or at least about 9 nucleotides in length, or at least about 10 nucleotides in length, or at least about 11 nucleotides in length, or at least about 12 nucleotides in length, or at least about 13 nucleotides in length, or at least about 14 nucleotides in length, or at least about 15 nucleotides in length, or at least about 16 nucleotides in length, or at least about 17 nucleotides in length, or at least about 18 nucleotides in length, or at least about 19 nucleotides in length, or at least about 20 nucleotides in length in length, or at least about 21 nucleotides in length, or at least about 22 nucleotides in length, or at least about 23 nucleotides in length, or at least about 24 nucleotides in length, or at least about 25 nucleotides in length, or at least about 26 nucleotides in length, or at least about 27 nucleotides in length, or at least about 28 nucleotides in length, or at least about 29 nucleotides in length, or at least about 30 nucleotides in length.

**[00107]** In some aspects, a single-stranded 5' arm and/or single-stranded 3' arm can be about 1 nucleotide in length, about 2 nucleotides in length, or about 3 nucleotides in length, or about 4 nucleotides in length, or about 5 nucleotides in length, or about 6 nucleotides in length, or about 7 nucleotides in length, or about 8 nucleotides in length, or about 9 nucleotides in length, or about

10 nucleotides in length, or about 11 nucleotides in length, or about 12 nucleotides in length, or about 13 nucleotides in length, or about 14 nucleotides in length, or about 15 nucleotides in length, or about 16 nucleotides in length, or about 17 nucleotides in length, or about 18 nucleotides in length, or about 19 nucleotides in length, or about 20 nucleotides in length in length, or about 21 nucleotides in length, or about 22 nucleotides in length, or about 23 nucleotides in length, or about 24 nucleotides in length, or about 25 nucleotides in length, or about 26 nucleotides in length, or about 27 nucleotides in length, or about 28 nucleotides in length, or about 29 nucleotides in length, or about 30 nucleotides in length.

**[00108]** A single-stranded 5' arm and/or single-stranded 3' arm can comprise an amplification primer binding site that hybridizes to an amplification primer.

**[00109]** As would be appreciated by the skilled artisan, an amplification primer binding site is a nucleic acid sequence that is capable of being bound by a primer suitable for priming an amplification reaction using a nucleic acid polymerase. As would be appreciated by the skilled artisan, these amplification primer binding sites can be used to generate sequencing libraries using techniques that are standard in the art and well-known to the skilled artisan.

**[00110]** In some aspects, a partially double-stranded adapter molecule can comprise an identifier sequence, as is described above. In some aspects, an identifier sequence located in a partially double-stranded adapter molecule can be located in a double-stranded region of the partially double-stranded adapter molecule. In some aspects, an identifier sequence of a partially double-stranded adapter molecule can span the entire double-stranded region of a partially-double stranded adapter molecule. In some aspects, an identifier sequence of a partially double-stranded adapter molecule can span a region of the double-stranded region of a partially-double stranded adapter molecule.

**[00111]** In some aspects, the double-stranded region of a partially double-stranded adapter molecule can be at least about 1 nucleotide in length, at least about 2 nucleotides in length, or at least about 3 nucleotides in length, or at least about 4 nucleotides in length, or at least about 5 nucleotides in length, or at least about 6 nucleotides in length, or at least about 7 nucleotides in length, or at least about 8 nucleotides in length, or at least about 9 nucleotides in length, or at least about 10 nucleotides in length, or at least about 11 nucleotides in length, or at least about 12 nucleotides in length, or at least about 13 nucleotides in length, or at least about 14 nucleotides in length, or at least about 15 nucleotides in length, or at least about 16 nucleotides in length, or at

least about 17 nucleotides in length, or at least about 18 nucleotides in length, or at least about 19 nucleotides in length, or at least about 20 nucleotides in length, or at least about 21 nucleotides in length, or at least about 22 nucleotides in length, or at least about 23 nucleotides in length, or at least about 24 nucleotides in length, or at least about 25 nucleotides in length, or at least about 26 nucleotides in length, or at least about 27 nucleotides in length, or at least about 28 nucleotides in length, or at least about 29 nucleotides in length, or at least about 30 nucleotides in length.

**[00112]** In some aspects, the double-stranded region of a partially double-stranded adapter molecule can be about 1 nucleotide in length, about 2 nucleotides in length, or about 3 nucleotides in length, or about 4 nucleotides in length, or about 5 nucleotides in length, or about 6 nucleotides in length, or about 7 nucleotides in length, or about 8 nucleotides in length, or about 9 nucleotides in length, or about 10 nucleotides in length, or about 11 nucleotides in length, or about 12 nucleotides in length, or about 13 nucleotides in length, or about 14 nucleotides in length, or about 15 nucleotides in length, or about 16 nucleotides in length, or about 17 nucleotides in length, or about 18 nucleotides in length, or about 19 nucleotides in length, or about 20 nucleotides in length, or about 21 nucleotides in length, or about 22 nucleotides in length, or about 23 nucleotides in length, or about 24 nucleotides in length, or about 25 nucleotides in length, or about 26 nucleotides in length, or about 27 nucleotides in length, or about 28 nucleotides in length, or about 29 nucleotides in length, or about 30 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded adapter molecule is about 9 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded adapter molecule is about 10 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded adapter molecule is about 11 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded adapter molecule is about 12 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded adapter molecule is about 19 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded adapter molecule is about 20 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded adapter molecule is about 21 nucleotides in length. In some aspects, the double-stranded region of a partially double-stranded adapter molecule is about 22 nucleotides in length.

**[00113]** In some aspects, the partially double-stranded adapter molecules of the present disclosure comprise a single-stranded 5' arm, a single-stranded 3' arm, a double-stranded region and a 3' overhang. An exemplary schematic of the preceding partially double-stranded adapter molecule is shown in the top panel of FIG. 3.

**[00114]** In some aspects, the partially double-stranded adapter molecules of the present disclosure comprise a single-stranded 5' arm, a single-stranded 3' arm, a double-stranded region and a 5' overhang. An exemplary schematic of the preceding partially double-stranded adapter molecule is shown in the top panel of FIG. 3.

**[00115]** In the preceding partially double-stranded adapter molecules, the single-stranded 5' arm, the single-stranded 3' arm, or both the single-stranded 5' arm and the single-stranded 3' arm can comprise amplification primer binding sites.

**[00116]** In the preceding partially double-stranded adapter molecules, the double-stranded region can comprise an identifier sequence.

**[00117]** The present disclosure provides pluralities of partially double-stranded adapter molecules.

**[00118]** The present disclosure provides pluralities of partially double-stranded adapter molecules, wherein the plurality comprises at least about one, or at least about two, or at least about three, or at least about four, or at least about five, or at least about six, or at least about seven, or at least about eight, or at least about nine, or at least about 10, or at least about 11, or at least about 12, or at least about 13, or at least about 14, or at least about 15, or at least about 16, or at least about 17, or at least about 18, or at least about 19, or at least about 20, or at least about 21, or at least about 22, or at least about 23, or at least about 24, or at least about 25, or at least about 26, or at least about 27, or at least about 28, or at least about 29, or at least about 30, or at least about 31, or at least about 32, or at least about 33, or at least about 34, or at least about 35, or at least about 36, or at least about 37, or at least about 38, or at least about 39, or at least about 40, or at least about 41, or at least about 42, or at least about 43, or at least about 44, or at least about 45, or at least about 46, or at least about 47, or at least about 48, or at least about 49, or at least about 50, or at least about 51, or at least about 52, or at least about 53, or at least about 54, or at least about 55, or at least about 56, or at least about 57, or at least about 58, or at least about 59, or at least about 60, or at least about 61, or at least about 62, or at least about 63, or at least about 64, or at least about 65, or at least about 66, or at least about 67, or at least about 68, or at least about 69, or at least about 70, or at least about 71, or at least about 72, or at least about 73, or at least about 74, or at least

about 75, or at least about 76, or at least about 77, or at least about 78, or at least about 79, or at least about 80, or at least about 81, or at least about 82, or at least about 83, or at least about 84, or at least about 85, or at least about 86, or at least about 87, or at least about 88, or at least about 89, or at least about 90, or at least about 91, or at least about 92, or at least about 93, or at least about 94, or at least about 95, or at least about 96, or at least about 97, or at least about 98, or at least about 99, or at least about 100 species of partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter molecules in the plurality.

**[00119]** The present disclosure provides pluralities of partially double-stranded adapter molecules, wherein the plurality comprises about one, or about two, or about three, or about four, or about five, or about six, or about seven, or about eight, or about nine, or about 10, or about 11, or about 12, or about 13, or about 14, or about 15, or about 16, or about 17, or about 18, or about 19, or about 20, or about 21, or about 22, or about 23, or about 24, or about 25, or about 26, or about 27, or about 28, or about 29, or about 30, or about 31, or about 32, or about 33, or about 34, or about 35, or about 36, or about 37, or about 38, or about 39, or about 40, or about 41, or about 42, or about 43, or about 44, or about 45, or about 46, or about 47, or about 48, or about 49, or about 50, or about 51, or about 52, or about 53, or about 54, or about 55, or about 56, or about 57, or about 58, or about 59, or about 60, or about 61, or about 62, or about 63, or about 64, or about 65, or about 66, or about 67, or about 68, or about 69, or about 70, or about 71, or about 72, or about 73, or about 74, or about 75, or about 76, or about 77, or about 78, or about 79, or about 80, or about 81, or about 82, or about 83, or about 84, or about 85, or about 86, or about 87, or about 88, or about 89, or about 90, or about 91, or about 92, or about 93, or about 94, or about 95, or about 96, or about 97, or about 98, or about 99, or about 100 species of partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter molecules in the plurality.

**[00120]** In some aspects of the pluralities of partially double-stranded adapter molecules, each of the species of partially double-stranded adapter molecules can be present in the same amount, or different species of partially double-stranded adapter molecules can be present in different amounts.

**[00121]** In some aspects, any of the partially double-stranded nucleic acid molecules described herein, including partially double-stranded identifier molecules and partially double-stranded adapter molecules can comprise at least one modified nucleic acid. In some aspects, a modified nucleic acid can comprise methylated cytidine. In some aspects, a modified nucleic acid can comprise 5mC (5-methylcytosine), 5hmC (5-hydroxymethylcytosine), 5fC (5-formylcytosine), 3mA (3-methyladenine), 5-fU (5-formyluridine), 5-hmU (5-hydroxymethyluridine), 5-hoU (5-hydroxyuridine), 7mG (7-methylguanine), 8oxoG (8-oxo-7,8-dihydroguanine), AP (apurinic/apyrimidinic sites), CPDs (Cyclobutane pyrimidine dimers), dI (deoxyinosine), dR5P (deoxyribose 5'-phosphate), dU (deoxyuridine), dX (deoxyxanthosine), PA (3'-phospho- $\alpha$ ,  $\beta$ -unsaturated aldehyde), rN (ribonucleotides), Tg (Thymine Glycol), TT (TT dimer) and/or Mismatches including AP:A (apurinic/apyrimidinic site base paired with adenine), DHT:A (5,6-dihydrothymine base paired with an adenine), 5-hmU:A (5-hydroxymethyluracil base paired with an adenine), 5-hmU:G (5-hydroxymethyluracil base paired with a guanine), I:T (inosine base paired with a thymine), 6-MeA:T (6-methyladenine base paired with a thymine), 8-OG:C (8-oxoguanine base paired with a cytosine), 8-OG:G (8-oxoguanine base paired with a guanine), U:A (uridine base paired with an adenine) or U:G (uridine base paired with a guanine) or any combination thereof.

**[00122]** Kits

**[00123]** The present disclosure provides kits comprising the compositions of the present disclosure. These compositions include, but are not limited to, the any of the partially double-stranded nucleic acid molecules described herein, including, but not limited to, partially double-stranded identifier molecules and partially double-stranded adapter molecules; any of the pluralities of partially double-stranded nucleic acid molecules, including, but not limited to pluralities of partially double-stranded identifier molecules and pluralities of partially double-stranded adapter molecules.

**[00124]** Accordingly, the present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about one, or at least about two, or at least about three, or at least about four, or at least about five, or at least about six, or at least about seven, or at least about eight, or at least about nine, or at least about 10, or at least about 11, or at least about 12, or at least about 13, or at least about 14, or at least about 15, or at least about 16, or at least about 17, or at least about 18, or at least about 19, or at

least about 20, or at least about 21, or at least about 22, or at least about 23, or at least about 24, or at least about 25, or at least about 26, or at least about 27, or at least about 28, or at least about 29, or at least about 30, or at least about 31, or at least about 32, or at least about 33, or at least about 34, or at least about 35, or at least about 36, or at least about 37, or at least about 38, or at least about 39, or at least about 40, or at least about 41, or at least about 42, or at least about 43, or at least about 44, or at least about 45, or at least about 46, or at least about 47, or at least about 48, or at least about 49, or at least about 50, or at least about 51, or at least about 52, or at least about 53, or at least about 54, or at least about 55, or at least about 56, or at least about 57, or at least about 58, or at least about 59, or at least about 60, or at least about 61, or at least about 62, or at least about 63, or at least about 64, or at least about 65, or at least about 66, or at least about 67, or at least about 68, or at least about 69, or at least about 70, or at least about 71, or at least about 72, or at least about 73, or at least about 74, or at least about 75, or at least about 76, or at least about 77, or at least about 78, or at least about 79, or at least about 80, or at least about 81, or at least about 82, or at least about 83, or at least about 84, or at least about 85, or at least about 86, or at least about 87, or at least about 88, or at least about 89, or at least about 90, or at least about 91, or at least about 92, or at least about 93, or at least about 94, or at least about 95, or at least about 96, or at least about 97, or at least about 98, or at least about 99, or at least about 100 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00125]** The present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about one, or about two, or about three, or about four, or about five, or about six, or about seven, or about eight, or about nine, or about 10, or about 11, or about 12, or about 13, or about 14, or about 15, or about 16, or about 17, or about 18, or about 19, or about 20, or about 21, or about 22, or about 23, or about 24, or about 25, or about 26, or about 27, or about 28, or about 29, or about 30, or about 31, or about 32, or about 33, or about 34, or about 35, or about 36, or about 37, or about 38, or about 39, or about 40, or about 41, or about 42, or about 43, or about 44, or about 45, or about 46, or about 47, or about 48, or about 49, or about 50, or about 51, or about 52, or about 53, or about 54, or about 55, or about 56, or about 57, or about 58, or about 59, or about 60, or about 61, or about 62, or about

63, or about 64, or about 65, or about 66, or about 67, or about 68, or about 69, or about 70, or about 71, or about 72, or about 73, or about 74, or about 75, or about 76, or about 77, or about 78, or about 79, or about 80, or about 81, or about 82, or about 83, or about 84, or about 85, or about 86, or about 87, or about 88, or about 89, or about 90, or about 91, or about 92, or about 93, or about 94, or about 95, or about 96, or about 97, or about 98, or about 99, or about 100 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00126]** The present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about 12 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00127]** The present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about 24 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00128]** The present disclosure provides a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about 48 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00129]** The present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises at least about 96 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00130]** The present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about 12 species of partially double-



stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00131]** The present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about 24 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00132]** The present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about 48 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00133]** The present disclosure provides a kit comprising a plurality of partially double-stranded identifier molecules, wherein the plurality comprises about 96 species of partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00134]** In some aspects of the kits of the present disclosure, each species of partially double-stranded identifier molecules is kept physically separate from other species of partially double-stranded identifier molecules.

**[00135]** In a non-limiting example, physical separation can be accomplished by enclosing each species of partially double-stranded identifier molecules in a separate container (*e.g.* different wells in a microplate, different sample tubes, etc.). By physically separate each species of partially double-stranded identifier molecules, the kit allows the user to optimize the number of barcode combinations to be used with each sample that is to be analyzed using the kit.

**[00136]** The kits of the present disclosure can further comprise a plurality of partially double-stranded adapter molecules. In some aspects, the plurality of partially double-stranded adapter molecules comprises at least about one, or at least about two, or at least about three, or at least about four, or at least about five, or at least about six, or at least about seven, or at least about eight, or at least about nine, or at least about 10, or at least about 11, or at least about 12, or at

least about 13, or at least about 14, or at least about 15, or at least about 16, or at least about 17, or at least about 18, or at least about 19, or at least about 20, or at least about 21, or at least about 22, or at least about 23, or at least about 24, or at least about 25, or at least about 26, or at least about 27, or at least about 28, or at least about 29, or at least about 30, or at least about 31, or at least about 32, or at least about 33, or at least about 34, or at least about 35, or at least about 36, or at least about 37, or at least about 38, or at least about 39, or at least about 40, or at least about 41, or at least about 42, or at least about 43, or at least about 44, or at least about 45, or at least about 46, or at least about 47, or at least about 48, or at least about 49, or at least about 50, or at least about 51, or at least about 52, or at least about 53, or at least about 54, or at least about 55, or at least about 56, or at least about 57, or at least about 58, or at least about 59, or at least about 60, or at least about 61, or at least about 62, or at least about 63, or at least about 64, or at least about 65, or at least about 66, or at least about 67, or at least about 68, or at least about 69, or at least about 70, or at least about 71, or at least about 72, or at least about 73, or at least about 74, or at least about 75, or at least about 76, or at least about 77, or at least about 78, or at least about 79, or at least about 80, or at least about 81, or at least about 82, or at least about 83, or at least about 84, or at least about 85, or at least about 86, or at least about 87, or at least about 88, or at least about 89, or at least about 90, or at least about 91, or at least about 92, or at least about 93, or at least about 94, or at least about 95, or at least about 96, or at least about 97, or at least about 98, or at least about 99, or at least about 100 species of partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter molecules in the plurality. In some aspects, the plurality of partially double-stranded adapter molecules comprises about one, or about two, or about three, or about four, or about five, or about six, or about seven, or about eight, or about nine, or about 10, or about 11, or about 12, or about 13, or about 14, or about 15, or about 16, or about 17, or about 18, or about 19, or about 20, or about 21, or about 22, or about 23, or about 24, or about 25, or about 26, or about 27, or about 28, or about 29, or about 30, or about 31, or about 32, or about 33, or about 34, or about 35, or about 36, or about 37, or about 38, or about 39, or about 40, or about 41, or about 42, or about 43, or about 44, or about 45, or about 46, or about 47, or about 48, or about 49, or about 50, or about 51, or about 52, or about 53, or about 54, or about 55, or about 56, or about 57, or about 58, or about 59, or about 60, or about 61, or about 62, or about

63, or about 64, or about 65, or about 66, or about 67, or about 68, or about 69, or about 70, or about 71, or about 72, or about 73, or about 74, or about 75, or about 76, or about 77, or about 78, or about 79, or about 80, or about 81, or about 82, or about 83, or about 84, or about 85, or about 86, or about 87, or about 88, or about 89, or about 90, or about 91, or about 92, or about 93, or about 94, or about 95, or about 96, or about 97, or about 98, or about 99, or about 100 species of partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter molecules in the plurality.

**[00137]** The kits of the present disclosure can further comprise a plurality of enzymes to mediate end-repair on double-stranded DNA molecules. Such pluralities of enzymes are well-known to the skilled artisan and include, but are not limited to, pluralities comprising DNA polymerases (*e.g.* T4 DNA polymerase), klenow fragments, polynucleotide kinases (*e.g.* T4 polynucleotide kinase) or any combination thereof.

**[00138]** The kits of the present disclosure can further comprise a plurality of reagents suitable for the purification of nucleic acid molecules. Such pluralities of reagents are well-known to the skilled artisan.

**[00139]** The kits of the present disclosure can further comprise at least one DNA ligase. The DNA ligase can be any DNA ligase known in the art, including but not limited to, T4 DNA ligase, T7 DNA ligase or any other DNA ligase known in the art.

**[00140]** The kits of the present disclosure can further comprise a plurality of amplification primers that bind to one or more of the amplification primer binding sites located on partially double-stranded adapter molecules.

**[00141]** The kits of the present disclosure can further comprise at least one DNA polymerase. In some aspects, the at least one DNA polymerase is able to catalyze amplification via the amplification primers that bind to one or more of the amplification primer binding sites located on partially double-stranded adapter molecules.

**[00142]** The kits of the present disclosure can further comprise written instructions for the performance of the methods of the present disclosure.

**[00143]** Methods of sequencing target nucleic acids

**[00144]** The present disclosure provides methods for sequencing target nucleic acids. The sequencing methods, compositions and kits of the present disclosure exhibit superior properties

as compared to existing NGS methods that use pre-pooled unique molecular identifiers (UMIs). As would be appreciated by the skilled artisan, existing NGS methods rely on the expensive synthesis of an entire adapter molecule per each barcode sequence that is to be used in an experiment. Moreover, in existing pre-pooled barcoded adapter products, there is no flexibility in the number and the length of barcodes that are used for individual samples. Moreover, existing pooled barcodes increase the risk of cross-talk and have a maximum hamming distance of one, so error-correction of barcodes is not possible. In contrast to existing methods, the sequencing composition, kits and methods of the present disclosure are more cost-effective, as only a single adapter needs to be synthesized for use with all identifier sequences. Moreover, the compositions, kits and methods of the present disclosure allow for a fully customizable number of barcodes to be used for each sample. That is, the number of barcodes used for a particular sample can be optimized for that particular sample type and/or experimental objective. The compositions, kits and methods of the present disclosure allow for all identifier sequences to remain completely independent, reducing the risk of crosstalk. Finally, the identifier sequences of the compositions, kits and methods of the present disclosure having hamming distances of at least two, allowing for error-correction and increased barcode fidelity. FIG. 7 shows a schematic comparison between existing next generation sequencing barcode compositions and methods and the compositions and the methods of the present disclosure.

**[00145]** In the methods of the present disclosure, the ligation of a partially double-stranded identifier molecule of the present disclosure to each of a transcript in a plurality of target nucleic acids results in the creation of a UMI sequence that is ligated to that transcript. In other words, the transcript becomes tagged with a combination of two identifier sequences through the ligation of partially double-stranded identifier molecules to each end. For example, in a method wherein four species of partially double-stranded identifier molecules are used, say *identifier x*, *identifier y*, *identifier w*, and *identifier z*, the random ligation of one of partially double-stranded identifier molecules to each end of the transcript could create one of 16 UMIs, as shown in Table 2.

**[00146]** Table 2. Exemplary UMI Combinations:

<i>Identifier x + Identifier x</i>
<i>Identifier y + Identifier y</i>
<i>Identifier w + Identifier w</i>

<i>Identifier z + Identifier z</i>
<i>Identifier x + Identifier y</i>
<i>Identifier x + Identifier w</i>
<i>Identifier x + Identifier z</i>
<i>Identifier y + Identifier x</i>
<i>Identifier y + Identifier w</i>
<i>Identifier y + Identifier z</i>
<i>Identifier z + Identifier x</i>
<i>Identifier z + Identifier w</i>
<i>Identifier z + Identifier y</i>
<i>Identifier w + Identifier x</i>
<i>Identifier w + Identifier z</i>
<i>Identifier w + Identifier y</i>

**[00147]** These UMI sequences that are created by the ligation steps of the methods of the present disclosure can then be used in analysis using methods standard in the art, including, but not limited to, error correction, consensus sequence creation, etc.

**[00148]** In some aspects, target nucleic acids are double-stranded nucleic acid molecules.

**[00149]** In some aspects, target nucleic acids can comprise DNA, RNA or a combination of DNA and RNA.

**[00150]** Target nucleic acids can be derived from any source, including, but not limited to any biological sample. Target nucleic acids can be extracted from biological samples using techniques that are standard in the art. After extraction from a biological samples, target nucleic acids and be processed using techniques that are standard in the art prior to being subjected to the methods of the present disclosure. These processing methods can include, but are not limited to, fragmentation, reverse transcription, end-repair or any other nucleic acid processing technique known in the art. In aspects wherein RNA is extracted from a biological sample, the RNA can be reverse transcribed into DNA prior to being subjected to the methods of the present disclosure.

**[00151]** The sequencing methods of the present disclosure can comprise: a) ligating a first partially double-stranded identifier molecule to one end of a target nucleic acid; b) ligating a second partially double-stranded identifier molecule to the other end of the target nucleic acid; c)

ligating a first partially double-stranded adapter molecule to the first partially double-stranded identifier molecule; and d) ligating a second partially double-stranded adapter molecule to the second partially double-stranded identifier molecule.

**[00152]** In some aspects of the preceding method, steps (a) and (b) can be performed sequentially. In some aspects of the preceding method, steps (a) and (b) can be performed concurrently. In some aspects of the preceding method, steps (c) and (d) can be performed sequentially. In some aspects of the preceding method, steps (c) and (d) can be performed concurrently. A non-limiting example of the preceding method is shown in FIG. 1 and FIG. 2. In the top panel of FIG. 2, a first partially double-stranded identifier molecule and a second partially double-stranded identifier molecule are ligated to the ends of a target nucleic acid. In the middle panel, a first partially double-stranded adapter molecule and a second partially double-stranded adapter molecule are ligated to the first partially double-stranded identifier molecule and the second partially double-stranded identifier molecule, respectively.

**[00153]** The present disclosure provides methods of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids; b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00154]** The present disclosure provides methods of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the plurality of partially double-stranded identifier molecules of the present disclosure comprises at least about two species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially

double-stranded identifier molecules in the plurality, wherein the ligation products comprise each of the at least four combinations of two species of partially double-stranded identifier molecules; b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00155]** The present disclosure provides methods of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the plurality of partially double-stranded identifier molecules of the present disclosure comprises at least about 12 species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality, wherein the ligation products comprise each of the at least 144 combinations of two species of partially double-stranded identifier molecules; b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00156]** The present disclosure provides methods of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the plurality of partially double-stranded identifier molecules of the present disclosure comprises at least about 24 species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality, wherein the ligation products comprise each

of the at least 576 combinations of two species of partially double-stranded identifier molecules; b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00157]** The present disclosure provides methods of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the plurality of partially double-stranded identifier molecules of the present disclosure comprises at least about 96 species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality, wherein the ligation products comprise each of the at least 9,216 combinations of two species of partially double-stranded identifier molecules; b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00158]** The present disclosure provides a method of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the plurality of partially double-stranded identifier molecules of the present disclosure comprises at least about two species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality, wherein the ligation products comprise at least about 5%, or at least about 10%, or at least about 15%, or at least about 20%, or at least



about 25%, or at least about 30%, or at least about 35%, or at least about 40%, or at least about 45%, or at least about 50%, or at least about 55%, or at least about 60%, or at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95% of the at least four combinations of two species of partially double-stranded identifier molecules; b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00159]** The present disclosure provides a method of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the plurality of partially double-stranded identifier molecules of the present disclosure comprises at least about 12 species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality, wherein the ligation products comprise at least about 5%, or at least about 10%, or at least about 15%, or at least about 20%, or at least about 25%, or at least about 30%, or at least about 35%, or at least about 40%, or at least about 45%, or at least about 50%, or at least about 55%, or at least about 60%, or at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95% of the at least 144 combinations of two species of partially double-stranded identifier molecules; b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00160]** The present disclosure provides a method of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each

end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the plurality of partially double-stranded identifier molecules of the present disclosure comprises at least about 24 species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality, wherein the ligation products comprise at least about 5%, or at least about 10%, or at least about 15%, or at least about 20%, or at least about 25%, or at least about 30%, or at least about 35%, or at least about 40%, or at least about 45%, or at least about 50%, or at least about 55%, or at least about 60%, or at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95% of the at least 576 combinations of two species of partially double-stranded identifier molecules; b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00161]** The present disclosure provides a method of sequencing a plurality of double-stranded target nucleic acids comprising: a) contacting the plurality of double-stranded target nucleic acids with a plurality of partially double-stranded identifier molecules of the present disclosure and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the plurality of partially double-stranded identifier molecules of the present disclosure comprises at least about 96 species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality, wherein the ligation products comprise at least about 5%, or at least about 10%, or at least about 15%, or at least about 20%, or at least about 25%, or at least about 30%, or at least about 35%, or at least about 40%, or at least about 45%, or at least about 50%, or at least about 55%, or at least about 60%, or at least about 65%, or at least about 70%, or at least about 75%, or at least about 80%, or at least about 85%, or at least about 90%, or at least about 95% of the at least 9,216 combinations of two species of partially double-stranded identifier molecules; b) contacting the products of step (a) with a plurality of

partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and c) sequencing the products of step (b).

**[00162]** In some aspects of the preceding methods, sequencing can be performed using any sequencing method known in the art, including, but not limited to, next generation sequencing methods, sequencing-by-synthesis methods, sequencing by ligation methods, single-molecule real-time sequencing methods, ion semiconductor sequencing methods, pyrosequencing methods, combinatorial probe anchor synthesis sequencing methods, nanopore sequencing methods, genanpsys sequencing methods, sanger sequencing methods or any other sequencing method known in the art.

**[00163]** In some aspects of the preceding methods, the methods can further comprise after step (b) and prior to step (c), constructing a sequencing library using the products of step (b). The sequencing library can be constructed using standard library construction techniques known in the art. These library construction techniques can comprise amplifying the products of step (b) by contacting the products of step (b) with amplification primers that bind to amplification primer binding sites and at least one polymerase. The amplification can comprise the introduction of sequencing adapters that are suitable for use in the sequencing method of choice. In some aspects, the library construction techniques can comprise nucleic acid purification techniques that are known in the art.

**[00164]** In some aspects, the preceding method can further comprise determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c). As would be appreciated by the skilled artisan, the identifier sequences of the ligated partially double-stranded identifier molecules can be used in the analysis of the sequencing data to determine the abundance of specific transcripts by allowing the skilled artisan to correct various errors introduced during the sequencing process (including, but not limited to, amplification errors) using methods standard in the art. As would be appreciated by the skilled artisan, the identifier sequences of the ligated partially double-stranded identifier molecules can be used in the analysis of the sequencing data to determine the identity of specific transcripts by allowing the skilled artisan to create consensus sequences using methods standard in the art.

**[00165]** In some aspects, determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) can comprise grouping the sequencing reads obtained in step (c) by the ligated identifier sequences in the sequencing reads.

**[00166]** In some aspects, determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequence data generated in step (c) can comprise grouping the sequencing reads obtained in step (c) by the specific genomic sequence that the sequencing reads most likely correspond to.

**[00167]** In some aspects, determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) can comprise first grouping the sequencing reads obtained in step (c) by the ligated identifier sequences in the sequencing reads and then further grouping by the specific genomic sequence that the sequencing reads most likely correspond to.

**[00168]** Without wishing to be bound by theory, for existing pre-pooled, degenerate UMIs, the number of UMIs available should be larger than the number of molecules present within the initial sample. This ensures each molecule gets a unique UMI. This approach leads to a large majority of UMIs containing only a single read. With a minimum requirement of at least two reads per UMI, to generate a consensus sequence, the UMIs containing only a single read are discarded. The inability to produce a consensus read for a large majority of the available UMIs, means that very high sequencing depths are required for each region of interest. These limitations are overcome by the methods of the present disclosure by: 1) giving the user the ability to modulate the number of discrete UMIs used for the initial sample; 2) using the sequence from the region of interest as a UMI itself.

**[00169]** In some aspects of the methods of the present disclosure, determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) can comprising grouping together aligned sequencing reads based on their absolute alignment to a reference sequence (*e.g.* a known genomic sequence). In some aspects, these aligned sequencing reads can then be further grouped based on their similarity from the reference sequence. In some aspects, the aligned sequencing reads can then further be sub-divided by their UMIs. Because the methods of the present disclosure allow the number of initial UMIs to be modulated, the number of sequencing reads per UMI can be

modulated to have on average at least two sequencing reads per UMI. Optimization of the number of reads per UMI, allows the majority of reads (and therefore UMIs) to produce usable consensus reads, therefore reducing the coverage required per region of interest.

**[00170]** Accordingly, in the methods of the present disclosure, the number of species of partially double-stranded identifier molecules that are used can be selected such that there is on average at least about two, or at least about three, or at least about four, or at least about five, or at least about six, or at least about seven, or at least about eight, or at least about nine, or at least about ten sequencing reads per UMI (*i.e.* combination of two species of double-stranded identifier molecules ligated onto a target transcript).

**[00171]** Accordingly, in the methods of the present disclosure, the number of species of partially double-stranded identifier molecules that are used can be selected such that there is at least about two, or at least about three, or at least about four, or at least about five, or at least about six, or at least about seven, or at least about eight, or at least about nine, or at least about ten sequencing reads per UMI (*i.e.* combination of two species of double-stranded identifier molecules ligated onto a target transcript).

**[00172]** An exemplary sequencing data analysis workflow is shown in FIG. 4.

**[00173]** In some aspects, determining the abundance and/or identify of specific transcripts in the plurality of double-stranded target nucleic acids can comprise determining the frequency of one or more mutations in a specific transcript in the plurality of double-stranded target nucleic acid. Mutations can include, but are not limited to one or more substitutions, one or more deletions, one or more insertions, one or more deletion-insertions, one or more duplications, one or more inversions, one or more repeat expansions or any combination thereof.

**[00174]** In some aspects of the preceding methods, the number of species of partially double-stranded identifier molecules in the plurality of partially double-stranded identifier molecules can be optimized to provide the appropriate sensitivity for the specific sequencing application. Without wishing to be bound by theory, in sequencing applications with the aim of identifying and/or quantifying rare mutations, the number of species can be increased to provide an increased number of possible barcode combinations.

**[00175]** The present disclosure provides methods for sequencing collections of double-stranded nucleic acid molecules using randomly paired adapter DNA constructs that together create combinatorial barcodes. Barcodes are used to identify and quantify individual variant molecules

within a complex DNA sample. The method comprises the steps: (a) affixing individual identifier molecules (containing discrete hemi-barcodes) to both ends of double stranded DNA fragments, while also affixing either an individual adapter molecule or an individual identifier adapter molecule onto the identifier molecule, to create a double stranded DNA fragment that contains a pair of identifier molecules and a pair of adapter molecules or a pair of identifier adapter molecules; (b) a single identifier molecule contains a sequence, that allows specific sticky-end ligation and is compatible with the adapter molecule or identifier adapter molecule. A second sequence that allows distinct, specific sticky-end ligation and is compatible with the target DNA fragments; (c) a single identifier molecule also contains a degenerate, semi-degenerate or discrete (non-degenerate) nucleic acid sequence which creates a relatively unique barcode; (d) The single adapter molecule contains a double stranded hybridized region, a sequence, that allows specific sticky-end ligation compatible with a sticky-end sequence on the identifier molecule, a single-stranded 5' arm, and a single stranded 3' arm, with further identifier molecules being affixed to the DNA-adapter fragment via amplification; (e) The single identifier adapter molecule contains a double stranded hybridized region, a sequence, that allows specific sticky-end ligation compatible with a sticky-end sequence on the identifier molecule, a single-stranded 5' arm with a single stranded identifier, and a single stranded 3' arm with a single stranded identifier.

**[00176]** The present disclosure provides a plurality of molecules is obtained by: (a) amplification of a single strand or both strands of the target DNA fragments prior to applying adapter molecules to the double stranded DNA targets; (b) amplification of a single strand or both strands of the target DNA-adapter product subsequent to applying adapter molecules to the double stranded DNA targets; (c) a combination of amplifications of either a single strand or both strands of the target DNA fragments prior to and/or subsequent to applying adapter molecules with index identifiers to the double stranded DNA targets; (d) Sequencing the amplified DNA-adapter products, thereby obtaining the association of each DNA target molecules with their corresponding barcodes to allow for downstream process such as error correction of barcodes, determining plurality of reads per barcode, determining of and correcting for errors associated with the sample preparation and sequencing of the target DNA molecules, determining true identities of target DNA sequences from potential false identities.

**[00177]** *Exemplary Embodiments*

**[00178]** Embodiment 1. A composition comprising:

a plurality of partially double-stranded identifier molecules; and  
a plurality of partially double-stranded adapter molecules.

**[00179]** Embodiment 2. The composition of embodiment 1, wherein the partially double-stranded identifier molecules comprise nucleic acid sequences about 11-20 nucleotides in length.

**[00180]** Embodiment 3. The composition of any of the preceding embodiments, wherein the partially double-stranded identifier molecules comprise at least one 5' overhang.

**[00181]** Embodiment 4. The composition of embodiment 3, wherein the partially double-stranded identifier molecules comprise two 5' overhangs.

**[00182]** Embodiment 5. The composition of embodiment 3, wherein the 5' overhang(s) is/are about 3 to about 5 nucleotides in length.

**[00183]** Embodiment 6. The composition of any one of embodiments 3-5, wherein at least one 5' overhang is capable of ligation to the partially double-stranded adapter molecules.

**[00184]** Embodiment 7. The composition of any one of embodiments 3-6, wherein at least one 5' overhang is capable of ligation to a target nucleic acid obtained from a biological sample.

**[00185]** Embodiment 8. The composition of any of the preceding embodiments, wherein the partially double-stranded adapter molecules comprise a double-stranded hybridized region.

**[00186]** Embodiment 9. The composition of any of the preceding embodiments, wherein the partially double-stranded adapter molecules comprise at least one overhang.

**[00187]** Embodiment 10. The composition of embodiment 10, wherein the overhang is capable of ligation to the partially double-stranded identifier molecules.

**[00188]** Embodiment 11. The composition of any of the preceding embodiments, wherein the partially double-stranded adapter molecules comprise a single-stranded 5' arm.

**[00189]** Embodiment 12. The composition of any of the preceding embodiments, wherein the partially double-stranded adapter molecules comprise a single-stranded 3' arm.

**[00190]** Embodiment 13. A kit comprising the composition of any of the preceding embodiments.

**[00191]** Embodiment 14. The kit of embodiment 13, further comprising a plurality of enzymes to mediate end-repair on double stranded DNA targets.

**[00192]** Embodiment 15. The kit of embodiment 13 or embodiment 14, further comprising a DNA ligase to mediate ligation of the adapter molecule or identifier adapter molecule and identifier molecule.

**[00193]** Embodiment 16. The kit of any one of embodiments 13-15, further comprising a set of primers suitable for the amplification of the DNA-adapter molecules.

**[00194]** Embodiment 17. The kit of any one of embodiments 13-16, further comprising a DNA polymerase to mediate the amplification of the DNA-adapter molecules.

**[00195]** Embodiment 18. The kit of any one of embodiments 13-17, further comprising reagents suitable for the purification of the end-repaired double stranded DNA targets and/or ligated DNA-adapter molecules and/or amplified DNA-adapter molecules.

**[00196]** Embodiment 19. The kit of any one of embodiments 13-18, further comprising buffers suitable to perform the appropriate enzymatic and purification steps.

**[00197]** Embodiment 20. The kit of any one of embodiments 13-19, further comprising written instructions.

**[00198]** Embodiment 21. A method for sequencing collections of double-stranded nucleic acid molecules using randomly paired adapter DNA constructs that together create combinatorial barcodes, wherein barcodes are used to identify and quantify individual variant molecules within a complex DNA sample, the method comprising:

a) affixing at least one partially double-stranded identifier molecule (containing discrete hemi-barcodes) to both ends of a target DNA fragment, wherein the identifier molecule comprises a discrete hemi-barcode,

b) affixing either at least one adapter molecule or identifier adapter molecule onto the identifier molecules, thereby producing a double stranded DNA fragment comprising a pair of identifier molecules and a pair of adapter molecules or identifier adapter molecules,

**[00199]** Embodiment 22. The method of embodiment 21, wherein the at least one identifier molecule comprises a degenerate, semi-degenerate or discrete (non-degenerate) nucleic acid sequence which creates a unique barcode.

**[00200]** Embodiment 23. The method of embodiment 21 or embodiment 22, wherein the at least one adapter molecule comprises a double stranded hybridized region, a sequence, that allows specific sticky-end ligation compatible with a sticky-end sequence on the at least one identifier molecule, a single-stranded 5' arm, and a single stranded 3' arm.

**[00201]** Embodiment 24. The method of any one of embodiments 21-23, the method further comprising affixing additional identifier molecules to the target DNA-adapter fragment via amplification.



**[00202]** Embodiment 25. The method of any one of embodiments 21-24, wherein the at least one identifier adapter molecule comprises a double stranded hybridized region, a sequence, which allows specific sticky-end ligation compatible with a sticky-end sequence on the at least one identifier molecule, a single-stranded 5' arm with a single stranded identifier, and a single stranded 3' arm with a single stranded identifier.

**[00203]** Embodiment 26. The method of any one of embodiments 21-25, the method further comprising amplifying a single strand or both strands of the target DNA fragments prior to applying adapter molecules to the double stranded DNA targets.

**[00204]** Embodiment 27. The method of any one of embodiments 21-26, the method further comprising amplifying a single strand or both strands of the target DNA-identifier product subsequent to applying adapter molecules to the double stranded DNA targets.

**[00205]** Embodiment 28. The method of any one of embodiments 21-27, the method further comprising sequencing the amplified DNA-adapter products, thereby obtaining the association of each DNA target molecules with their corresponding barcodes.

**[00206]** Embodiment 29. The method of embodiment 28, wherein association of each DNA molecules with their corresponding barcodes allow for at least one downstream process, wherein the downstream process is selected from error correction of barcodes, determining plurality of reads per barcode, determining of and correcting for errors associated with the sample preparation and sequencing of the target DNA molecules, determining true identities of target DNA sequences from potential false identities.

**[00207]** Embodiment 30. The method of any one of embodiments 21-29, wherein the at least one adapter molecule or identifier adapter molecule comprises a primer binding site.

**[00208]** Embodiment 31. The method of embodiment 30, wherein the primer binding site comprises a nucleotide sequence that permits for the linear or exponential amplification.

**[00209]** Embodiment 32. The method of any one of embodiments 21-30, wherein the at least one identifier molecule contains an error correctable, discrete hemi-barcodes.

**[00210]** Embodiment 33. A partially double-stranded identifier molecule comprising:  
a double-stranded region; and  
a first overhang.

**[00211]** Embodiment 34. The partially double-stranded identifier molecule of the embodiment 33, further comprising a second overhang.

**[00212]** Embodiment 35. The partially double-stranded identifier molecule of any one of embodiments 33-34, wherein the first and second overhangs are 5' overhangs.

**[00213]** Embodiment 36. The partially double-stranded identifier molecule of any one of embodiments 33-34, wherein the first and second overhangs are 3' overhangs.

**[00214]** Embodiment 37. The partially double-stranded identifier molecule of any one of embodiments 33-36, wherein the double-stranded region comprises an identifier sequence.

**[00215]** Embodiment 38. The partially double-stranded identifier molecule of embodiment 37, wherein the identifier sequence spans the entire double-stranded region.

**[00216]** Embodiment 39. The partially double-stranded identifier molecule of embodiment 37, wherein the identifier sequence spans a portion of the double-stranded region.

**[00217]** Embodiment 40. The partially double-stranded identifier molecule of any one of embodiments 37-39, wherein the identifier sequence is about 9 nucleotides in length.

**[00218]** Embodiment 41. The partially double-stranded identifier molecule of any one of embodiments 37-39, wherein the identifier sequence is about 10 nucleotides in length.

**[00219]** Embodiment 42. The partially double-stranded identifier molecule of any one of embodiments 37-39, wherein the identifier sequence is about 11 nucleotides in length.

**[00220]** Embodiment 43. The partially double-stranded identifier molecule of any one of embodiments 37-39, wherein the identifier sequence is about 12 nucleotides in length.

**[00221]** Embodiment 44. The partially double-stranded identifier molecule of any one of embodiments 37-39, wherein the identifier sequence is about 19 nucleotides in length.

**[00222]** Embodiment 45. The partially double-stranded identifier molecule of any one of embodiments 37-39, wherein the identifier sequence is about 20 nucleotides in length.

**[00223]** Embodiment 46. The partially double-stranded identifier molecule of any one of embodiments 37-39, wherein the identifier sequence is about 21 nucleotides in length.

**[00224]** Embodiment 47. The partially double-stranded identifier molecule of any one of embodiments 37-39, wherein the identifier sequence is about 22 nucleotides in length.

**[00225]** Embodiment 48. The partially double-stranded identifier molecule of any one of embodiments 33-47, wherein the first overhang is about 1 nucleotide in length.

**[00226]** Embodiment 49. The partially double-stranded identifier molecule of embodiment 48, wherein the first overhang is an adenine or a thymine.

**[00227]** Embodiment 50. The partially double-stranded identifier molecule of any one of embodiments 33-47, wherein the first overhang is about 2 nucleotides in length.

**[00228]** Embodiment 51. The partially double-stranded identifier molecule of any one of embodiments 33-47, wherein the first overhang is about 3 nucleotides in length.

**[00229]** Embodiment 52. The partially double-stranded identifier molecule of any one of embodiments 33-47, wherein the first overhang is about 4 nucleotides in length.

**[00230]** Embodiment 53. The partially double-stranded identifier molecule of any one of embodiments 33-47, wherein the first overhang is about 5 nucleotides in length.

**[00231]** Embodiment 54. The partially double-stranded identifier molecule of any one of embodiments 34-53, wherein the second overhang is about 1 nucleotide in length.

**[00232]** Embodiment 55. The partially double-stranded identifier molecule of embodiment 54, wherein the second overhang is an adenine or a thymine.

**[00233]** Embodiment 56. The partially double-stranded identifier molecule of any one of embodiments 34-53, wherein the second overhang is about 2 nucleotides in length.

**[00234]** Embodiment 57. The partially double-stranded identifier molecule of any one of embodiments 34-53, wherein the second overhang is about 3 nucleotides in length.

**[00235]** Embodiment 58. The partially double-stranded identifier molecule of any one of embodiments 34-53, wherein the second overhang is about 4 nucleotides in length.

**[00236]** Embodiment 59. The partially double-stranded identifier molecule of any one of embodiments 34-53, wherein the second overhang is about 5 nucleotides in length.

**[00237]** Embodiment 60. The partially double-stranded identifier molecule of any one of embodiments 33-59, wherein the partially double-stranded identifier molecule comprises DNA.

**[00238]** Embodiment 61. A plurality of the partially double-stranded identifier molecules of any one of embodiments 33-60, wherein the plurality comprises at least about 12 species of the partially double-stranded identifier molecules, wherein each species of partially double-stranded identifier molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00239]** Embodiment 62. The plurality of embodiment 61, wherein the plurality comprises at least about 24 species of the partially double-stranded identifier molecules.

**[00240]** Embodiment 63. The plurality of embodiment 62, wherein the plurality comprises at least about 48 species of the partially double-stranded identifier molecules.

**[00241]** Embodiment 64. The plurality of embodiment 63, wherein the plurality comprises at least about 96 species of the partially double-stranded identifier molecules.

**[00242]** Embodiment 65. The plurality of any one of embodiments 61-64, wherein the identifier sequence of one species of partially double-stranded identifier molecules will have a hamming distance of at least about two to any other identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

**[00243]** Embodiment 66. A partially double-stranded adapter molecule comprising:

a double-stranded region;

an overhang;

a single-stranded 5' arm; and

a single-stranded 3' arm.

**[00244]** Embodiment 67. The partially double-stranded adapter molecule of embodiment 66, wherein the overhang is a 5' overhang.

**[00245]** Embodiment 68. The partially double-stranded adapter molecule of embodiment 66, wherein the overhang is a 3' overhang.

**[00246]** Embodiment 69. The partially double-stranded adapter molecule of any one of embodiments 66-68, wherein the overhang is about 1 nucleotide in length.

**[00247]** Embodiment 70. The partially double-stranded adapter molecule of embodiment 69, wherein the overhang is an adenine or a thymine.

**[00248]** Embodiment 71. The partially double-stranded adapter molecule of any one of embodiments 66-68, wherein the overhang is about 2 nucleotides in length.

**[00249]** Embodiment 72. The partially double-stranded adapter molecule of any one of embodiments 66-68, wherein the overhang is about 3 nucleotides in length.

**[00250]** Embodiment 73. The partially double-stranded adapter molecule of any one of embodiments 66-68, wherein the overhang is about 4 nucleotides in length.

**[00251]** Embodiment 74. The partially double-stranded adapter molecule of any one of embodiments 66-68, wherein the overhang is about 5 nucleotides in length.

**[00252]** Embodiment 75. The partially double-stranded adapter molecule of any one of embodiments 66-74, wherein the double-stranded region comprises an identifier sequence.

**[00253]** Embodiment 76. The partially double-stranded adapter molecule of embodiment 75, wherein the identifier sequence is about 9 nucleotides in length.

**[00254]** Embodiment 77. The partially double-stranded adapter molecule of embodiment 75, wherein the identifier sequence is about 10 nucleotides in length.

**[00255]** Embodiment 78. The partially double-stranded adapter molecule of embodiment 75, wherein the identifier sequence is about 11 nucleotides in length.

**[00256]** Embodiment 79. The partially double-stranded adapter molecule of embodiment 75, wherein the identifier sequence is about 12 nucleotides in length.

**[00257]** Embodiment 80. The partially double-stranded adapter molecule of embodiment 75, wherein the identifier sequence is about 19 nucleotides in length.

**[00258]** Embodiment 81. The partially double-stranded adapter molecule of embodiment 75, wherein the identifier sequence is about 20 nucleotides in length.

**[00259]** Embodiment 82. The partially double-stranded adapter molecule of embodiment 75, wherein the identifier sequence is about 21 nucleotides in length.

**[00260]** Embodiment 83. The partially double-stranded adapter molecule of embodiment 75, wherein the identifier sequence is about 22 nucleotides in length.

**[00261]** Embodiment 84. The partially double-stranded adapter molecule of any one of embodiments 66-83, wherein the single-stranded 5' arm comprises at least one amplification primer binding site.

**[00262]** Embodiment 85. The partially double-stranded adapter molecule of any one of embodiments 66-84, wherein the single-stranded 3' arm comprises at least one amplification primer binding site.

**[00263]** Embodiment 86. The partially double-stranded adapter molecule of any one of embodiments 66-85, wherein the partially double-stranded adapter molecule comprises DNA.

**[00264]** Embodiment 87. A plurality of the partially double-stranded adapter molecules of any one of embodiments 66-85, wherein the plurality comprises at least about 12 species of the partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter molecules in the plurality.

**[00265]** Embodiment 88. The plurality of embodiment 87, wherein the plurality comprises at least about 24 species of the partially double-stranded adapter molecules.

**[00266]** Embodiment 89. The plurality of embodiment 88, wherein the plurality comprises at least about 48 species of the partially double-stranded adapter molecules.

[00267] Embodiment 90. The plurality of embodiment 89, wherein the plurality comprises at least about 96 species of the partially double-stranded adapter molecules.

[00268] Embodiment 91. The plurality of any one of embodiments 87-90, wherein the identifier sequence of one species of partially double-stranded identifier molecules have a hamming distance of at least about two to any other identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.

[00269] Embodiment 92. A kit comprising the plurality of any one of embodiments 61-65.

[00270] Embodiment 93. The kit of embodiment 92, further comprising the plurality of any one of embodiments 87-91.

[00271] Embodiment 94. The kit of embodiment 92 or 93, further comprising a plurality of enzymes to mediate end-repair on double-stranded.

[00272] Embodiment 95. The kit of any one of embodiments 92-94, further comprising a plurality of reagents for the purification of nucleic acid molecules.

[00273] Embodiment 96. The kit of any one of embodiments 92-95, further comprising at least one DNA polymerase.

[00274] Embodiment 97. The kit of any one of embodiments 92-96, further comprising a plurality of amplification primers.

[00275] Embodiment 98. The kit of embodiment 97, wherein the amplification primers in the plurality bind to the amplification primer binding sites present in the partially double-stranded adapter molecules.

[00276] Embodiment 99. The kit of any one of embodiments 92-98, further comprising at least one DNA ligase.

[00277] 100. The kit of any one of embodiments 92-99, further comprising written instructions for the performance of one of the methods of the present disclosure.

[00278] Embodiment 101. A method of sequencing a plurality of double-stranded target nucleic acids comprising:

a) contacting the plurality of double-stranded target nucleic acids with the plurality of partially double-stranded identifier molecules of any one of embodiments 61-65 and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids;

b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and

c) sequencing the products of step (b).

**[00279]** Embodiment 102. A method of sequencing a plurality of double-stranded target nucleic acids comprising:

a) contacting the plurality of double-stranded target nucleic acids with the plurality of partially double-stranded identifier molecules of any one of embodiments 61-65 and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the ligation products comprise each of the combinations of two species of partially double-stranded identifier molecules;

b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and

c) sequencing the products of step (b).

**[00280]** Embodiment 103. A method of sequencing a plurality of double-stranded target nucleic acids comprising:

a) contacting the plurality of double-stranded target nucleic acids with the plurality of partially double-stranded identifier molecules of any one of embodiments 61-65 and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids, wherein the ligation products comprise at least 10% of the combinations of two species of partially double-stranded identifier molecules;

b) contacting the products of step (a) with a plurality of partially double-stranded adapter molecules of the present disclosure and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and

c) sequencing the products of step (b).

**[00281]** Embodiment 104. The method of embodiment 103, wherein the ligation products in step (a) comprise at least 20% of the combinations of two species of partially double-stranded identifier molecules.

**[00282]** Embodiment 105. The method of embodiment 104, wherein the ligation products in step (a) comprise at least 30% of the combinations of two species of partially double-stranded identifier molecules.

**[00283]** Embodiment 106. The method of embodiment 105, wherein the ligation products in step (a) comprise at least 40% of the combinations of two species of partially double-stranded identifier molecules.

**[00284]** Embodiment 107. The method of embodiment 106, wherein the ligation products in step (a) comprise at least 50% of the combinations of two species of partially double-stranded identifier molecules.

**[00285]** Embodiment 108. The method of embodiment 107, wherein the ligation products in step (a) comprise at least 60% of the combinations of two species of partially double-stranded identifier molecules.

**[00286]** Embodiment 109. The method of embodiment 108, wherein the ligation products in step (a) comprise at least 70% of the combinations of two species of partially double-stranded identifier molecules.

**[00287]** Embodiment 110. The method of embodiment 109, wherein the ligation products in step (a) comprise at least 80% of the combinations of two species of partially double-stranded identifier molecules.

**[00288]** Embodiment 111. The method of embodiment 110, wherein the ligation products in step (a) comprise at least 90% of the combinations of two species of partially double-stranded identifier molecules.

**[00289]** Embodiment 112. The method of any one of embodiments 101-111, the method further comprising after step (b) and prior to step (c), constructing a sequencing library using the products of step (b).

**[00290]** Embodiment 113. The method of any one of embodiments 101-112, the method further comprising after step (b) and prior to step (c), amplifying the products of step (b).

**[00291]** Embodiment 114. The method of embodiment 113, wherein amplifying the products of step (b) comprises contacting the products of step b with amplification primers that bind to amplification primer binding sites in the partially double-stranded adapter molecules and at least one polymerase.



**[00292]** Embodiment 115. The method of any one of embodiments 101-114, wherein the method further comprising determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c).

**[00293]** Embodiment 116. The method of embodiment 115, wherein determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises correcting for errors using the identifier sequences of the ligated partially double-stranded identifier molecules.

**[00294]** Embodiment 117. The method of embodiment 116, wherein the errors comprise amplification errors, sample preparation errors, sequencing errors or any combination thereof.

**[00295]** Embodiment 118. The method of any one of embodiments 115-117, wherein determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises creating consensus sequences using identifier sequences of the ligated partially double-stranded identifier molecules.

**[00296]** Embodiment 119. The method of any one of embodiments 115-118, wherein determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises grouping the sequencing reads obtained in step (c) by the ligated identifier sequences in the sequencing reads.

**[00297]** Embodiment 120. The method of any one of embodiments 115-119, wherein determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequence data generated in step (c) comprises grouping the sequencing reads obtained in step (c) by the specific genomic sequence that the sequencing reads most likely correspond to.

**[00298]** Embodiment 121. The method of any one of embodiments 115-120, wherein determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises first grouping the sequencing reads obtained in step (c) by the ligated identifier sequences in the sequencing reads and then further grouping by the specific genomic sequence that the sequencing reads most likely correspond to.

[00299] Embodiment 122. The method of any one of embodiments 115-121, wherein determining the abundance and/or identify of specific transcripts in the plurality of double-stranded target nucleic acids can comprise determining the frequency of one or more mutations in a specific transcript in the plurality of double-stranded target nucleic acid.

[00300] Embodiment 123. The method of embodiment 122, wherein the one or more mutations comprise one or more insertions, one or more deletion-insertions, one or more duplications, one or more inversions, one or more repeat expansions or any combination thereof.

[00301] Embodiment 124. The method of any one of embodiments 101-123, wherein the number of species of partially double-stranded identifier molecules in the plurality is selected such that there is on average at least about two sequencing reads for each UMI that is measured.

[00302] Embodiment 125. The method of any one of embodiments 101-124, wherein the number of species of partially double-stranded identifier molecules in the plurality is selected such that there at least about two sequencing reads for each UMI that is measured.

[00303] *Examples*

[00304] Example 1—ligation of partially double-stranded identifier molecules and partially double-stranded adapter molecules of the present disclosure.

[00305] The following is a non-limiting example of the ligation of partially double-stranded identifier molecules with varying overhang sizes and partially double-stranded adapter molecules of the present disclosure with varying overhang sizes and varying sizes of the double-strand regions to target nucleic acids using the methods of the present disclosure.

[00306] The double-stranded sequences of the partially double-stranded identifier molecules and partially double-stranded adapter molecules are shown below and in FIG. 5 and FIG. 6

[00307] **Partially double-stranded adapter molecules with varying overhang sizes**

Uni_Short	TACTACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 193)
IDX_1bp_Overhang	CACTGACCTCAAGTCTGCACACGAGAAGGCTAG (SEQ ID NO: 194)
Uni_Short	TACTACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 193)
IDX_2bp_Overhang	CACTGACCTCAAGTCTGCACACGAGAAGGCTA (SEQ ID NO: 195)
Uni_Short	TACTACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 193)
IDX_3bp_Overhang	CACTGACCTCAAGTCTGCACACGAGAAGGCT (SEQ ID NO: 196)

```

Uni_Short          TACTACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 193)
                    |||
IDX_4bp_Overhang  CACTGACCTCAAGTCTGCACACGAGAAGGC (SEQ ID NO: 197)

Uni_Short          TACTACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 193)
                    |||
IDX_5bp_Overhang  CACTGACCTCAAGTCTGCACACGAGAAGG (SEQ ID NO: 198)
    
```

**[00308] Partially double-stranded identifier molecules with varying overhang sizes**

```

BC01a_Overhang      ATCGAGTGCACAT (SEQ ID NO: 199)
BC01b_1bp_Overhang  ATAGCTCACGTGT (SEQ ID NO: 200)

BC01a_Overhang      ATCGAGTGCACAT (SEQ ID NO: 199)
BC01b_2bp_Overhang  GATAGCTCACGTGT (SEQ ID NO: 201)

BC01a_Overhang      ATCGAGTGCACAT (SEQ ID NO: 199)
BC01b_3bp_Overhang  AGATAGCTCACGTGT (SEQ ID NO: 202)

BC01a_Overhang      ATCGAGTGCACAT (SEQ ID NO: 199)
BC01b_4bp_Overhang  TAGATAGCTCACGTGT (SEQ ID NO: 203)

BC01a_Overhang      ATCGAGTGCACAT (SEQ ID NO: 199)
BC01b_5bp_Overhang  CTAGATAGCTCACGTGT (SEQ ID NO: 204)
    
```

**[00309] Primers for 300 bp PCR Product**

```

pUC19_300bp_Frw_primer  CGAGGGAGCTTCCAGGGGAAAC (SEQ ID NO: 205)
pUC19_300bp_Rev_primer  TTGGGCGCTCTTCCGCTTCCTC (SEQ ID NO: 206)
    
```

**[00310] Partially double-stranded adapter molecules with varying sizes of the double-stranded region and corresponding partially double-stranded identifier molecules**

```

12bp_Stem_A-T_UNI  TACTACTCTTTCCCTACACGACGCTCTTCCGATC (SEQ ID NO: 207)
                    |||
12bp_Stem_A-T_IDX  CACTGACCTCAAGTCTGCACACGAGAAGGCTAGA (SEQ ID NO: 208)

12bp_Stem_A-T_BC01a      TATCGAGTGCACAT (SEQ ID NO: 209)
                    |||
12bp_Stem_A-T_BC01b      TAGCTCACGTGT (SEQ ID NO: 210)

11-bp Stem_C-G_UNI  TACTACTCTTTCCCTACACGACGCTCTTCCGATC (SEQ ID NO: 211)
                    |||
11-bp Stem_C-G_IDX  CACTGACCTCAAGTCTGCACACGAGAAGGCTA (SEQ ID NO: 212)

11-bp Stem_C-G_BC01a      TATCGAGTGCACAT (SEQ ID NO: 209)
                    |||
11-bp Stem_C-G_BC01b      GATAGCTCACGTGT (SEQ ID NO: 213)

11-bp Stem_G-C_UNI  TACTACTCTTTCCCTACACGACGCTCTTCCGAT (SEQ ID NO: 214)
                    |||
    
```

```

11-bp Stem_G-C_IDX  CACTGACCTCAAGTCTGCACACGAGAAGGCTAG (SEQ ID NO: 215)
11-bp Stem_G-C_BC01a      CTATCGAGTGCACAT (SEQ ID NO: 216)
                        |||
11-bp Stem_G-C_BC01b      ATAGCTCACGTGT (SEQ ID NO: 217)
10-bp Stem_A-T_UNI  TACTACTCTTTCCCTACACGACGCTCTTCCGA (SEQ ID NO: 218)
                        |||
10-bp Stem_A-T_IDX  CACTGACCTCAAGTCTGCACACGAGAAGGCTA (SEQ ID NO: 219)
10-bp Stem_A-T_BC01a      TCTATCGAGTGCACAT (SEQ ID NO: 220)
                        |||
10-bp Stem_A-T_BC01b      GATAGCTCACGTGT (SEQ ID NO: 221)
9-bp Stem_A-T_UNI  TACTACTCTTTCCCTACACGACGCTCTTCCGA (SEQ ID NO: 222)
                        |||
9-bp Stem_A-T_IDX  CACTGACCTCAAGTCTGCACACGAGAAGGC (SEQ ID NO: 223)
9-bp Stem_A-T_BC01a      TCTATCGAGTGCACAT (SEQ ID NO: 224)
                        |||
9-bp Stem_A-T_BC01b      TAGATAGCTCACGTGT (SEQ ID NO: 225)
    
```

**[00311]** The ligation reactions were performed as followed:

**[00312]** Amplification and End-Repair of a 300 bp PCR Product from pUC19

- 1) Q5® High-Fidelity DNA Polymerase (NEB M0491) - Used standard manufacturer's protocol.
- 2) NEBNext® Ultra™ II End Repair/dA-Tailing Module (NEB E7546) - Used standard manufacturer's protocol.
- 3) dA-tailed pcr products were purified using the standard 2X SPRI beads protocol and quantified by QuBit.

**[00313]** Phosphorylation & Annealing of Barcodes

- 1) 250 pMols of each single stranded barcode was phosphoylated at the 5'-terminus using 0.5 U/μL of T4 PNK (NEB) for 30 min @ 37 °C. In 1X T4 DNA ligase buffer (NEB)
- 2) The barcodes we paired together before denaturing for 2 min @ 80 °C then annealing by cooling to 4 °C and incubating for 10 mins

**[00314]** Adapter/Barcode Ligation reaction

- 1) Barcode and adapter pairs were diluted to 30 μM in 10 mM Tris-HCl and 50 mM NaCl (pH 8.0), before denaturing for 2 min @ 80 °C then annealing by cooling to 4 °C and incubating for 10 mins.

- 2) Ligation of the 300 bp PCR (50 ng) product from pUC19, with both the barcode (0.25  $\mu$ M) and adapter (0.25  $\mu$ M) sequences was achieved in 1x T4 DNA ligase Buffer (NEB), 40 U/ $\mu$ L of T4 DNA ligase (NEB) and incubated for 30 min @ 21 °C.
- 3) The Ligation product was purified using the standard 1X SPRI beads protocol.
- 4) Indexing PCR was performed with the NEBNext Multiplex Oligos for Illumina kit using the manufacturer's protocol.
- 5) The indexed PCR product was purified using the standard 1X SPRI beads protocol, before visualizing on an agarose gel.

**[00315]** The agarose gel analysis of the ligation reactions described above are shown in FIG. 5 and FIG. 6. As shown in FIG. 5 and FIG. 6, the partially double-stranded identifier molecules and partially double-stranded adapter molecules can be efficiently ligated to target nucleic acids in the sequencing methods of the present disclosure.

**[00316]** Example 2—sequencing genomic regions of interest using the sequencing methods of the present disclosure

**[00317]** The following is a non-limiting example of using the compositions and methods of the present disclosure to sequence a plurality of double-stranded target nucleic acid molecules. More specifically, regions of interest were amplified from genomic DNA and analyzed using the compositions and methods of the present disclosure, as well as existing NGS methods, to compare the results of both methods.

**[00318]** Amplification of the Regions of Interest

- 1) Region of interests were amplified from 5 ng of gDNA (Quantitative Multiplex Reference Standard, Horizon Discovery) using multiplex AmpliSeq PCR primers (0.5  $\mu$ M), 1X Q5 Reaction Buffer (NEB), 1x Taq Buffer (NEB), 0.2 mM dNTPs (NEB) 1.0U Q5 Polymerase (NEB) and 1.25 U Taq polymerase (NEB). The PCR mixture was amplified for 2 min at 98 °C, then 30 cycles of 30s at 98 °C, 90s at 60 °C and 30s at 72 °C and final 5 min at 72°C.
- 2) The PCR products were purified using the standard 3X SPRI beads protocol.

**[00319]** Adapter/Barcode Ligation Reaction and Unique Dual Indexing

- 1) Ligation of the PCR product (50 ng), with both the barcode (0.25  $\mu$ M) and adapter (0.25  $\mu$ M) sequences was achieved in 1x T4 DNA ligase Buffer (NEB), 40 U/ $\mu$ L of T4 DNA ligase (NEB) and incubated for 30 min @ 21 °C.

- 2) The Ligation product was purified using the standard 1X SPRI beads protocol.
- 3) Indexing PCR was performed with the NEBNext Multiplex Oligos for Illumina kit using the manufacturer's protocol.
- 4) The indexed PCR product was purified using the standard 1X SPRI beads protocol, before visualizing on an agarose gel and sequenced

**[00320]** Generation of Duplex Consensus Sequence BAM files, starting with Paired-end Fastq files

- 1) Fastq files generated from Next-generation Sequences were processed according to the workflow shown in FIG. 4, including the CorrectUmis and GroupBySeq steps:
  - a) CorrectUmis - The partially double stranded identifier molecules and adapter molecules used to generate UMIs (*i.e.* the specific combination of the two species of partially double-stranded identifier molecules that are ligated to a single transcript), can further be used for error correction (once assembled libraries have been sequenced), due to the fact that the UMIs created are discrete with a defined distance between each UMI (*see* FIG 7.). Error-correction is not normally applied to UMIs generated from a pre-pooled, degenerate method due to there being no control over the distance between UMIs.
  - b) GroupBySeq – Using EGFR4 as an example and focusing on the G>A variant present at an allelic frequency of 24.5%, single stranded consensus sequences (SSCS) were generated using CallMolecularConsensus (fgbios tools) and visualized via IGV (Broad Institute). Overall, base calling at the known G>A variant position, was of a lower quality for SSCS generated without the GroupBySeq step (top), than compared to SSCS generated with the GroupBySeq step.

Without wishing to be bound by theory, for existing pre-pooled, degenerate UMIs, the number of UMIs available should be larger than the number of molecules present within the initial sample. This ensures each molecule gets a unique UMI. This approach leads to a large majority of UMIs containing only a single read. With a minimum requirement of at least two reads per UMI, to generate a consensus sequence, the UMIs containing only a single read are discarded. The inability to produce a consensus read for a large majority of the

available UMIs, means that very high sequencing depths are required for each region of interest. These limitations are overcome by the methods of the present disclosure by:-

- 1) giving the user the ability to modulate the number of discrete UMIs used for the initial sample;
- 2) using the sequence from the region of interest as a UMI itself. Aligned reads can be grouped together based on their absolute alignment to the reference using GroupBySeq, within GroupBySeq reads are grouped based on their similarity/difference from the reference. The GroupBySeq reads can then further be sub-divided by their UMIs. Because the number of initial UMIs can be modulated, the number of reads per UMI can be modulated to have on average at least two reads per UMI (once the GroupBySeq step has been performed). Optimization of the number of reads per UMI, allows the majority of reads (and therefore UMIs) to produce usable consensus reads, therefore reducing the coverage required per region of interest.

**[00321]** The sequencing libraries created using the sequencing methods of the present disclosure were analyzed using a Bioanalyzer system. The results of the analysis are shown in FIG. 9. As shown in FIG. 9, minimal adapter dimers were observed in the 100-150 bp region.

**[00322]** The analysis of the sequencing results for specific mutations in 11 genes are shown in FIGs. 10-20, including the results using existing NGS methods (top panel) and the results using the sequencing methods of the present disclosure (denoted gSynth Duplex Sequencing in FIGs. 10-20). FIGs. 10-20 also show the expected allelic fraction of the mutation that is being analyzed, and the number of different UMIs (barcodes) that are possible based on the number of species of partially double-stranded identifier molecules that were used in the sequence (*e.g.* 12 species yield 144 possible barcodes, 24 species yield 576 possible barcodes, 48 species yield 2,304 possible barcodes, etc.).

**[00323]** FIG. 10 shows the sequencing results for the EGFR4 gene and the measured mutant frequencies for a DNA base change of GGC→AGC.

**[00324]** FIG. 11 shows the sequencing results for the PI3KCA10 gene and the measured mutant frequencies for a DNA base change of CAT→CGT.

[00325] FIG. 12 shows the sequencing results for the KRAS1 gene and the measured mutant frequencies for a DNA base change of GGC→GAC.

[00326] FIG. 13 shows the sequencing results for the NRAS gene and the measured mutant frequencies for a DNA base change of CAA→AAA.

[00327] FIG. 14 shows the sequencing results for the BRAF gene and the measured mutant frequencies for a DNA base change of CTG→CAG.

[00328] FIG. 15 shows the sequencing results for the KIT gene and the measured mutant frequencies for a DNA base change of GAC→GTC.

[00329] FIG. 16 shows the sequencing results for the PI3KCA7 gene and the measured mutant frequencies for a DNA base change of GAG→AAG.

[00330] FIG. 17 shows the sequencing results for the KRAS1 gene and the measured mutant frequencies for a DNA base change of GGT→GAT.

[00331] FIG. 18 shows the sequencing results for the EGFR8 gene and the measured mutant frequencies for a DNA base change of CTG→CGG.

[00332] FIG. 19 shows the sequencing results for the EGFR5 gene and the measured mutant frequencies for a DNA base change of AAGGAATTAAGAGAAGCA→AA.

[00333] FIG. 20 shows the sequencing results for the EGFR6 gene and the measured mutant frequencies for a DNA base change of ACG→ATG.

[00334] As shown in FIGs. 10-20, the sequencing results obtained by the methods of the present disclosure, and more specifically the mutation frequency measured using the sequencing methods of the present disclosure was more accurate as compared to the results obtained using existing NGS methods. Moreover, the sequencing results obtained by the methods of the present disclosure exhibited less noise as compared to the sequencing results obtained by existing NGS methods. Accordingly, the results presented in this example demonstrate that the sequencing compositions and methods of present disclosure provide superior sequencing results, including mutation frequency measurements, as compared to existing NGS methods.

### EQUIVALENTS

[00335] The foregoing description has been presented only for the purposes of illustration and is not intended to limit the disclosure to the precise form disclosed. The details of one or more embodiments of the disclosure are set forth in the accompanying description above. Although any methods and materials similar or equivalent to those described herein can be used in the



practice or testing of the present disclosure, the preferred methods and materials are now described. Other features, objects, and advantages of the disclosure will be apparent from the description and from the claims. In the specification and the appended claims, the singular forms include plural referents unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. All patents and publications cited in this specification are incorporated by reference.

What is claimed is:

1. A plurality of partially double-stranded identifier molecules,
  - wherein the partially double-stranded identifier molecules comprise:
    - a double-stranded region comprising an identifier sequence; and
    - a first overhang;
  - wherein the plurality comprises at least about 12 species of the partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter molecules in the plurality,
    - wherein the identifier sequence of one species of partially double-stranded identifier molecules will have a hamming distance of at least about two to any other identifier sequence of any other species of partially double-stranded identifier molecules in the plurality.
2. The plurality of partially double-stranded identifier molecules of claim 1, wherein the partially double-stranded identifier molecules further comprise a second overhang.
3. The plurality of partially double-stranded identifier molecules of any of the preceding claims, wherein the first and/or second overhangs are
  - a) 5' overhangs; or
  - b) 3' overhangs.
4. The plurality of partially double-stranded identifier molecules of any of the preceding claims, wherein:
  - a) the identifier sequence spans the entire double-stranded region; or
  - b) the identifier sequence spans a portion of the double-stranded region.
5. The plurality of partially double-stranded identifier molecules of any of the preceding claims, wherein the identifier sequence is:
  - a) about 9 nucleotides in length;
  - b) about 10 nucleotides in length;
  - c) about 11 nucleotides in length;

- d) about 12 nucleotides in length;
  - e) about 19 nucleotides in length;
  - f) about 20 nucleotides in length;
  - g) about 21 nucleotides in length; or
  - h) about 22 nucleotides in length.
6. The plurality of partially double-stranded identifier molecules of any of the preceding claims, wherein the first overhang and/or the second overhang is about 1 nucleotide in length, preferably wherein the first overhang and/or the second overhang is:
- a) an adenine or a thymine; or
  - b) a guanosine or a cytosine.
7. The plurality of partially double-stranded identifier molecules of any of the preceding claims, wherein the first overhang and/or the second overhang is:
- a) about 2 nucleotides in length;
  - b) about 3 nucleotides in length;
  - c) about 4 nucleotides in length; or
  - d) about 5 nucleotides in length.
8. The plurality of partially double-stranded identifier molecules of any of the preceding claims, wherein the partially double-stranded identifier molecules comprise DNA.
9. The plurality of partially double-stranded identifier molecules of any of the preceding claims, wherein the plurality comprises:
- a) at least about 24 species of the partially double-stranded identifier molecules;
  - b) at least about 48 species of the partially double-stranded identifier molecules; or
  - c) at least about 96 species of the partially double-stranded identifier molecules.
10. A plurality of partially double-stranded adapter molecules,
- wherein the partially double-stranded adapter molecules comprise:
- a double-stranded region;

- an overhang;
- a single-stranded 5' arm; and
- a single-stranded 3' arm;

wherein the single-stranded 5' arm comprises at least one amplification primer binding site and the single-stranded 3' arm comprises at least one amplification primer binding site.

11. The plurality of partially double-stranded adapter molecules of claim 10, wherein the double-stranded region comprises an identifier sequence,

wherein the plurality comprises at least about 12 species of the partially double-stranded adapter molecules, wherein each species of partially double-stranded adapter molecules comprises an identifier sequence that is different from the identifier sequence of any other species of partially double-stranded adapter molecules in the plurality.

12. The plurality of partially double-stranded adapter molecules of any one of claims 10-11, wherein the overhang is:

- a) a 5' overhang; or
- b) a 3' overhang.

13. The plurality of partially double-stranded adapter molecules of any one of claims 10-12, wherein the overhang is about 1 nucleotide in length, preferably wherein the overhang is:

- a) an adenine or a thymine; or
- b) a guanosine or cytosine.

14. The plurality of partially double-stranded adapter molecules of any one of claims 10-12, wherein the overhang is:

- a) about 2 nucleotides in length;
- b) about 3 nucleotides in length;
- c) about 4 nucleotides in length;
- d) about 5 nucleotides in length.

15. The plurality of partially double-stranded adapter molecules of any one of claims 11-14, wherein the identifier sequence is

- a) about 9 nucleotides in length;
- b) about 10 nucleotides in length;
- c) about 11 nucleotides in length;
- d) about 12 nucleotides in length;
- e) about 19 nucleotides in length;
- f) about 20 nucleotides in length;
- g) about 21 nucleotides in length; or
- h) about 22 nucleotides in length.

16. The plurality of partially double-stranded adapter molecules of any one of claims 10-15, wherein the partially double-stranded adapter molecules comprise DNA.

17. A method of sequencing a plurality of double-stranded target nucleic acids comprising:

a) contacting the plurality of double-stranded target nucleic acids with the plurality of partially double-stranded identifier molecules of any one of claims 1-9 and at least one ligase such that a partially double-stranded identifier molecule is ligated to each end of the double-stranded target nucleic acids in the plurality of double-stranded target nucleic acids;

b) contacting the products of step (a) with the plurality of partially double-stranded adapter molecules of any one of claims 10-16 and at least one ligase such that a partially double-stranded adapter molecule is ligated to each end of the products of step (a); and

c) sequencing the products of step (b).

18. The method of claim 17, wherein the ligation products in step (a) comprise:

a) at least 10% of the combinations of two species of partially double-stranded identifier molecules;

b) at least 20% of the combinations of two species of partially double-stranded identifier molecules;

c) at least 30% of the combinations of two species of partially double-stranded identifier molecules;

d) at least 40% of the combinations of two species of partially double-stranded identifier molecules;

e) at least 50% of the combinations of two species of partially double-stranded identifier molecules;

f) at least 60% of the combinations of two species of partially double-stranded identifier molecules;

g) at least 70% of the combinations of two species of partially double-stranded identifier molecules;

h) at least 80% of the combinations of two species of partially double-stranded identifier molecules;

i) at least 90% of the combinations of two species of partially double-stranded identifier molecules; or

j) each of the combinations of two species of partially double-stranded identifier molecules.

19. The method of claim 17 or claim 18, the method further comprising after step (b) and prior to step (c), constructing a sequencing library using the products of step (b).

20. The method of any one of claims 17-19, wherein step (a) and step (b):

a) are performed sequentially; or

b) are performed concurrently.

21. The method of any one of claims 17-20, the method further comprising after step (b) and prior to step (c), amplifying the products of step (b), preferably wherein amplifying the products of step (b) comprises contacting the products of step (b) with amplification primers that bind to amplification primer binding sites in the partially double-stranded adapter molecules and at least one polymerase.

22. The method of any one of claims 17-21, wherein the method further comprising determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c), preferably wherein

determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises correcting for errors using the identifier sequences of the ligated partially double-stranded identifier molecules, preferably wherein the errors comprise amplification errors, sample preparation errors, sequencing errors or any combination thereof; and/or

determining the abundance and/or the identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises creating consensus sequences using identifier sequences of the ligated partially double-stranded identifier molecules.

23. The method of claim 22, wherein determining the abundance and/or identity of specific transcripts in the plurality of double-stranded target nucleic acids using the sequencing data generated in step (c) comprises

a) grouping the sequencing reads obtained in step (c) by the ligated identifier sequences in the sequencing reads;

b) grouping the sequencing reads obtained in step (c) by the specific genomic sequence that the sequencing reads most likely correspond to; or

c) any combination thereof.

24. The method of claim 22 or claim 23, wherein determining the abundance and/or identify of specific transcripts in the plurality of double-stranded target nucleic acids can comprise determining the frequency of one or more mutations in a specific transcript in the plurality of double-stranded target nucleic acid, preferably wherein the one or more mutations comprise one or more insertions, one or more deletion-insertions, one or more duplications, one or more inversions, one or more repeat expansions or any combination thereof.

25. A kit comprising the plurality of partially double-stranded identifier molecules of any one of claims 1-9.

26. The kit of claim 25, wherein the kit further comprises the plurality of partially double-stranded adapter molecules of any one of claims 10-16.



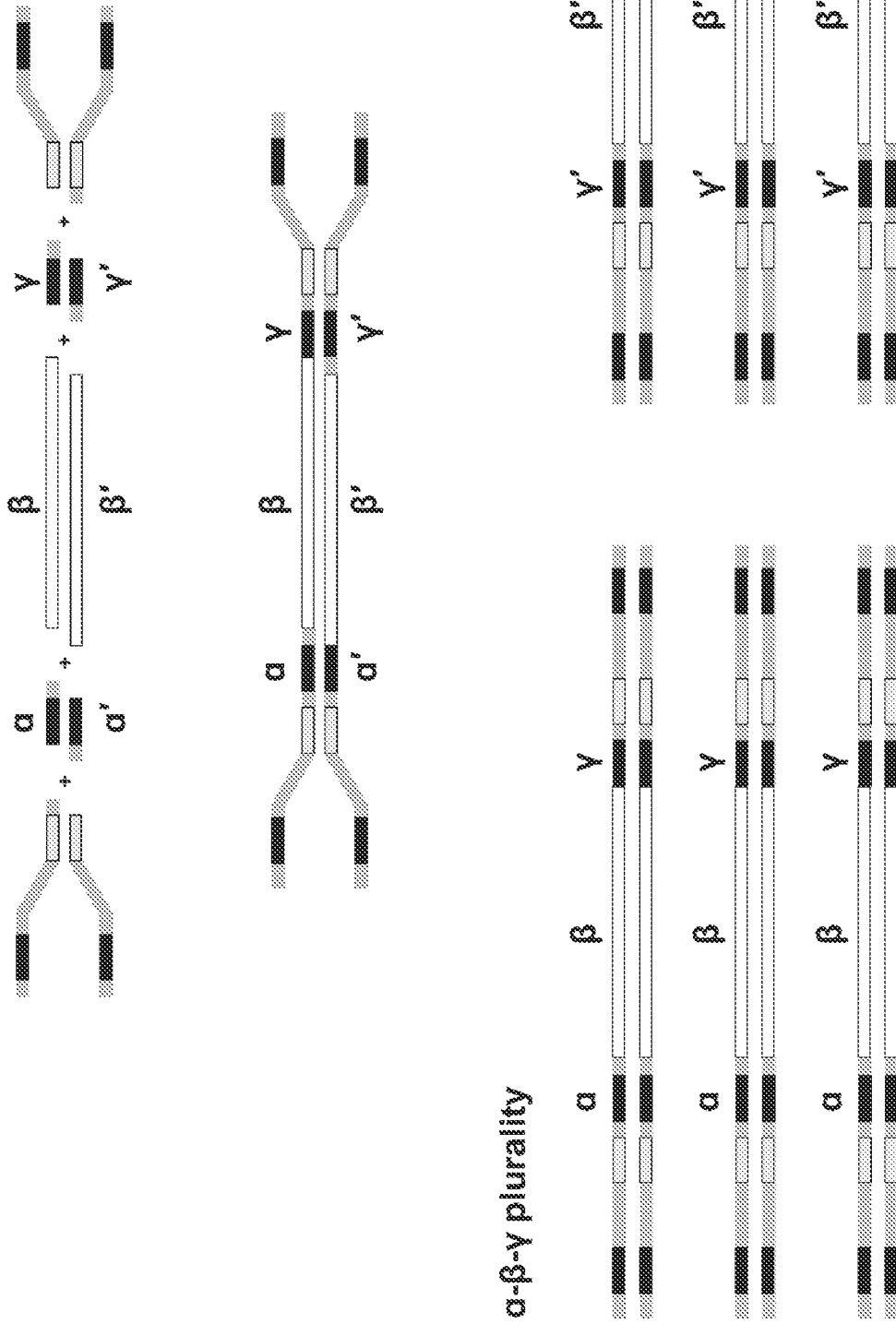


FIG. 1

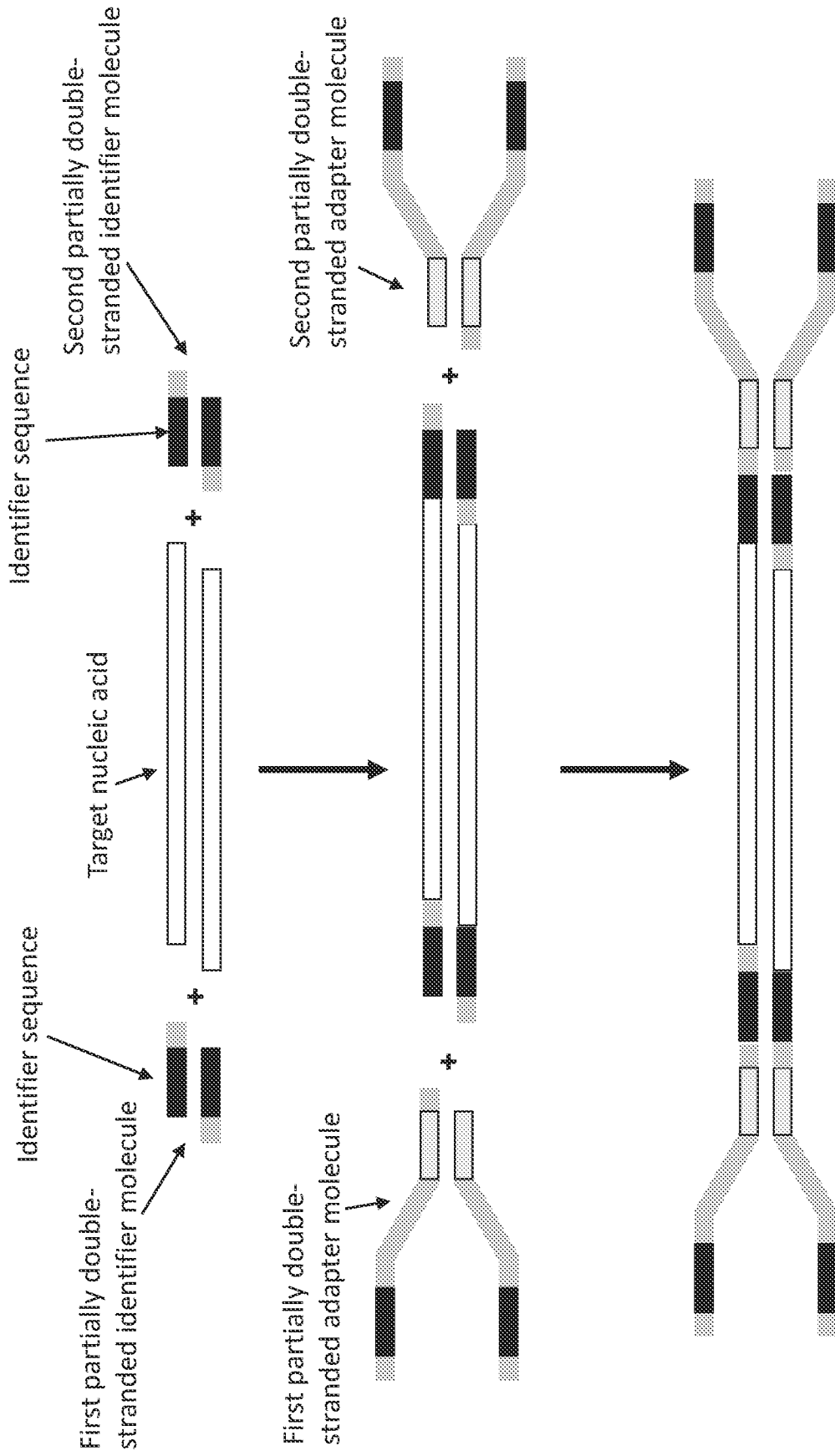


FIG. 2

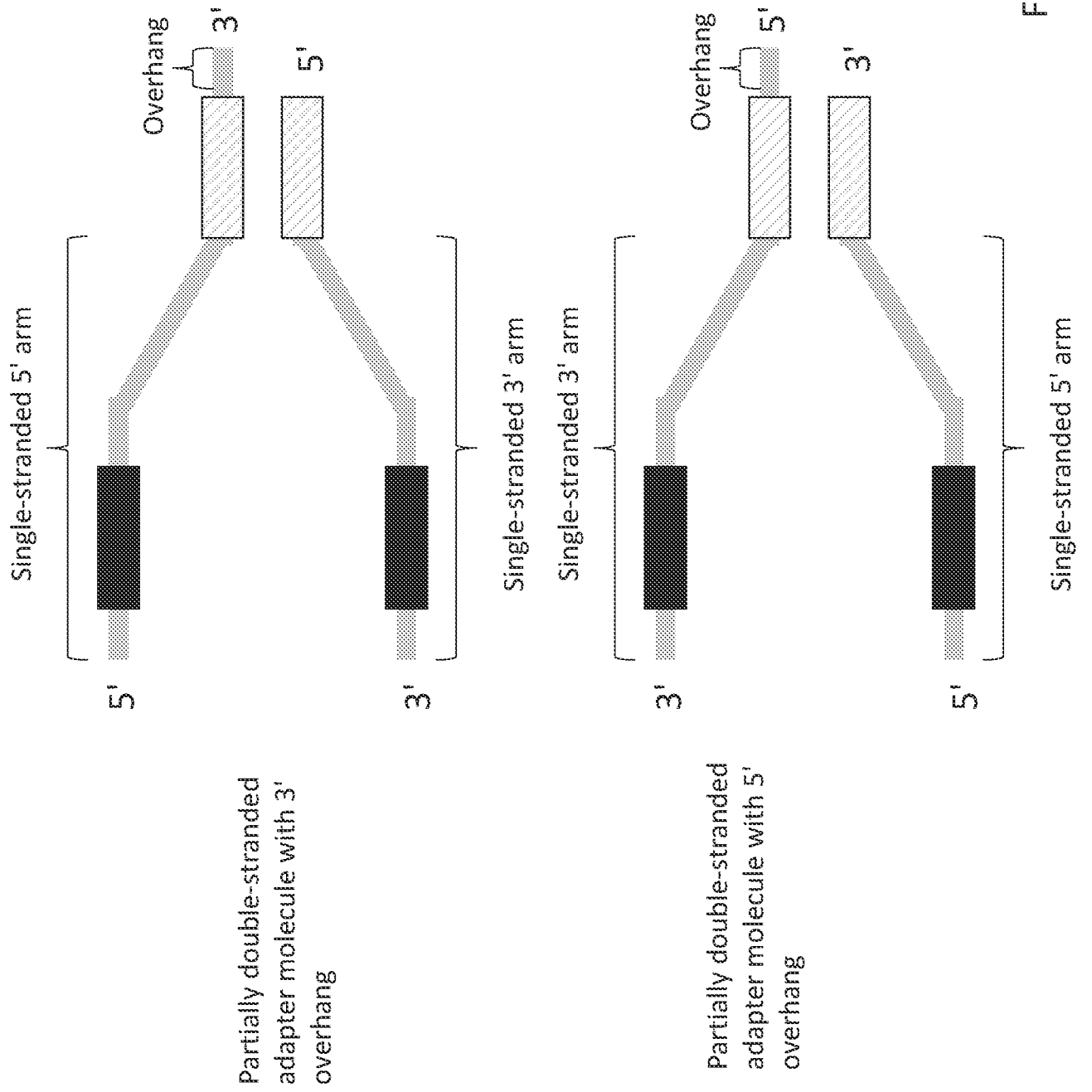


FIG. 3

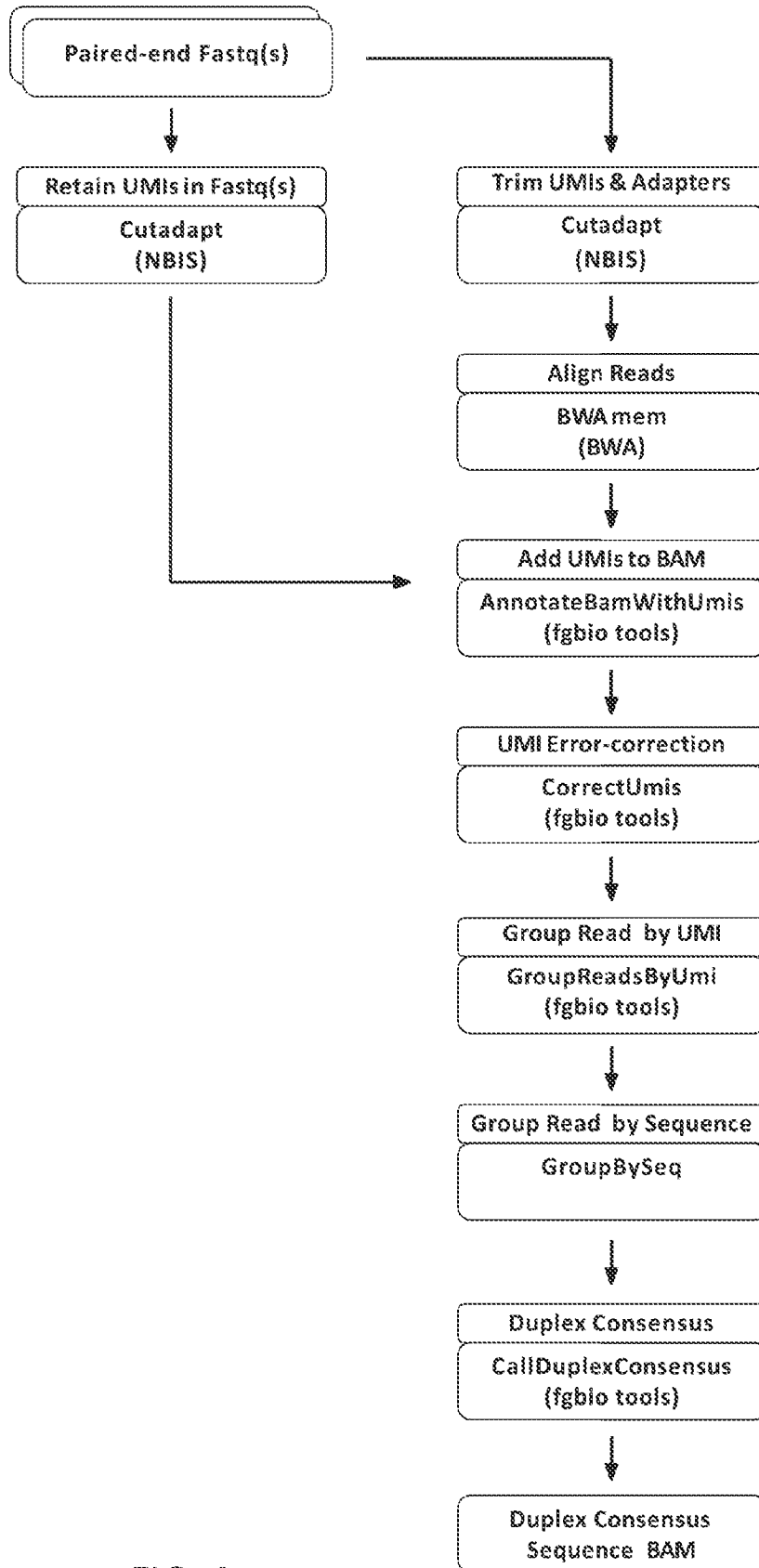


FIG. 4

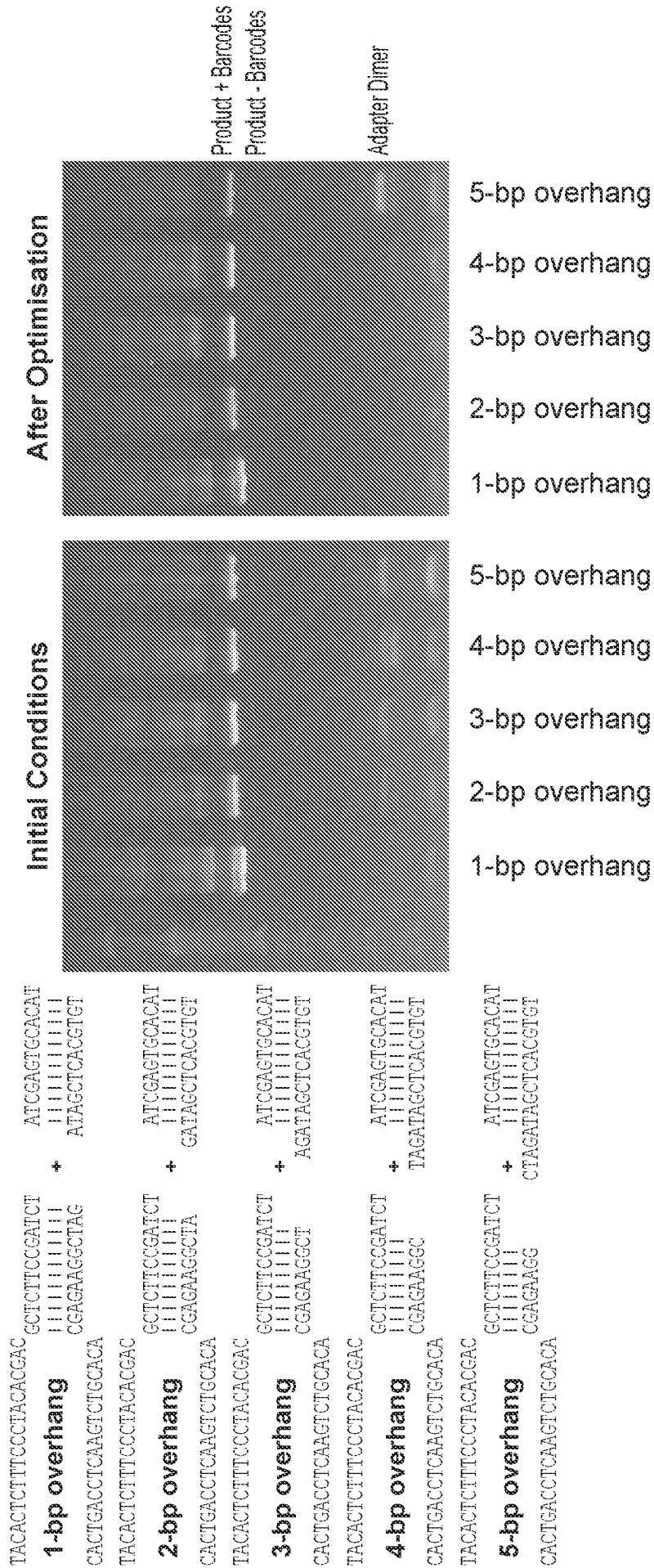
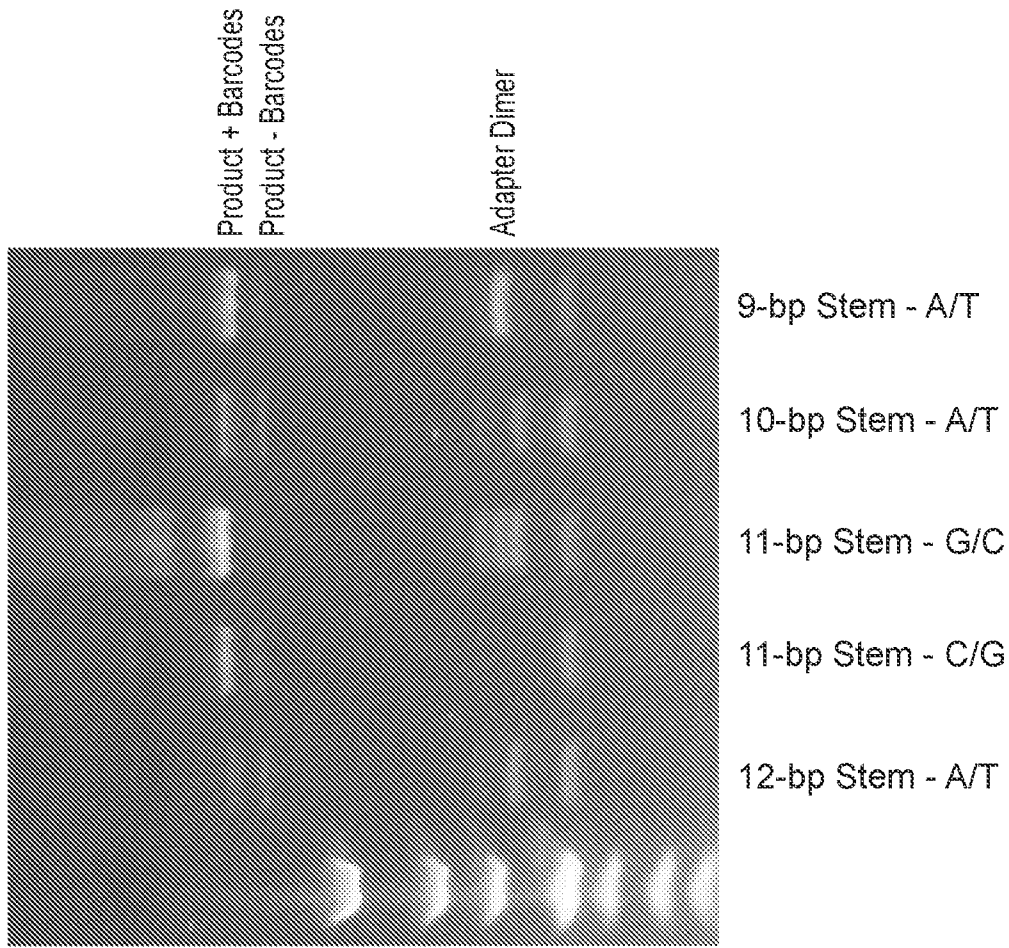


FIG. 5



TACACTCTTTC CCTACAGCAGC	GCTCTTCCGATC	TATCGAGTGCACAT
<b>12-bp Stem - A/T</b>		
CACTGACCTCAAGTCTGCACA	CGAGAAGGCTAGA	TAGCTCACGTGT
TACACTCTTTC CCTACAGCAGC	GCTCTTCCGATC	TATCGAGTGCACAT
<b>11-bp Stem - C/G</b>		
CACTGACCTCAAGTCTGCACA	CGAGAAGGCTA	GATAGCTCACGTGT
TACACTCTTTC CCTACAGCAGC	GCTCTTCCGAT	CTATCGAGTGCACAT
<b>11-bp Stem - G/C</b>		
CACTGACCTCAAGTCTGCACA	CGAGAAGGCTAG	ATAGCTCACGTGT
TACACTCTTTC CCTACAGCAGC	GCTCTTCCGA	TCTATCGAGTGCACAT
<b>10-bp Stem - A/T</b>		
CACTGACCTCAAGTCTGCACA	CGAGAAGGCTA	GATAGCTCACGTGT
TACACTCTTTC CCTACAGCAGC	GCTCTTCCGA	TCTATCGAGTGCACAT
<b>9-bp Stem - A/T</b>		
CACTGACCTCAAGTCTGCACA	CGAGAAGGC	TAGATAGCTCACGTGT

FIG. 6



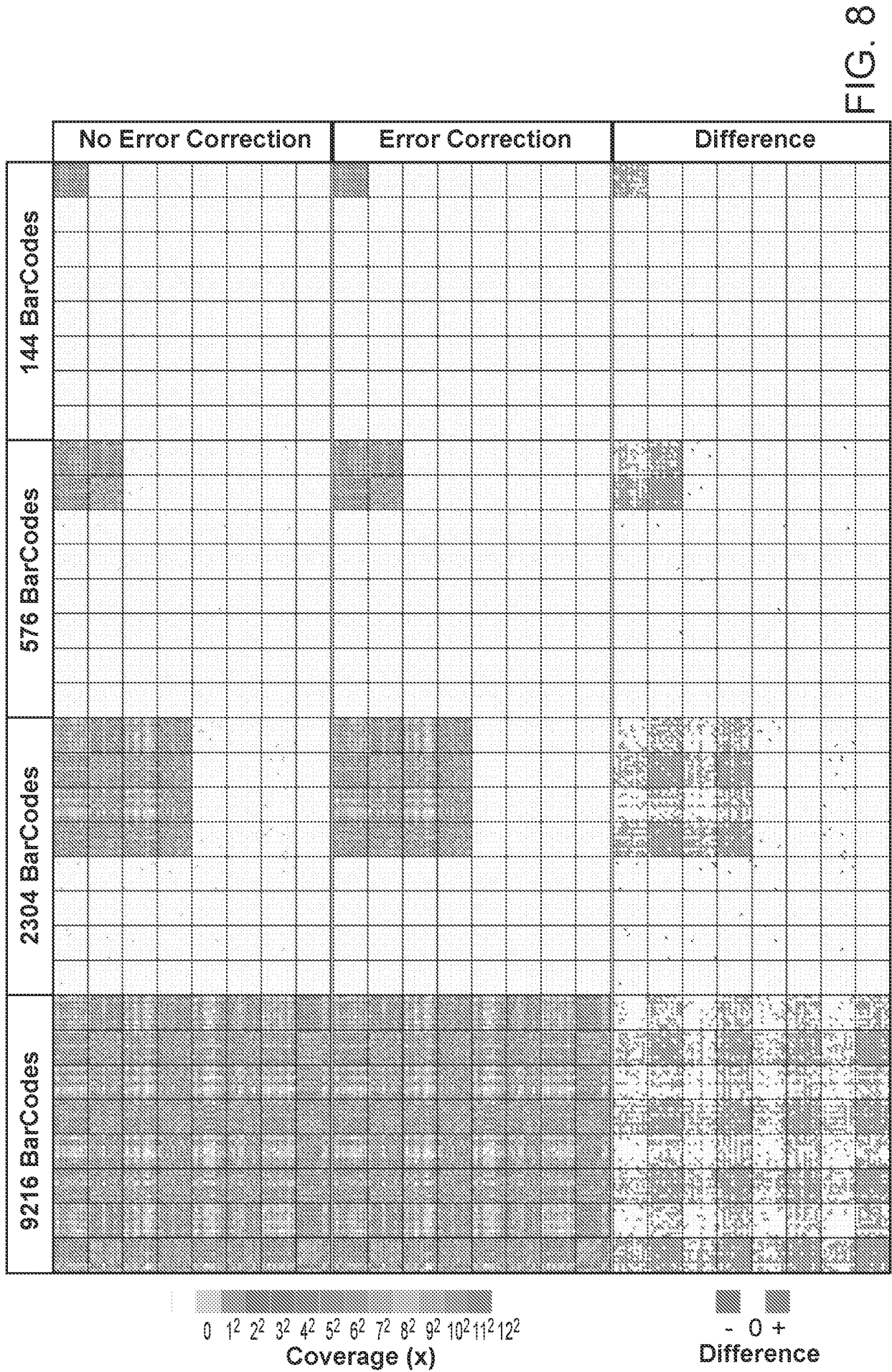


FIG. 8



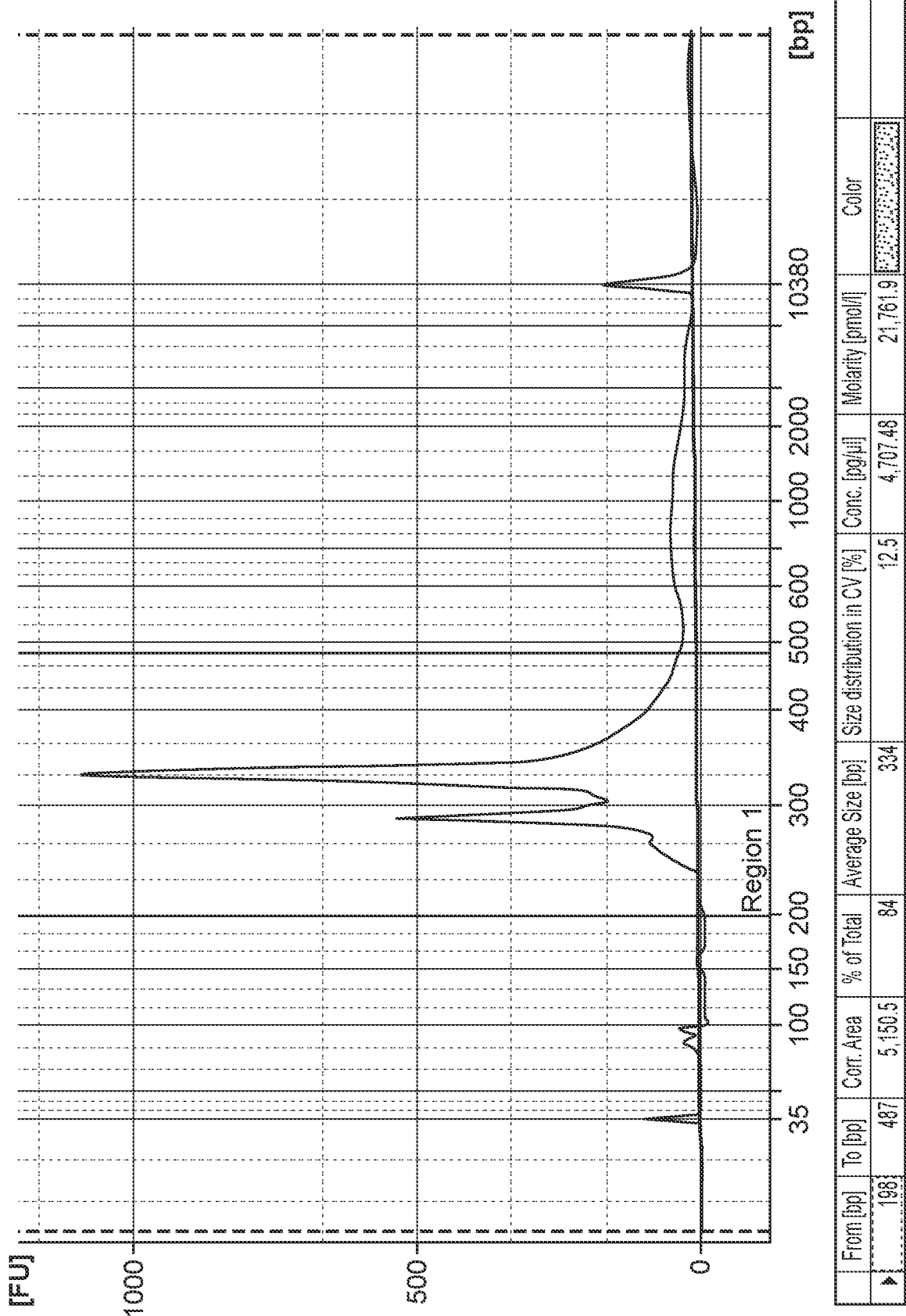


FIG. 9

Gene01 = EGFR4 | Expected Allelic Fraction = 24.5%  
DNA Base Change GGC → AGC | Variant Location Chr7:55,241,707  
Barcodes = 144 | Coverage = 432X | Average Reads per Barcode = 2.0

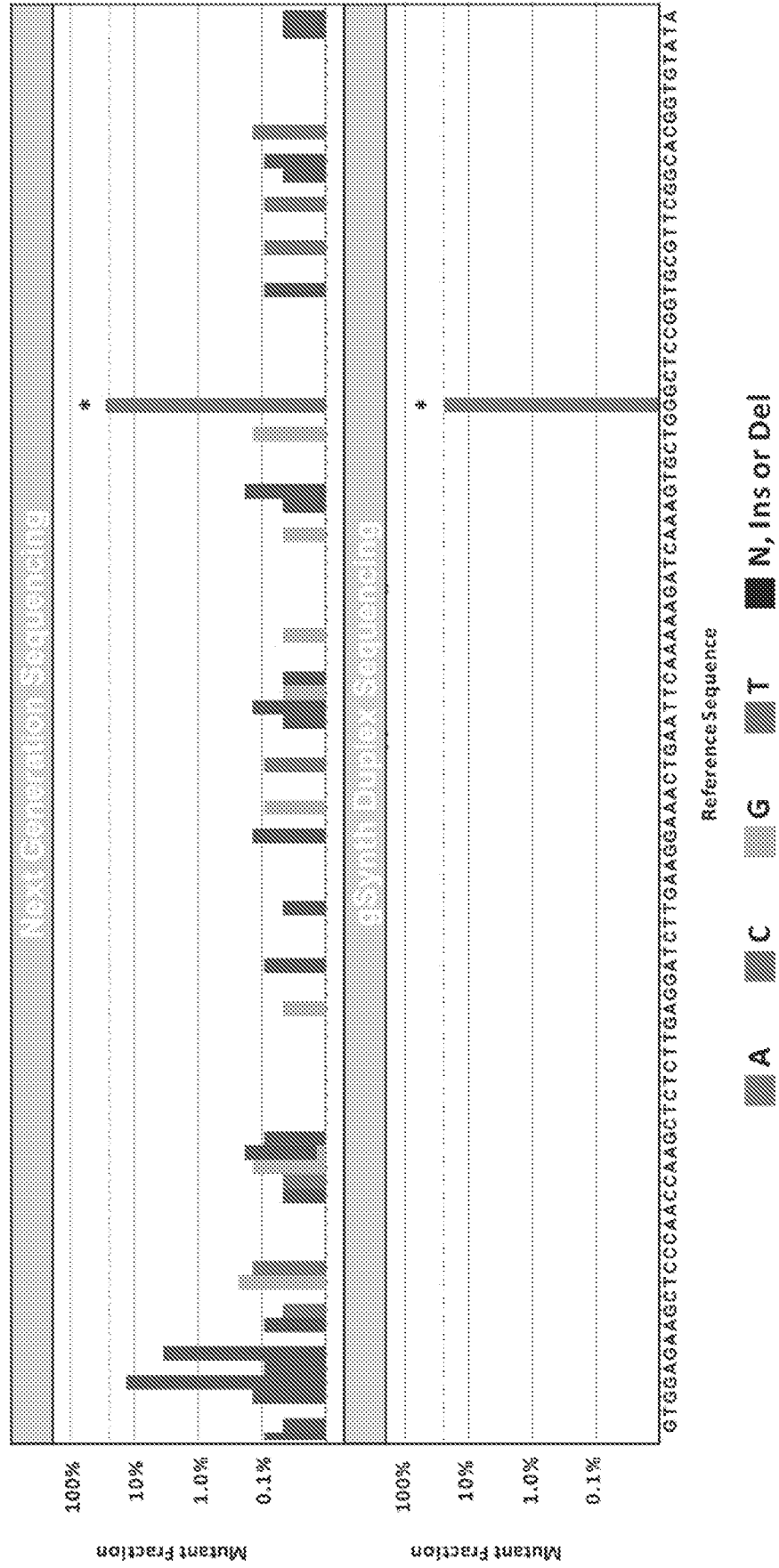


FIG. 10

Gene02 = PI3KCA10 | Expected Allelic Fraction = 17.5%  
DNA Base Change CAT → CGT | Variant Location Chr3:178,952,085  
Barcodes = 144 | Coverage = 432X | Average Reads per Barcode = 1.5

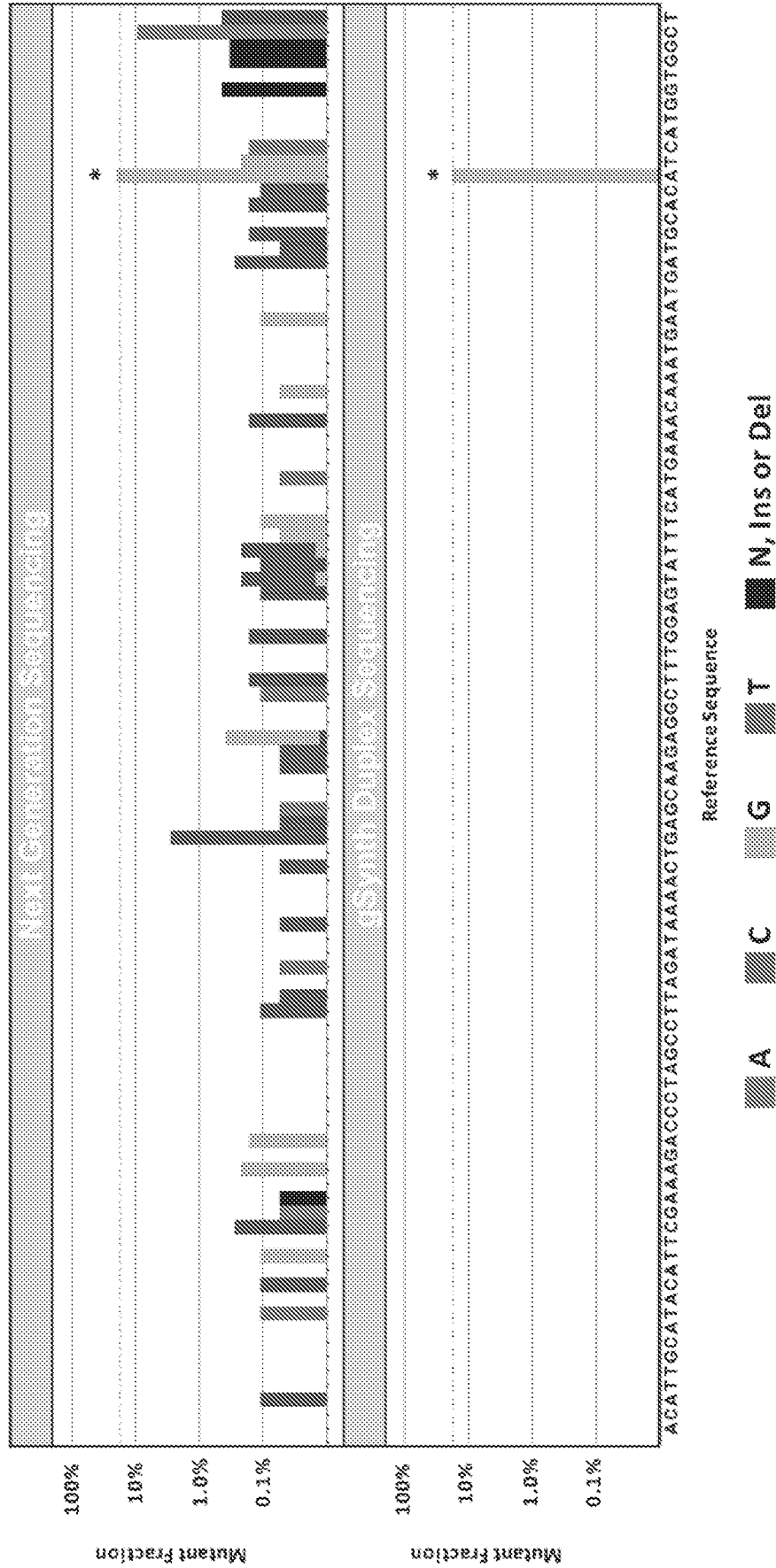


FIG. 11

Gene03 = KRAS1 | Expected Allelic Fraction = 15.0%  
DNA Base Change GGC → GAC | Variant Location Chr12:25,398,281

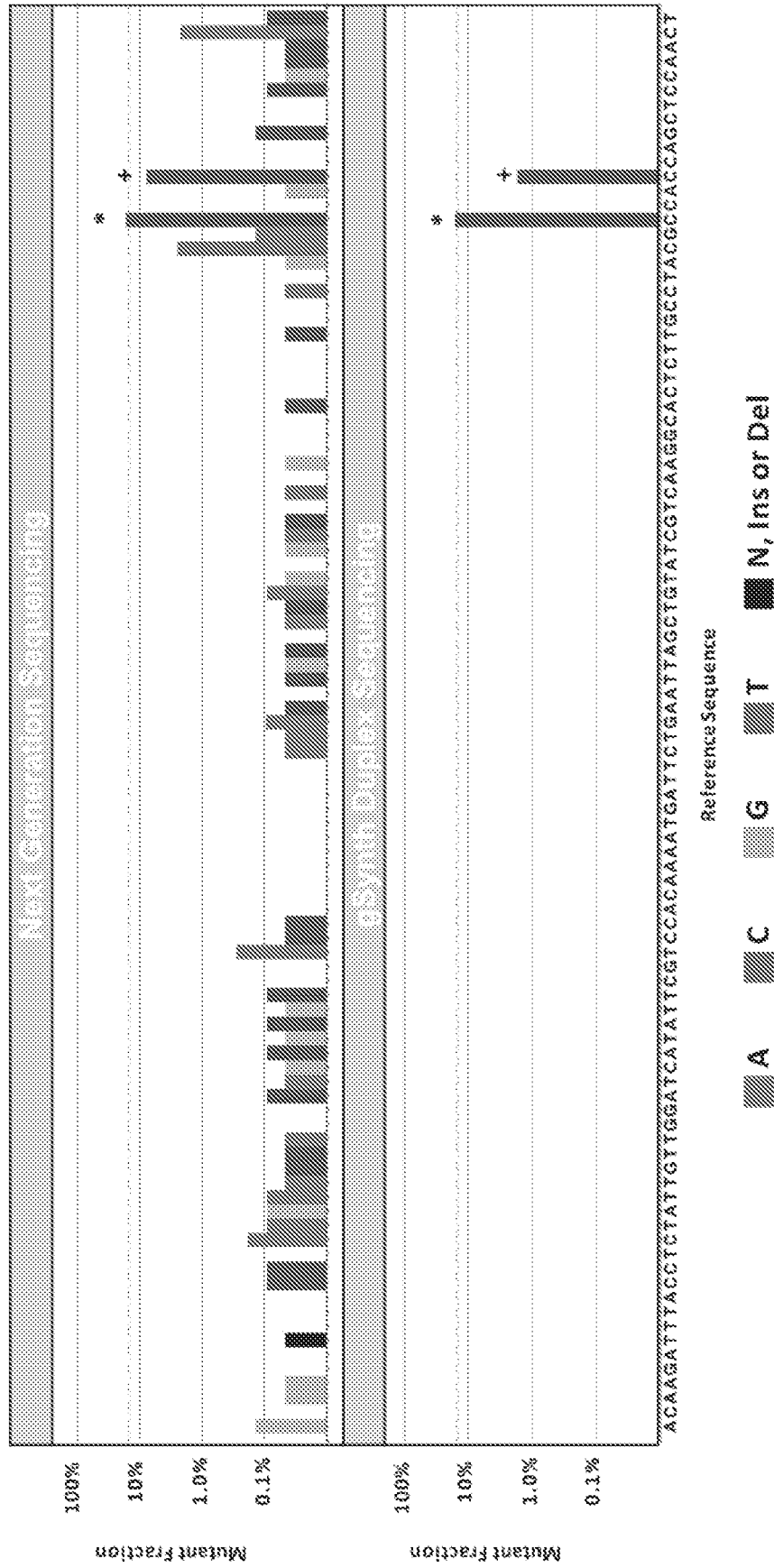


FIG. 12

Gene04 = NRAS | Expected Allelic Fraction = 12.5%  
 DNA Base Change CAA → AAA | Variant Location Chr1:115,256,530  
 Barcodes = 576 | Coverage = 2,304X | Average Reads per Barcode = 4.0

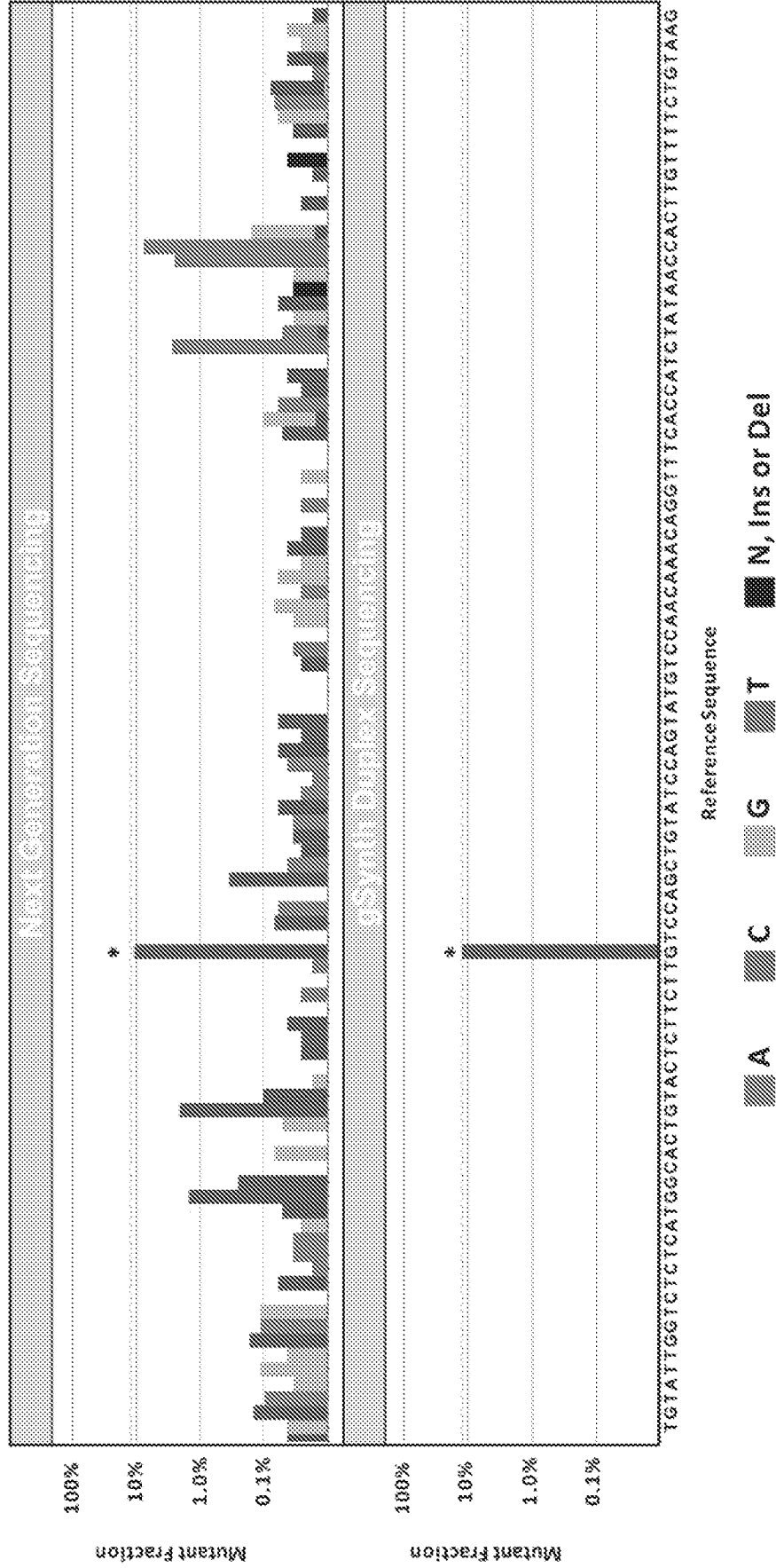


FIG. 13

Gene05 = BRAF | Expected Allelic Fraction = 10.5%  
 DNA Base Change GTG → GAG | Variant Location Chr7:140,453,136  
 Barcodes = 576 | Coverage = 1,728X | Average Reads per Barcode = 3.0

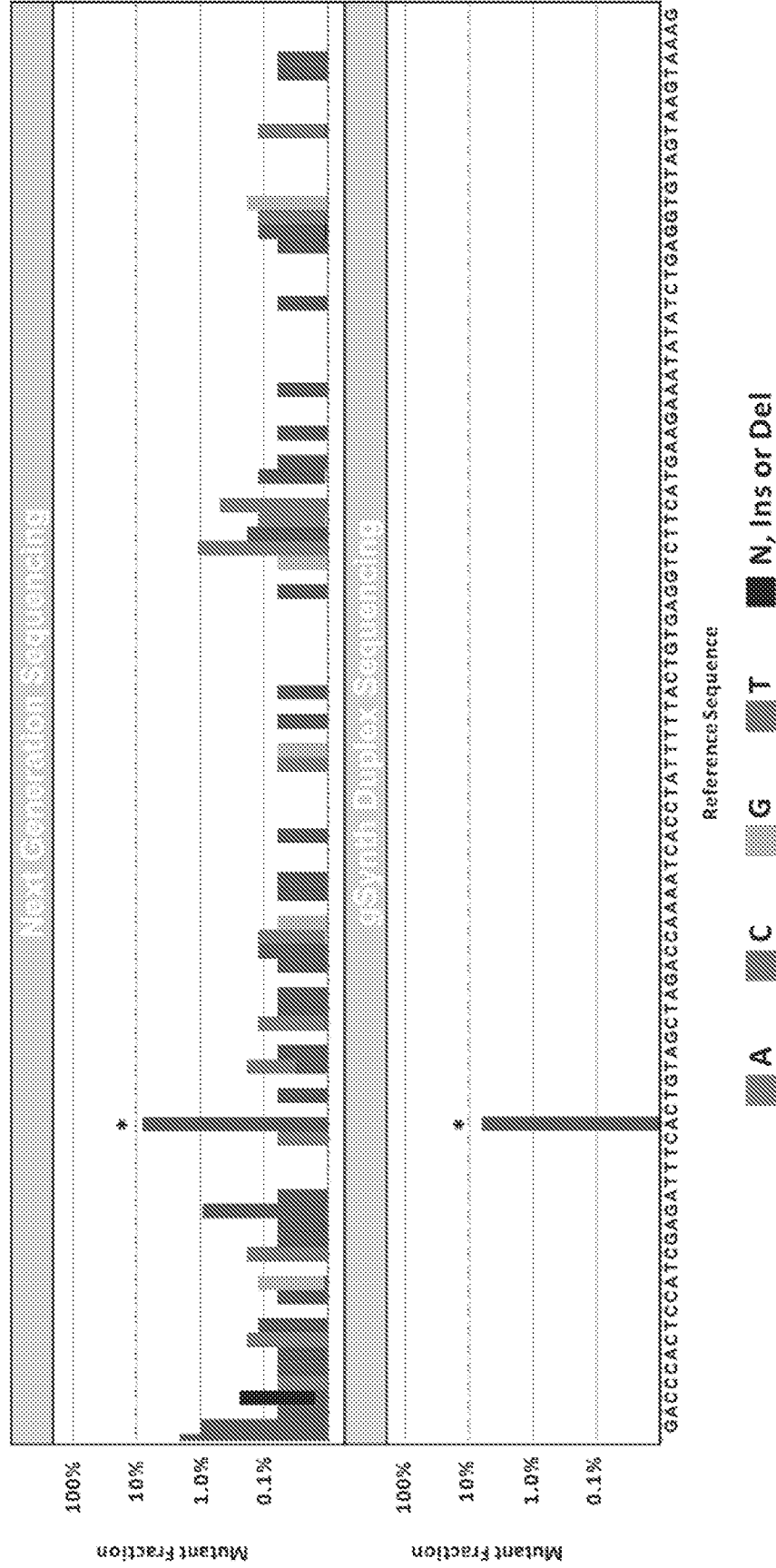


FIG. 14

Gene06 = KIT | Expected Allelic Fraction = 10.0%  
 DNA Base Change GAC → GTC | Variant Location Chr4:55,599,321  
 Barcodes = 576 | Coverage = 1,728X | Average Reads per Barcode = 3.0

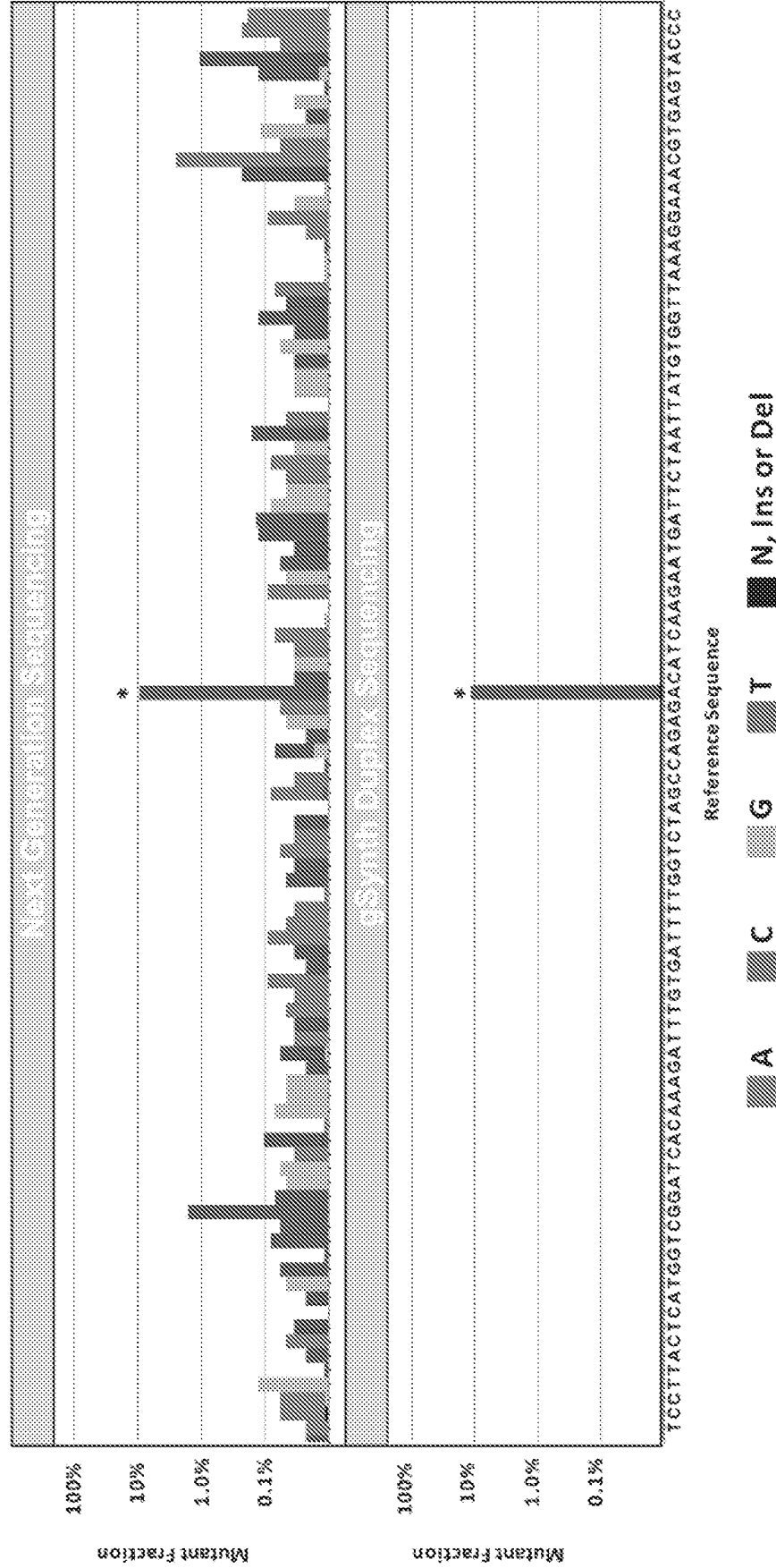


FIG. 15

Gene07 = PI3KCA7 | Expected Allelic Fraction = 9.0%  
 DNA Base Change GAG → AAG | Variant Location Chr3:178,936,091  
 Barcodes = 576 | Coverage = 2,304X | Average Reads per Barcode = 4.5

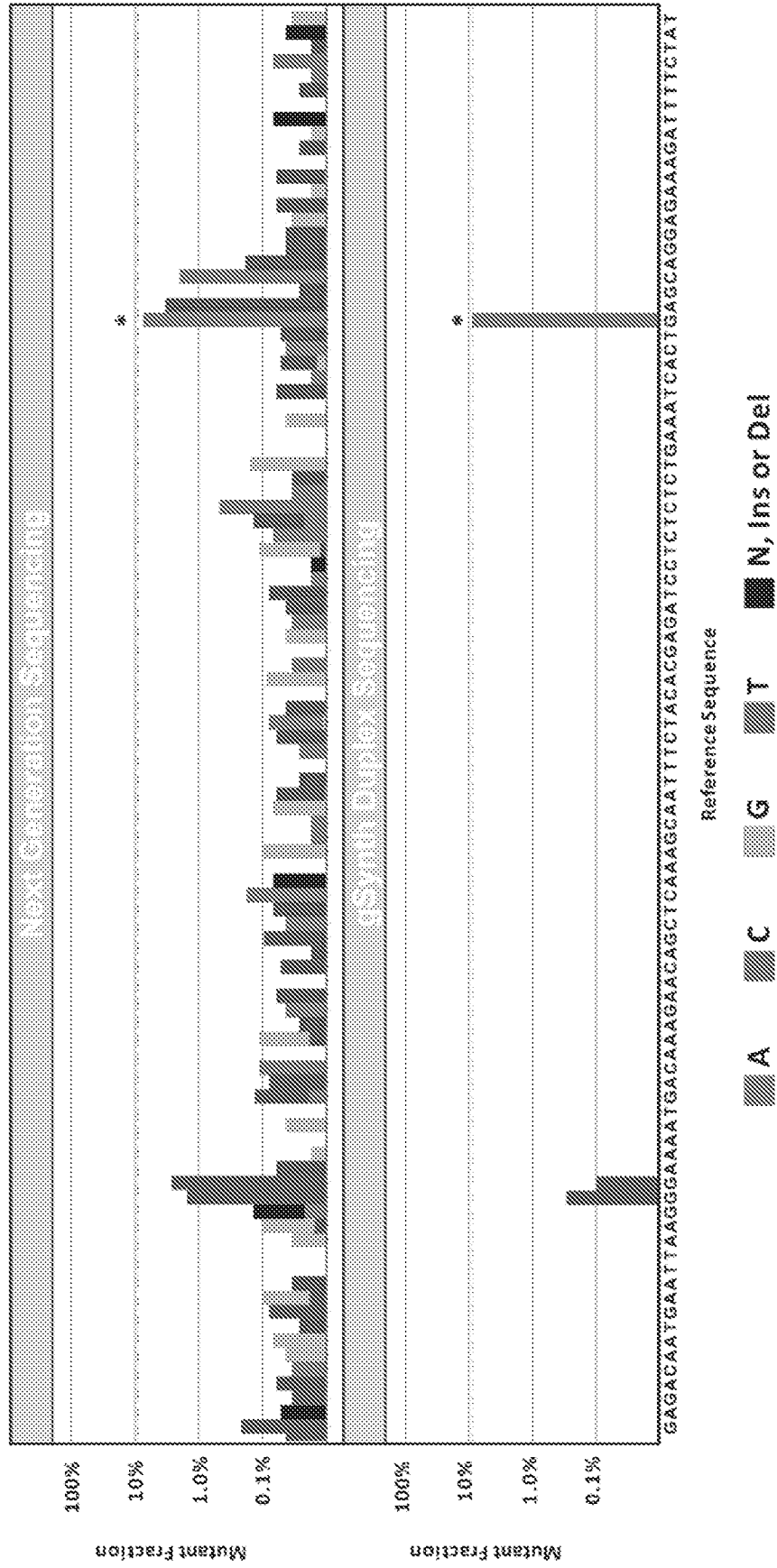


FIG. 16



Gene08 = KRAS1 | Expected Allelic Fraction = 6.0%  
DNA Base Change GGT → GAT | Variant Location Chr12:25,398,284  
Barcodes = 576 | Coverage = 2,304X | Average Reads per Barcode = 4.0

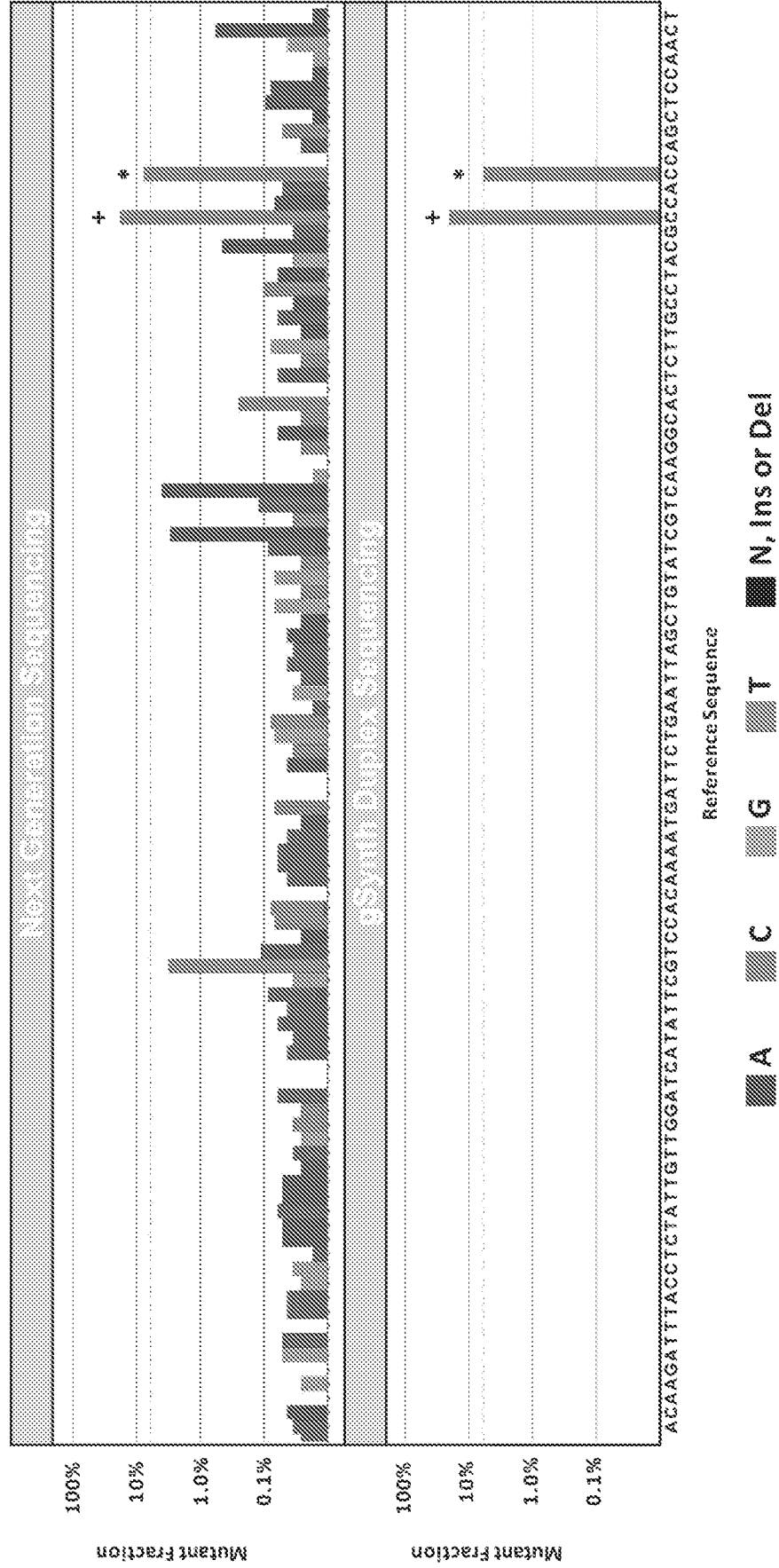


FIG. 17



Gene10 = EGFR5 | Expected Allelic Fraction = 2.0%  
DNA Base AAGGAATTAAGAGAAGCA → AA | Variant Location Chr3:55,242,464  
Barcodes = 2,304 | Coverage = 13,824X | Average Reads per Barcode = 6.0

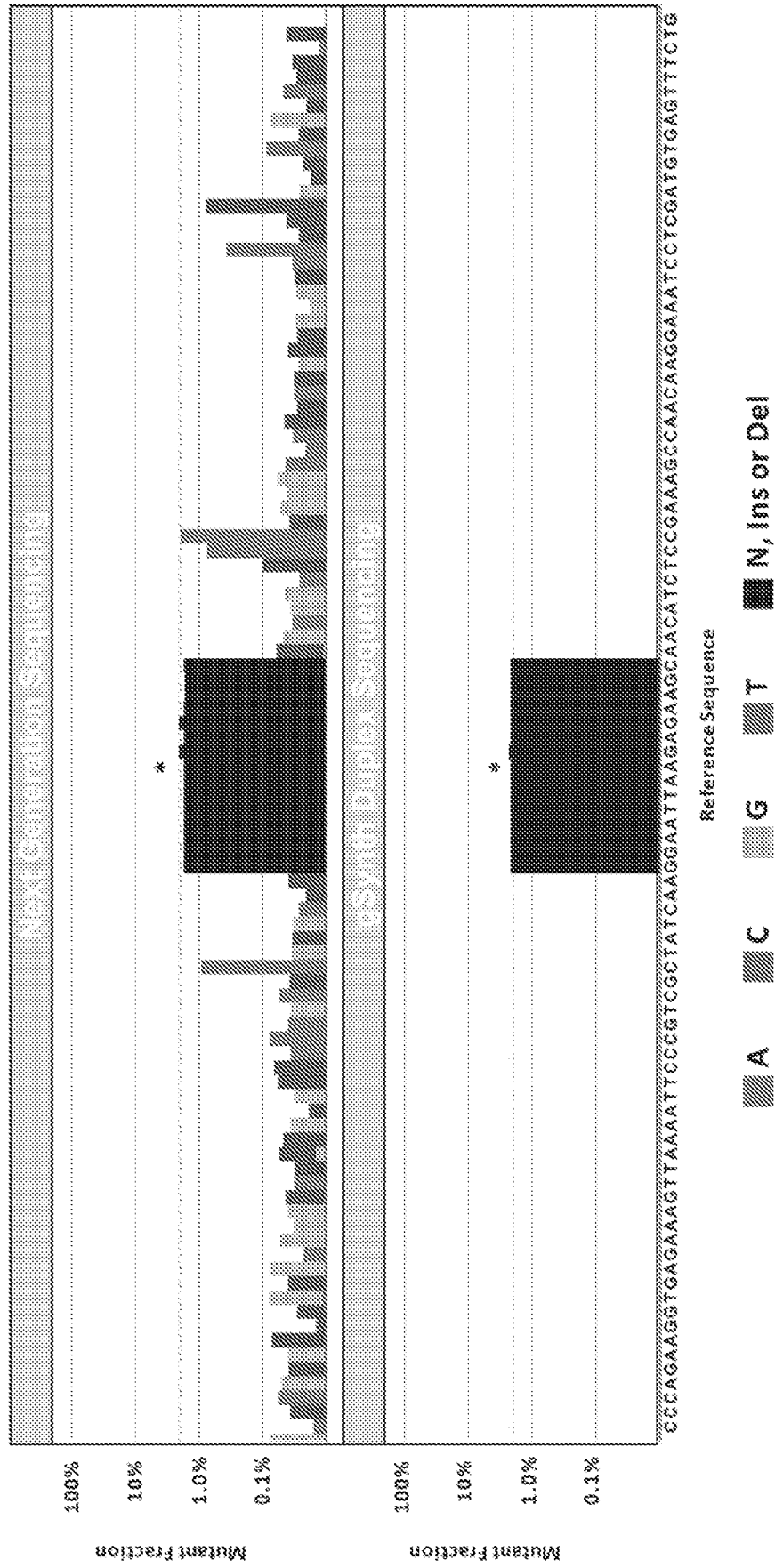


FIG. 19

Gene11 = EGFR6 | Expected Allelic Fraction = 1.0%  
DNA Base Change ACG → ATG | Variant Location Chr3:55,249,071  
Barcodes = 2,304 | Coverage = 23,040X | Average Reads per Barcode = 10

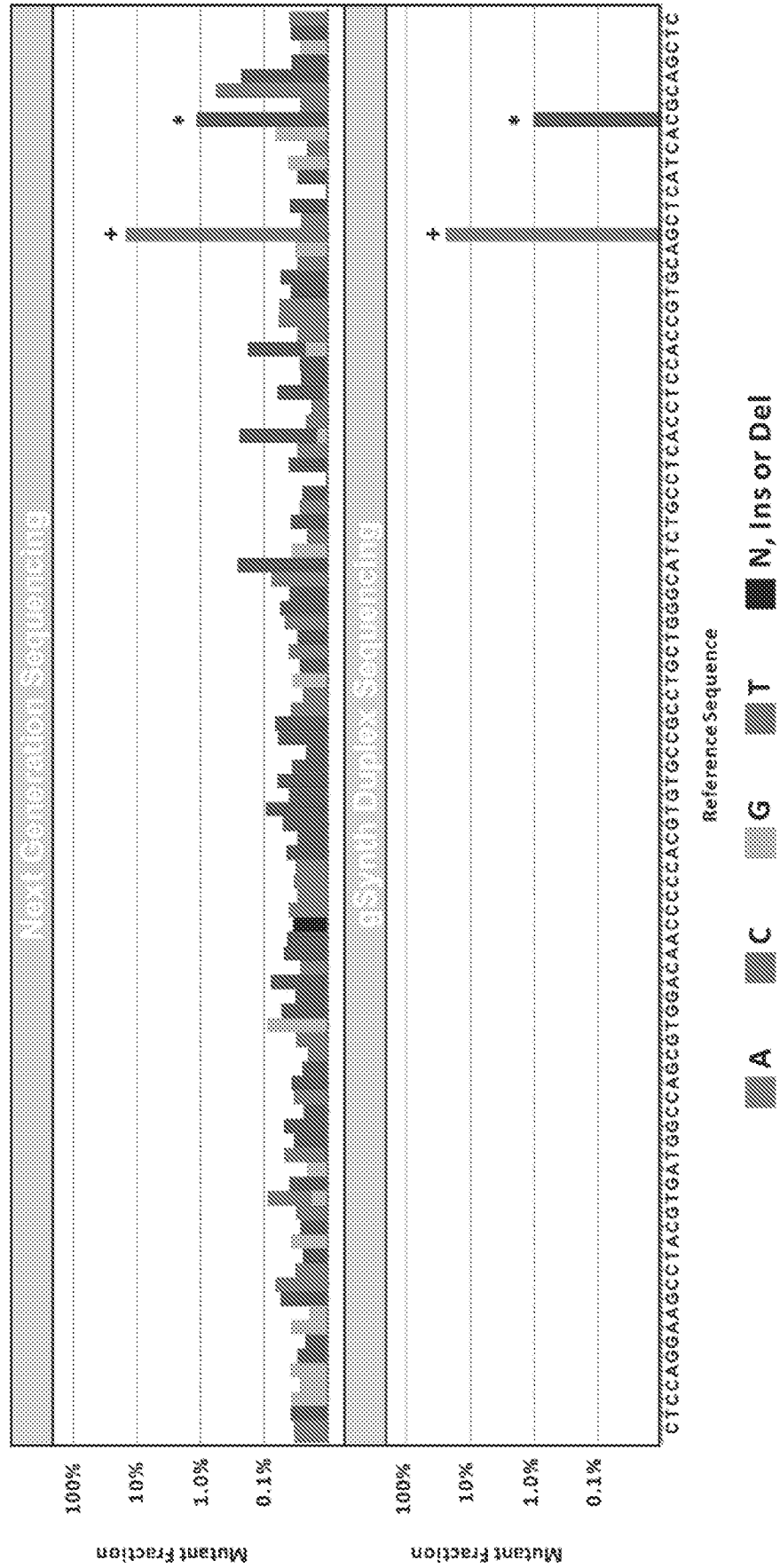


FIG. 20

**INTERNATIONAL SEARCH REPORT**

International application No  
**PCT/US2021/060328**

**A. CLASSIFICATION OF SUBJECT MATTER**  
**INV. C12Q1/6806 C12Q1/6869**  
**ADD.**

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

**B. FIELDS SEARCHED**

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

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

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

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
<b>X</b>	<b>WO 2013/142389 A1 (UNIV WASHINGTON CT COMMERCIALI [US])</b> <b>26 September 2013 (2013-09-26)</b>	<b>1, 3-16, 25, 26</b>
<b>A</b>	<b>figures; page 8, [0022]-[0023]; claims</b> -----	<b>17-24</b>
<b>X</b>	<b>WO 2020/185967 A1 (RED GENOMICS INC [US])</b> <b>17 September 2020 (2020-09-17)</b>	<b>1-16, 25, 26</b>
<b>A</b>	<b>figure 4; claims; page 6; examples</b> -----	<b>17-24</b>
<b>X</b>	<b>WO 2019/094651 A1 (TWINSTRAND BIOSCIENCES INC [US])</b> <b>16 May 2019 (2019-05-16)</b>	<b>1, 3-16, 25, 26</b>
<b>A</b>	<b>figures; page 4, [0011]-page 5 [0014]; pages 7-8</b> -----	<b>17-24</b>
<b>X</b>	<b>WO 2018/050722 A1 (INIVATA LTD [GB])</b> <b>22 March 2018 (2018-03-22)</b>	<b>1, 3-16, 25, 26</b>
<b>A</b>	<b>figures; pages 8-9</b> -----	<b>17-24</b>
	-/--	

Further documents are listed in the continuation of Box C.       See patent family annex.

\* Special categories of cited documents :

<p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"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)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p>	<p>"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</p> <p>"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</p> <p>"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</p> <p>"&amp;" document member of the same patent family</p>
---	---

Date of the actual completion of the international search <b>3 March 2022</b>	Date of mailing of the international search report <b>14/03/2022</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>Hennard, Christophe</b>
--	--

# INTERNATIONAL SEARCH REPORT

International application No

PCT/US2021/060328

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
<b>X</b>	<b>WO 2018/175997 A1 (UNIV WASHINGTON [US])</b>	<b>1, 3-16,</b>
	<b>27 September 2018 (2018-09-27)</b>	<b>25, 26</b>
<b>A</b>	<b>figures; [0093]</b>	<b>17-24</b>
	-----	

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2021/060328

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013142389	A1	26-09-2013	DK 2828218 T3 02-11-2020
			EP 2828218 A1 28-01-2015
			EP 3744857 A1 02-12-2020
			ES 2828661 T3 27-05-2021
			HU E051845 T2 29-03-2021
			PL 2828218 T3 11-01-2021
			PT 2828218 T 11-11-2020
			US 2015044687 A1 12-02-2015
			US 2018142293 A1 24-05-2018
			US 2018363051 A1 20-12-2018
			US 2018363052 A1 20-12-2018
			US 2018363053 A1 20-12-2018
			US 2019093160 A1 28-03-2019
			US 2019093161 A1 28-03-2019
			US 2019093162 A1 28-03-2019
			US 2019119748 A1 25-04-2019
			US 2019119749 A1 25-04-2019
			US 2019271040 A1 05-09-2019
			US 2019284626 A1 19-09-2019
			US 2019284627 A1 19-09-2019
			US 2019292597 A1 26-09-2019
			US 2019323082 A1 24-10-2019
			US 2019338358 A1 07-11-2019
			US 2019352714 A1 21-11-2019
			US 2020318185 A1 08-10-2020
			US 2020392580 A1 17-12-2020
			US 2021324470 A1 21-10-2021
			US 2021371920 A1 02-12-2021
			US 2021371921 A1 02-12-2021
			US 2021371922 A1 02-12-2021
			US 2021371923 A1 02-12-2021
			US 2021371924 A1 02-12-2021
			US 2021381048 A1 09-12-2021
US 2022010376 A1 13-01-2022			
US 2022017961 A1 20-01-2022			
WO 2013142389 A1 26-09-2013			
-----			
WO 2020185967	A1	17-09-2020	US 2021017572 A1 21-01-2021
			WO 2020185967 A1 17-09-2020
-----			
WO 2019094651	A1	16-05-2019	AU 2018366213 A1 14-05-2020
			CA 3081027 A1 16-05-2019
			EP 3707273 A1 16-09-2020
			SG 11202003885U A 28-05-2020
			US 2020362390 A1 19-11-2020
			WO 2019094651 A1 16-05-2019
-----			
WO 2018050722	A1	22-03-2018	US 2019203283 A1 04-07-2019
			WO 2018050722 A1 22-03-2018
-----			
WO 2018175997	A1	27-09-2018	AU 2018240559 A1 19-09-2019
			CA 3057867 A1 27-09-2018
			CN 110520542 A 29-11-2019
			EP 3601598 A1 05-02-2020
			JP 2020511966 A 23-04-2020
			US 2020131561 A1 30-04-2020
			WO 2018175997 A1 27-09-2018
-----			