



- (51) **International Patent Classification:**
C12Q 1/6806 (2018.01) *C12Q 1/6869* (2018.01)
- (21) **International Application Number:**
PCT/EP2024/076525
- (22) **International Filing Date:**
20 September 2024 (20.09.2024)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/583,994 20 September 2023 (20.09.2023) US
63/584,039 20 September 2023 (20.09.2023) US
- (71) **Applicant: ILLUMINA, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).
- (71) **Applicant (for MU only): MARKS & CLERK LLP** [GB/GB]; 15 Fetter Lane, London EC4A 1BW (GB).
- (72) **Inventors: KARUNAKARAN, Aathavan;** Berkeley, California (US). **CARRAMI, Eli;** Cambridge (GB). **SIU, Merek;** Alameda, California (US).

- (74) **Agent: MARKS & CLERK LLP;** 15 Fetter Lane, London EC4A 1BW (GB).

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

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

- (54) **Title:** CONCURRENT SEQUENCING USING NICK TRANSLATION

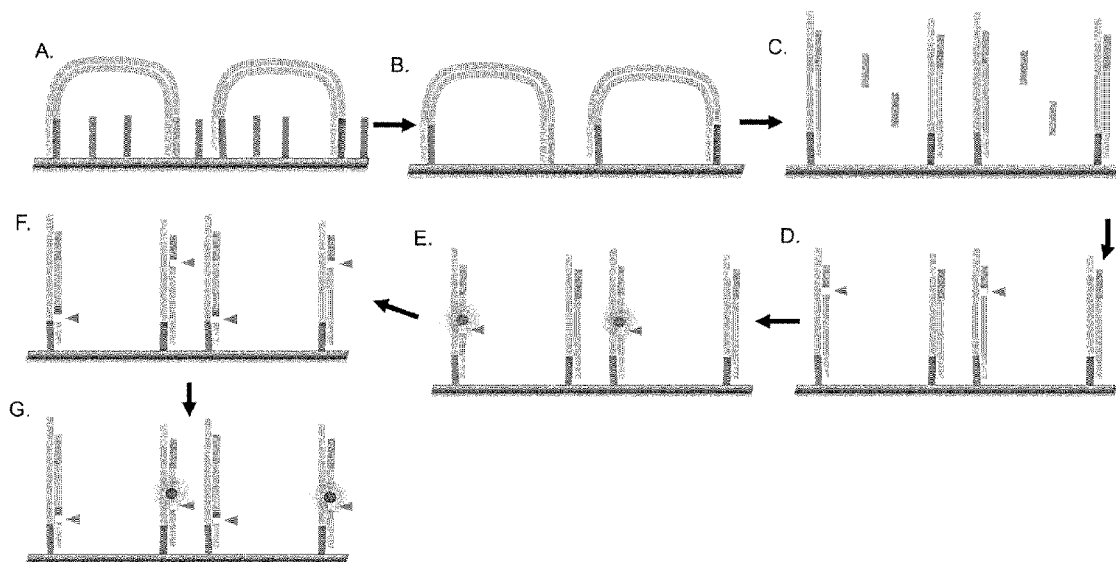


Figure 13

- (57) **Abstract:** The invention relates to methods and kits for use in nucleic acid sequencing, in particular methods for use in double-stranded sequencing, and in particular for use in double-stranded sequence by synthesis (SBS). In a further embodiment, the methods of the invention can be used in concurrent sequencing.

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

Published:

- *with international search report (Art. 21(3))*
- *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))*
- *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

Concurrent sequencing using nick translation

SUMMARY OF THE INVENTION

5 The invention relates to methods and kits for use in nucleic acid sequencing, in particular methods for use in double-stranded sequencing, and in particular for use in double-stranded sequence by synthesis (SBS). In a further embodiment, the methods of the invention can be used in concurrent sequencing.

10 The invention also relates to methods and kits for use in nucleic acid sequencing, in particular methods for use in concurrent sequencing using nick translation, and in particular using a polymerase with 5' to 3' exonuclease activity.

BACKGROUND

15 In some types of next-generation sequencing (NGS) technologies, a nucleic acid cluster is created on a flow cell by amplifying an original template nucleic acid strand. Sequencing cycles may be performed as complementary strands of the template nucleic acids are being synthesized, i.e., using sequencing-by-synthesis (SBS) processes.

20 In each sequencing cycle, deoxyribonucleic acid analogs conjugated to fluorescent labels are hybridized to the template nucleic acids, and excitation light sources are used to excite the fluorescent labels on the deoxyribonucleic acid analogs. Detectors capture fluorescent emissions from the fluorescent labels and identify the deoxyribonucleic acid analogs. As a result, the sequence of the template nucleic acids may be determined by repeatedly performing such sequencing cycles.

25 NGS allows for the sequencing of a number of different template nucleic acids simultaneously, which has significantly reduced the cost of sequencing in the last twenty years. However, there remains a desire for further improvements in sequencing technology, and in particular a desire to improve accuracy.

30 SUMMARY OF THE INVENTION

The accuracy of SBS sequencing can be affected by sequence-specific errors, such as secondary structure elements including G-quads and homopolymers that are formed in single-stranded DNA after the step of linearization using standard SBS protocols.

G-Quads are highly stable secondary structures formed by G-rich sequences in DNA and RNA. The structures of guanine quartets sterically hinder DNA polymerase's movement along a polynucleotide and diminish the enzyme's kinetics and fidelity. As such, reducing the formation of G-Quads is highly desirable. However, the high stability of G-quads means these structures are difficult to remove without affecting the integrity of the template strand.

As described herein, one solution may be to use cleavable sequencing primers to perform double-stranded SBS, consequently improving the accuracy of the sequence read.

According to an aspect of the present invention, there is provided a method of sequencing a first and second polynucleotide sequence, wherein the method comprises preparing the first and second polynucleotide sequence for sequencing and sequencing the nucleobases in the first and second polynucleotide sequence, wherein the first and second polynucleotide sequences are sequenced using nick translation.

In one embodiment, the method of preparing a first polynucleotide sequence and a second polynucleotide sequence for sequencing, comprises

- a. hybridising the first polynucleotide sequence to first immobilised primers on a solid support and hybridising the second polynucleotide sequence to second immobilised primers on a solid support;
- b. synthesising a plurality of first and second polynucleotide sequences by conducting an amplification reaction to extend the first and second immobilised primers;
- c. selectively removing first and second immobilised primers that have not been extended;
- d. applying a plurality of first and second sequencing primers, wherein the first sequencing primers comprise a first cleavable site and wherein the second sequencing primers comprise a second cleavable site and conducting an amplification reaction to extend the first and second sequencing primers; and
- e. selectively cleaving the first cleavable site in the first sequencing primer and/or selectively cleaving the second cleavable site in the second sequencing primer.

In one embodiment, the first and second immobilised primers that have not been extended are removed using a single-stranded exonuclease.

5 The first and second cleavable sites may be an unnatural nucleotide, wherein the unnatural nucleotide may be selected from oxoguanine, optionally 8-oxoguanine, a hypoxanthine or optionally inosine, xanthine, a methylguanine, a methyladenines, a modified cytosine, dihydrouracil, and uracil.

10 In one embodiment, the first and second cleavable sites are the same.

In another embodiment, the first and second cleavable sites are different.

15 In one embodiment, the first or second sequencing primers comprise a mixture of cleavable sequencing primers and non-cleavable sequencing primers.

The method may further comprise a step of linearising the first polynucleotide sequence and the second polynucleotide sequence following step (b).

20 In one embodiment, the first polynucleotide sequence is a forward strand of a double-stranded polynucleotide to be identified and the second polynucleotide sequence is a reverse strand of a double-stranded polynucleotide to be identified.

In one embodiment, there is provided a method of sequencing a first and second polynucleotide sequence, wherein the method comprises:

25 preparing first and second polynucleotide sequences for sequencing using a method of the invention; and
 sequencing nucleobases in the first polynucleotide sequence and the second polynucleotide sequence.

30 In one embodiment the method comprises cleaving the first cleavable site in the first sequencing primer and sequencing the nucleobases in the first polynucleotide sequence.

35 The method may further comprise subsequently cleaving the second cleavable site in the second sequencing primer and sequencing the nucleobases in the first polynucleotide sequence.

The method may further comprise blocking the 3' end of the first polynucleotide sequence prior to sequencing the second polynucleotide.

5 Alternatively, the method comprises cleaving the first cleavable site in the first sequencing primer and cleaving the second cleavable site in the second sequencing primer and sequencing nucleobases in the first polynucleotide sequence and the second polynucleotide sequence, wherein preferably the method comprises applying a plurality of sequencing primers, wherein the first or second sequencing primers comprise a mixture of cleavable sequencing primers and non-cleavable sequencing primers.

10

In one embodiment, the first or second sequencing primers comprise a 1:1 mixture of cleavable sequencing primers and non-cleavable sequencing primers.

15

In one embodiment, the step of concurrently sequencing nucleobases comprises performing sequencing-by-synthesis or sequencing-by-ligation. The step of concurrently sequencing nucleobases may comprise treatment with a polymerase and a 5'-3' exonuclease or a polymerase with 5' to 3' exonuclease activity. Alternatively, the step of concurrently sequencing nucleobases comprises treatment with a strand displacement polymerase (e.g. phi29).

20

Alternatively, in one embodiment, there is provided a method of preparing a first polynucleotide sequence and a second polynucleotide sequence for, preferably, simultaneous sequencing, wherein the method comprises

25

a. attaching a first adaptor to a 5' end of the first polynucleotide sequence and attaching a first adaptor to a 5' end of the second polynucleotide sequence;

30

b. attaching a second adaptor to a 3' end of the first polynucleotide sequence and attaching a second adaptor to a 3' end of the second polynucleotide sequence; wherein the first adaptor comprises at least one cleavable site and the second adaptor comprises a complement of at least one cleavable site, and wherein the first and second adaptors comprises an immobilised primer-binding sequence and a sequencing primer binding sequence or complements thereof;

35

c. hybridising the first and second polynucleotide sequences to first and second immobilised primers on a solid support, wherein optionally, the first and second immobilised primers comprise at least one cleavable site;

- d. forming a cluster of amplified first and second polynucleotide sequences;
and
- e. cleaving at least one cleavable sites to provide sequencing sites for nick translation.

5

Accordingly, in this embodiment, there is provided a method of sequencing a target polynucleotide that provides all the advantages of double-stranded SBS sequencing, without the need to reconfigure the immobilised and/or sequencing primers.

10 The first adaptor may comprise two or at least two cleavable sites and the second adaptor comprise two or at least two cleavable sites.

The first adaptor may comprise one cleavable site and the second adaptor may comprise one complement of a cleavable site wherein the first and second immobilised primers
15 comprise one cleavable site.

The method may further comprise a step of removing a or the sequence between the cleavable sites.

20 The method may further comprise blocking 3'-ends of the first polynucleotide sequences and the second polynucleotide sequences.

The method may further comprise applying a plurality of first and second sequencing primers, wherein the first or second sequencing primers comprise a mixture of blocked
25 sequencing primers and unblocked sequencing primers.

The blocked sequencing primer may comprise a blocking group at a 3' end of the sequencing primer. Optionally, the blocking group may be selected from the group consisting of: a hairpin loop, a deoxynucleotide, a deoxyribonucleotide, a hydrogen atom
30 instead of a 3'-OH group, a phosphate group, a phosphorothioate group, a propyl spacer, a modification blocking the 3'-hydroxyl group, or an inverted nucleobase.

The first or second sequencing primers may comprise a mixture of blocked sequencing primers and unblocked sequencing primers wherein the ratio of blocked to unblocked
35 sequencing primers is 1:1.

In one embodiment, the cleavable sites may be restriction sites for a nicking endonuclease.

In one embodiment, the cluster may be formed by bridge amplification.

5

In one embodiment, the step of cleaving at the cleavage sites is carried out following bridge amplification and before linearization of the extended sequences.

10

In one embodiment, the first polynucleotide sequence is a forward strand of a double-stranded polynucleotide to be identified and the second polynucleotide sequence is a reverse strand of a double-stranded polynucleotide to be identified.

In one embodiment, there is provided, a method of simultaneously sequencing a first and second polynucleotide sequence, comprising:

15

preparing polynucleotide sequences for simultaneous sequencing using a method according the invention; and
concurrently sequencing nucleobases in the first portion and the second portion.

20

In one embodiment, the step of concurrently sequencing nucleobases comprises performing sequencing-by-synthesis or sequencing-by-ligation.

In one embodiment, the step of concurrently sequencing nucleobases comprises treatment with a polymerase and a 5'-3' exonuclease or a polymerase with 5' to 3' exonuclease activity.

25

In another aspect of the invention there is provided a sequencing kit comprising a plurality of first and second sequencing primers, wherein the first and second sequencing primers comprise at least one cleavage site.

30

The kit may further comprise first and second sequencing primers wherein the first and second sequencing primers comprise a mixture of cleavable sequencing primers and un-cleavable sequencing primers.

The kit may further comprise a polymerase and a 5' to 3' exonuclease or a polymerase with 5' to 3' exonuclease activity. The kit may further comprise a single-stranded exonuclease and cleavable effectors, such as nickases.

5 In another aspect of the invention, there is provided a library preparation kit comprising a plurality of first adaptors and a plurality of second adaptors, wherein the first adaptors comprise at least one cleavable site and the second adaptors comprise a complement of at least one cleavable site, and wherein the first and second adaptors comprises an immobilised primer-binding sequence and a sequencing primer binding sequence or
10 complements thereof.

In a further aspect of the invention there is provided a library preparation kit, where the kit comprises a plurality of first and second sequencing primers, wherein the first and second sequencing primers comprise at least one cleavage site, and wherein the kit
15 comprises a plurality of first adaptors and a plurality of second adaptors, wherein the first adaptors comprise at least one cleavable site and the second adaptors comprise a complement of at least one cleavable site, and wherein the first and second adaptors comprises an immobilised primer-binding sequence and a sequencing primer binding sequence or complements thereof.

20 The kit may further comprise first and second sequencing primers wherein the first and second sequencing primers comprise a mixture of blocked sequencing primers and unblocked sequencing primers, and wherein the ratio of blocked to unblocked sequencing primers is 1:1.

25 The kit may also comprise a polymerase and a 5' to 3' exonuclease or a polymerase with 5' to 3' exonuclease activity.

30 In one aspect of the invention there is provided a data processing device comprising means for carrying out the method of the invention.

In another aspect of the invention there is provided a data processing device comprising means for carrying out a method of the invention. In one embodiment, the data processing device is a polynucleotide sequencer.

In another aspect of the invention, there is provided a computer-readable storage medium comprising instructions which, when executed by a processor, cause the processor to carry out a method of the invention.

5 In another aspect of the invention, there is provided a computer-readable data carrier having stored thereon a computer program product of the invention.

Accordingly, the present invention provides all the advantages of double-stranded SBS sequencing, without the need to reconfigure the immobilised and/or sequencing primers.

DESCRIPTION OF THE DRAWINGS

10

Features of examples of the present disclosure will become apparent by reference to the following detailed description and drawings, in which like reference numerals correspond to similar, though perhaps not identical, components. For the sake of brevity, reference numerals or features having a previously described function may or may not be described in connection with other drawings in which they appear.

15

Figure 1 shows a forward strand, reverse strand, forward complement strand, and reverse complement strand of a polynucleotide molecule.

20

Figure 2 shows an example of a polynucleotide sequence (or insert) with 5' and 3' adaptor sequences.

Figure 3 shows a typical polynucleotide with 5' and 3' adaptor sequences.

25

Figure 4 includes Panels A-I and shows an example of a method of sequencing a library strand. **(A)** Each duplex of the library is de-hybridized, and the single strands flowed across a solid support (e.g. a flow cell) to attach to the solid support via Watson-Crick binding to a complementary lawn primer (P5 or P7) and become immobilised. **(B)** The immobilised primers (P5 and P7) are then extended (using the hybridised strand as a "template") to generate a first or second immobilised template strand. For example, the first extended immobilised strand may comprise a first primer sequence at its 5' end (e.g. P5), and a first primer-binding sequence at its 3' end (e.g. P7'). **(C)** Displacing or de-hybridising the (non-immobilised) library strands from the first or second immobilised strands. **(D)** Hybridising the 3' primer binding sequence of the first library strand to a first lawn primer. **(E)** Carrying out an extension reaction to extend the lawn primers to generate a first or second immobilised (also referred to herein as extended) template

30

strand complementary to the library strand. **(F)** Considering two immobilised strands, the second extended immobilised strand may comprise a second primer sequence at its 5' end (e.g. P7), and a second primer-binding sequence at its 3' end (e.g. P5'). The first and second library strands may be the forward strand and reverse strand respectively of a polynucleotide duplex. **(G)** Hybridising the 3' primer binding sequence of the first library strand to a first lawn primer and hybridising the 3' primer binding sequence of the second library strand to a second lawn primer; and carrying out an extension reaction to extend the lawn primers to generate a first or second immobilised (also referred to herein as extended) template strand complementary to the library strands, wherein the immobilised strands comprise a 3' (second or first respectively) primer binding sequence. **(H)** Displacing or de-hybridising the (non-immobilised) library strands from immobilised strands. **(I)** Extension of the immobilised primers using the now bridged first or second extended strands as a template. This step is referred to as clustering. In one embodiment, the cluster is generated by bridge amplification. The resulting structure may be referred to herein as a sequence bridge.

Figure 5 shows a comparison between four workings of the conventional (A) NGS SBS workflow, (B) SPEAR workflow, (C) single-read dsDNA nick-translation, and (D) 16QaM SEAR using nick translation SBS workflows. These workflows are only examples and can be altered according to the embodiments described herein. The dark font highlights which stages of the conventional SBS and SPEAR SBS workflows are eliminated using nick translation of this invention.

Figure 6 shows an exemplar solid support 200 comprising a substrate 204 with a plurality of wells 203. Immobilised primers 201, 202 are found within the well.

Figure 7 shows a graphical representation of sixteen distributions of signals generated by polynucleotide sequences according to one embodiment.

Figure 8 shows the detection of nucleobases using 4-channel, 2-channel and 1-channel chemistry.

Figure 9 shows that by plotting relative intensities of light signals obtained from a first channel (ch1) and a second channel (ch2), a constellation of 16 clouds is obtained.

Figure 10 shows a flow diagram showing a method for base calling according to one embodiment.

Figure 11 shows g-quad coverage.

5

Figure 12 shows that dsDNA-SBS using the present invention reduces sequencing errors at G-quad regions. Read 1-5 represents dsDNA-SBS and read 6, standard SBS. Read 6 (standard SBS) had a greater error rate at G-quad regions (shown by the peak) compared to multiple reads using ds-DNA-nt-SBS (1-5).

10

Figure 13 illustrates the key stages of the single-read embodiment of this invention. A) Seeding and clustering DNA libraries onto conventional immobilized primers. B) Exonuclease treatment to remove excess surface primers. C) A further round(s) of amplification with in-solution sequencing primers containing at least one cleavage site. D) Primers are extended to complete dsDNA molecules. E) A cleavage effector, for example a nicking enzyme, is added and cleaves the sequencing primer. F) Nick translate SBS read 1 is completed, using a polymerase and a 5'→3' exonuclease. G) The 3' end of the growing first read is blocked, and the second read primer is nicked. H) Nick translate SBS read 2 is completed. This eliminates the need for cluster resynthesis for read 2, reducing the sequencing time.

15

20

Figure 14 illustrates the key stages of the simultaneous sequencing embodiment of this invention. A) Seeding and clustering DNA libraries onto conventional immobilized primers. B) Exonuclease treatment to remove excess surface primers. C) A further round(s) of amplification with in-solution sequencing primers containing at least one cleavage site, wherein one of the SBS primers is a mix of nickable and non-nickable forms. D) A cleavage effector, for example a nicking enzyme, is added and cleaves the nickable sequencing primers. E) Nick-translation SBS of read 1 and read 2 are completed simultaneously. A different intensity ratio of read 1 and read 2, resulting from the mix of nickable and non-nickable primers is decoded by 16 QAM.

25

30

Figure 15 shows preparation of a library using a loop fork method.

Figure 16 shows an example of a concatenated polynucleotide sequence comprising a first portion and a second portion, as well as terminal and internal adaptor sequences.

35

Figure 17 shows an example of a concatenated polynucleotide sequence comprising a first portion and a second portion, as well as terminal and internal adaptor sequences.

Figure 18 shows an example of possible cleavage positions (e.g. nicking sites) (arrows) within the 5' and 3' adaptors. Each adaptor sequence may comprise at least one or at least two of cleavage sites. Exemplar possible cleavage sites include at the junctions between each component forming the 3' or 5' adaptor or at positions within each component of the adaptor; for example, a cleavage site may be positioned at the junction between an index (I5 or I7) and a SBS primer site (SBS3 or SBS'12), or alternatively be positioned within the SBS primer site or within the index sequence itself.

Figure 19 shows the key steps of one embodiment of the present invention: A) clustering by bridge amplification of first and second polynucleotide strands each comprising two cleavage sites within the adaptor regions; B) cleaving the first and second polynucleotide strands at a cleavable site found in the 5' adaptor regions and melting the nicked region; C) Addition of first and second sequencing primers, wherein either the first or second sequencing primers comprise a mix of blocked and unblocked sequencing primers; D) Nick-translation sequencing-by-synthesis using a polymerase and a 5'-3' exonuclease or a polymerase comprising 5' to 3' exonuclease activity. The use of blocked and unblocked sequencing primers generates a different (preferably 2:1) intensity between read 1 and read 2, allowing for their simultaneous detection.

DETAILED DESCRIPTION OF THE INVENTION

The present invention will now be described by way of the following non-limiting examples.

All patents, patent applications, and other publications referred to herein, including all sequences disclosed within these references, are expressly incorporated herein by reference, to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference.

All documents cited are, in relevant part, incorporated herein by reference in their entireties for the purposes indicated by the context of their citation herein. However, the citation of any document is not to be construed as an admission that it is prior art with respect to the present disclosure.

The present invention can be used in sequencing, in particular duplex sequencing. Methodologies applicable to the present invention have been described in WO 08/041002, WO 07/052006, WO 98/44151, WO 00/18957, WO 02/06456, WO 07/107710, WO05/068656, US 13/661,524 and US 2012/0316086, the contents of which are herein incorporated by reference. Further information can be found in US 20060024681, US 20060292611, WO 06/110855, WO 06/135342, WO 03/074734, WO07/010252, WO 07/091077, WO 00/179553, WO 98/44152 and WO 2022/087150, the contents of which are herein incorporated by reference.

As used herein, the term “variant” refers to a variant polypeptide sequence or part of the polypeptide sequence that retains desired function of the full non-variant sequence. For example, a desired function of the immobilised primer retains the ability to bind (i.e. hybridise) to a target sequence.

As used in any aspect described herein, a “variant” has at least 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or at least 99% overall sequence identity to the non-variant nucleic acid sequence. The sequence identity of a variant can be determined using any number of sequence alignment programs known in the art. As an example, Emboss Stretcher from the EMBL-EBI may be used: ebi.ac.uk/Tools/psa/emboss_stretcher/ (using default parameters: pair output format, Matrix = BLOSUM62, Gap open = 1, Gap extend = 1 for proteins; pair output format, Matrix = DNAfull, Gap open = 16, Gap extend = 4 for nucleotides).

As used herein, the term “fragment” refers to a functionally active series of consecutive nucleic acids from a longer nucleic acid sequence. The fragment may be at least 99%, at least 95%, at least 90%, at least 80%, at least 70%, at least 60%, at least 50%, at least 40% or at least 30% the length of the longer nucleic acid sequence. A fragment as used herein may also retain the ability to bind (i.e. hybridise) to a target sequence.

Sequencing typically comprises four fundamental steps: 1) library preparation to form a plurality of target polynucleotides for identification; 2) cluster generation to form an array of amplified template polynucleotides; 3) sequencing the cluster array of amplified

template polynucleotides; and 4) data analysis to identify characteristics of the target polynucleotides from the amplified template polynucleotide sequences. These steps are described in greater detail below.

5 Library strands and template terminology

For a given double-stranded polynucleotide sequence 100 to be identified, the polynucleotide sequence 100 comprises a forward strand of the sequence 101 and a reverse strand of the sequence 102. See Figure 1.

10

When the polynucleotide sequence 100 is replicated (e.g. using a DNA/RNA polymerase), complementary versions of the forward strand 101 of the sequence 100 and the reverse strand 102 of the sequence 100 are generated. Thus, replication of the polynucleotide sequence 100 provides a double-stranded polynucleotide sequence 100a that comprises a forward strand of the sequence 101 and a forward complement strand of the sequence 101', and a double-stranded polynucleotide sequence 100b that comprises a reverse strand of the sequence 102 and a reverse complement strand of the sequence 102'.

15

20 The term "template" may be used to describe a complementary version of the double-stranded polynucleotide sequence 100. As such, the "template" comprises a forward complement strand of the sequence 101' and a reverse complement strand of the sequence 102'. Thus, by using the forward complement strand of the sequence 101' as a template for complementary base pairing, a sequencing process (e.g. a sequencing-by-synthesis or a sequencing-by-ligation process) reproduces information that was present in the original forward strand of the sequence 101. Similarly, by using the reverse complement strand of the sequence 102' as a template for complementary base pairing, a sequencing process (e.g. a sequencing-by-synthesis or a sequencing-by-ligation process) reproduces information that was present in the original reverse strand of the sequence 102.

25

30

The two strands in the template may also be referred to as a forward strand of the template 101' and a reverse strand of the template 102'. The complement of the forward strand of the template 101' is termed the forward complement strand of the template 101, whilst the complement of the reverse strand of the template 102' is termed the reverse complement strand of the template 102.

35

Generally, where forward strand, reverse strand, forward complement strand, and reverse complement strand are used herein without qualifying whether they are with respect to the original polynucleotide sequence 100 or with respect to the “template”, these terms may be interpreted as referring to the “template”.

Language for original polynucleotide sequence 100	Corresponding language for the “template”
Forward strand of the sequence 101	Forward complement strand of the template 101 (sometimes referred to herein as forward complement strand 101)
Reverse strand of the sequence 102	Reverse complement strand of the template 102 (sometimes referred to herein as reverse complement strand 102)
Forward complement strand of the sequence 101'	Forward strand of the template 101' (sometimes referred to herein as forward strand 101')
Reverse complement strand of the sequence 102'	Reverse strand of the template 102' (sometimes referred to herein as reverse strand 102')

Library preparation

Library preparation is the first step in any high-throughput sequencing platform. These libraries allow templates to be generated via complementary base pairing that can subsequently be clustered and amplified. During library preparation, nucleic acid sequences, for example genomic DNA sample, or cDNA or RNA sample, is converted into a sequencing library, which can then be sequenced. By way of example with a DNA sample, the first step in library preparation is random fragmentation of the DNA sample. Sample DNA is first fragmented and the fragments of a specific size (typically 200–500 bp, but can be larger) are ligated, sub-cloned or “inserted” in-between two oligo adaptors (adaptor sequences). The original sample DNA fragments are referred to as “inserts”. The target polynucleotides may advantageously also be size-fractionated prior to modification with the adaptor sequences.

As described herein, the templates to be generated typically include separate polynucleotide sequences, in particular a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion. Generating these templates from particular libraries may be performed according to methods known to persons of skill in the art. However, some example approaches of preparing libraries suitable for generation of such templates are described below.

In some embodiments, the library may be prepared by ligating adaptor sequences to double-stranded polynucleotide sequences, each comprising a forward strand of the sequence and a reverse strand of the sequence, as described in more detail in e.g. WO 07/052006, which is incorporated herein by reference. In some cases, "tagmentation" can be used to attach the sample DNA to the adaptors, as described in more detail in e.g. WO 10/048605, US 2012/0301925, US 2013/0143774 and WO 2016/189331, each of which are incorporated herein by reference. In tagmentation, double-stranded DNA is simultaneously fragmented and tagged with adaptor sequences and PCR primer binding sites. The combined reaction eliminates the need for a separate mechanical shearing step during library preparation. These procedures may be used, for example, for preparing templates including a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion, wherein the first portion is a forward strand of the template, and the second portion is a forward complement strand of the template – i.e. a copy of the forward strand (or alternatively, wherein the first portion is a reverse strand of the template, and the second portion is a reverse complement strand of the template).

Where features herein are described in relation to the "forward" strand, it should be considered that these features could equally be applied to the "reverse strand". Equally, where reference is made to the reverse strand, this can mean a complement of the forward strand, and the forward strand can mean a complement of the reverse strand.

Where libraries are prepared by ligating adaptor sequences to double-stranded polynucleotide sequences as described above, library preparation may comprise ligating a first primer-binding sequence 301' or an adaptor that comprises a primer-binding sequence 301' (e.g. P5', such as SEQ ID NO. 3) and a second terminal sequencing primer binding site 304 (e.g. SBS3', for example, SEQ ID NO. 8) to a 3'-end of a forward strand of a sequence 101. See Figure 2. The library preparation may be arranged such

that the second terminal sequencing primer binding site 304 is attached (e.g. directly attached) to the 3'-end of the forward strand of the sequence 101, and such that the first primer-binding sequence 301' is attached (e.g. directly attached) to the 3'-end of the second terminal sequencing primer binding site 304.

5

The library preparation may further comprise ligating a complement or an adaptor comprising a complement of first terminal sequencing primer binding site 303' (e.g. SBS12, such as SEQ ID NO. 9) (also referred to herein as a first terminal sequencing primer binding site complement 303') and a complement of a second primer-binding sequence 302 (also referred to herein as a second primer-binding complement sequence 302) (e.g. P7, such as SEQ ID NO. 2) to a 5'-end of the forward strand of the sequence 101. The library preparation may be arranged such that first terminal sequencing primer binding site complement 303' is attached (e.g. directly attached) to the 5'-end of the forward strand of the sequence 101, and such that second primer-binding complement sequence 302 is attached (e.g. directly attached) to the 5'-end of first terminal sequencing primer binding site complement 303'.

10

15

Thus, one strand of a polynucleotide within a polynucleotide library may comprise, in a 5' to 3' direction, a second primer-binding complement sequence 302 or a first adaptor comprising a second primer-binding complement sequence 302 (e.g. P7), a first terminal sequencing primer binding site complement 303' (e.g. SBS12), a forward strand of the sequence/insert 101, a second terminal sequencing primer binding site 304 or a second adaptor comprising a second terminal sequencing primer binding site 304 (e.g. SBS3'), and a first primer-binding sequence 301' (e.g. P5') (Figure 2 – bottom strand).

20

25

Although not shown in Figure 2, the strand may further comprise one or more index sequences. As such, a first index sequence (e.g. i7) may be provided between the second primer-binding complement sequence 302 (e.g. P7) and the first terminal sequencing primer binding site complement 303' (e.g. SBS12). Separately, or in addition, a second index complement sequence (e.g. i5') may be provided between the second terminal sequencing primer binding site 304 (e.g. SBS3') and the first primer-binding sequence 301' (e.g. P5'). Thus, in some embodiments, one strand of a polynucleotide within a polynucleotide library may comprise, in a 5' to 3' direction, a second primer-binding complement sequence 302 (e.g. P7), a first index sequence (e.g. i7), a first terminal sequencing primer binding site complement 303' (e.g. SBS12), a forward strand of the sequence 101, a second terminal sequencing primer binding site 304 (e.g. SBS3'),

30

35

a second index complement sequence (e.g. i5'), and a first primer-binding sequence 301' (e.g. P5'). A typical polynucleotide is shown in Figure 3.

5 When a double-stranded sequence 100 is used, the library preparation may also comprise ligating a second primer-binding sequence 302' or a third adaptor comprising a second primer-binding sequence 302' (e.g. P7') and a first terminal sequencing primer binding site 303 (e.g. SBS12') to a 3'-end of a reverse strand of a sequence 102. The library preparation may be arranged such that first terminal sequencing primer binding site 303 is attached (e.g. directly attached) to the 3'-end of the reverse strand of the
10 sequence 102, and such that the second primer-binding sequence 302' is attached (e.g. directly attached) to the 3'-end of first terminal sequencing primer binding site 303.

The library preparation may further comprise ligating a complement of a second terminal sequencing primer binding site 304' or a fourth adaptor comprising a complement of a
15 second terminal sequencing primer binding site 304' (e.g. SBS3) (also referred to herein as a second terminal sequencing primer binding site complement 304') and a complement of a first primer-binding sequence 301 (also referred to herein as a first primer-binding complement sequence 301) (e.g. P5) to a 5'-end of the reverse strand of the sequence 102. The library preparation may be arranged such that the second
20 terminal sequencing primer binding site complement 304' is attached (e.g. directly attached) to the 5'-end of the reverse strand of the sequence 102, and such that the first primer-binding complement sequence 301 is attached (e.g. directly attached) to the 5'-end of the second terminal sequencing primer binding site complement 304'.

25 Thus, another strand of a polynucleotide within a polynucleotide library may comprise, in a 5' to 3' direction, a first primer-binding complement sequence 301 or a third adaptor comprising a first primer-binding complement sequence 301 (e.g. P5), a second terminal sequencing primer binding site complement 304' (e.g. SBS3), a reverse strand of the sequence/insert 102, a first terminal sequencing primer binding site 303 or a fourth
30 adaptor comprising a first terminal sequencing primer binding site 303 (e.g. SBS12'), and a second primer-binding sequence 302' (e.g. P7') (Figure 2 – top strand).

Although not shown in Figure 2, the another strand may further comprise one or more index sequences. As such, a second index sequence (e.g. i5) may be provided between
35 the first primer-binding complement sequence 301 (e.g. P5) and the second terminal sequencing primer binding site complement 304' (e.g. SBS3). Separately, or in addition,

a first index complement sequence (e.g. i7') may be provided between the first terminal sequencing primer binding site 303 (e.g. SBS12') and the second primer-binding sequence 302' (e.g. P7'). Thus, in some embodiments, another strand of a polynucleotide within a polynucleotide library may comprise, in a 5' to 3' direction, a first primer-binding complement sequence 301 (e.g. P5), a second index sequence (e.g. i5), a second terminal sequencing primer binding site complement 304' (e.g. SBS3), a reverse strand of the sequence 102, a first terminal sequencing primer binding site 303 (e.g. SBS12'), a first index complement sequence (e.g. i7'), and a second primer-binding sequence 302' (e.g. P7'). A typical polynucleotide is shown in Figure 3 (top strand).

As used herein, the term "genetically unrelated" refers to portions which are not related in the sense of being any two of the group consisting of: forward strands, reverse strands, forward complement strands, and reverse complement strands. However, the "genetically unrelated" sequences could be different fragment sequences which are derived from the same source, but are different fragments from that source (e.g. from the same fragmented library preparation process). This includes sequences that can be overlapping in sequence (but not identical in sequence).

As will be described later, during clustering and amplification, further processes may be used to generate templates including a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion, wherein the first portion and the second portion are genetically unrelated.

In some embodiments, the library may be prepared using a loop fork method, which is described below. This procedure may be used, for example, for preparing templates including a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion, wherein the first portion is a forward strand of the template, and the second portion is a reverse complement strand of the template (or alternatively, wherein the first portion is a reverse strand of the template, and the second portion is a forward complement strand of the template). Such libraries may also be referred to as self-tandem inserts. A representative process for conducting a loop fork method is shown in Figure 15.

Starting from a double-stranded polynucleotide sequence comprising a forward strand of the sequence and a reverse strand of the sequence, adaptors may be ligated to a first end of the sequence (e.g. using processes as described in more detail in e.g. WO

07/052006, or "tagmentation" methods as described above). A second end of the sequence (different from the first end) may be ligated to a loop, which connects the forward strand of the sequence and the reverse strand of the sequence, thus generating a loop fork ligated polynucleotide sequence. Conducting PCR on the loop fork ligated polynucleotide sequence produces a new double-stranded polynucleotide sequence, one strand comprising the forward strand of the sequence and the reverse strand of the sequence, and the other strand comprising a forward complement strand of the sequence and a reverse complement strand of the sequence. The library is now ready for seeding, clustering and amplification.

As will be described later, during clustering and amplification, further processes may be used to generate templates including a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion, wherein the first portion is a forward strand of the template, and the second portion is a reverse complement strand of the template (or alternatively, wherein the first portion is a reverse strand of the template, and the second portion is a forward complement strand of the template).

In a further embodiment, the P5' and P7' primer-binding sequences are complementary to short primer sequences (or lawn primers) present on the surface of a flow cell. Binding of P5' and P7' to their complements (P5 and P7) on – for example – the surface of the flow cell, permits nucleic acid amplification. As used herein "′" denotes the complementary strand.

The primer-binding sequences in the adaptor that permit hybridisation to amplification primers (e.g. also referred to herein as the immobilised or lawn primers) will typically be around 20-40 nucleotides in length, although the invention is not limited to sequences of this length. The precise identity of the amplification primers (e.g. lawn primers), and hence the cognate sequences in the adaptors, are generally not material to the invention, as long as the primer-binding sequences are able to interact with the amplification primers in order to direct PCR amplification. The sequence of the amplification primers may be specific for a particular target nucleic acid that it is desired to amplify, but in other embodiments these sequences may be "universal" primer sequences which enable amplification of any target nucleic acid of known or unknown sequence which has been modified to enable amplification with the universal primers. The criteria for design of PCR primers are generally well known to those of ordinary skill in the art.

The index sequences (also known as a barcode or tag sequence) are unique short DNA (or RNA) sequences that are added to each DNA (or RNA) fragment during library preparation. The unique sequences allow many libraries to be pooled together and sequenced simultaneously. Sequencing reads from pooled libraries are identified and sorted computationally, based on their barcodes, before final data analysis. Library multiplexing is also a useful technique when working with small genomes or targeting genomic regions of interest. Multiplexing with barcodes can exponentially increase the number of samples analysed in a single run, without drastically increasing run cost or run time. Examples of tag sequences are found in WO05/068656, whose contents are incorporated herein by reference in their entirety. The invention is not limited by the number of reads per cluster, for example two reads per cluster: three or more reads per cluster are obtainable simply by dehybridising a first extended sequencing primer, and rehybridising a second primer before or after a cluster repopulation/strand resynthesis step. Methods of preparing suitable samples for indexing are described in, for example WO 2008/093098, which is incorporated herein by reference. Single or dual indexing may also be used. With single indexing, up to 48 unique 6-base indexes can be used to generate up to 48 uniquely tagged libraries. With dual indexing, up to 24 unique 8-base Index 1 sequences and up to 16 unique 8-base Index 2 sequences can be used in combination to generate up to 384 uniquely tagged libraries. Pairs of indexes can also be used such that every i5 index and every i7 index are used only one time. With these unique dual indexes, it is possible to identify and filter indexed hopped reads, providing even higher confidence in multiplexed samples.

The sequencing primer binding sites are sequencing and/or index primer binding sites and indicate the starting point of the sequencing read. In one embodiment, sequencing primer binding site is only required to be a binding site for a sequencing primer. During the sequencing process, a sequencing primer anneals (i.e. hybridises) to at least a portion of the sequencing primer binding site on the template strand. The polymerase enzyme binds to this site and incorporates complementary nucleotides base by base into the growing opposite strand.

As will be understood by the skilled person, a double-stranded nucleic acid will typically be formed from two complementary polynucleotide strands comprised of deoxyribonucleotides or ribonucleotides joined by phosphodiester bonds, but may additionally include one or more ribonucleotides and/or non-nucleotide chemical

moieties and/or non-naturally occurring nucleotides and/or non-naturally occurring backbone linkages. In particular, the double-stranded nucleic acid may include non-nucleotide chemical moieties, e.g. linkers or spacers, at the 5' end of one or both strands. By way of non-limiting example, the double-stranded nucleic acid may include methylated nucleotides, uracil bases, phosphorothioate groups, peptide conjugates etc. Such non-DNA or non-natural modifications may be included in order to confer some desirable property to the nucleic acid, for example to enable covalent, non-covalent or metal-coordination attachment to a solid support, or to act as spacers to position the site of cleavage an optimal distance from the solid support. A single stranded nucleic acid consists of one such polynucleotide strand. Where a polynucleotide strand is only partially hybridised to a complementary strand – for example, a long polynucleotide strand hybridised to a short nucleotide primer – it may still be referred to herein as a single stranded nucleic acid.

A sequence comprising at least a primer-binding sequence (a primer-binding sequence and a sequencing primer binding site, or in another aspect, a combination of a primer-binding sequence, an index sequence and a sequencing primer binding site) may be referred to herein as an adaptor sequence, and an insert is flanked by a 5' adaptor sequence and a 3' adaptor sequence. As used herein the 5' adaptor sequence is the adaptor sequence that is attached to the 5' end of the insert. Similarly, the 3' adaptor sequence is the adaptor sequence that is attached to the 3' end of the insert. The primer-binding sequence may also comprise a sequencing primer for the index read. However, one advantage of one embodiment of the present invention is that a separate sequencing primer for the index read is not needed, reducing the number of reagents/primers needed for a sequencing read.

As used herein, an “adaptor” refers to a sequence that comprises a short sequence-specific oligonucleotide that is ligated to the 5' and 3' ends of each DNA (or RNA) fragment in a sequencing library as part of library preparation. The adaptor sequence may further comprise non-peptide linkers.

In one embodiment, each adaptor sequence comprises at least one cleavable site or a complement of a (or the same) cleavable site. In one embodiment, the 5' adaptor sequence comprises a cleavable site and the 3' adaptor sequence comprises a complement of that cleavable site. This is shown, for example, in Figure 18.

As used herein by “complement of a cleavable site” is meant a sequence that is the complement or substantially complementary to a site (e.g. a sequence) that is cleavable. Unlike the cleavable site, the complement of the cleavable site may not be cleavable. It will be recognised that the adaptor sequences comprise complements of cleavage sites so that when the library strands are used as a template during amplification cleavable site are translated into the copied strand.

In another embodiment, each adaptor sequence comprises at least two cleavable sites or complements of each cleavable site. In one embodiment, the 5' adaptor sequences comprise two cleavable sites and the 3' adaptor sequences comprise two complements of those (same) cleavable sites. This is shown, for example, in Figure 18.

The at least one cleavable site or complement of the cleavable site may be 5' of, 3' of, or within one of the sequences selected from the primer-binding site, the index sequence or the sequencing primer binding site as shown in Figure 18. In one embodiment, the at least one cleavable site is 3' or 5' of the sequencing primer binding site. That is, the at least one cleavable site is between the insert and the sequencing primer binding site.

In one embodiment, the 5' adaptor sequence comprises at least one cleavable site, where the at least one cleavable site is 3' of the sequencing primer binding site. That is, the at least one cleavable site is between the insert and the sequencing primer binding site. In a further embodiment, the 3' adaptor sequence comprises at least one complement of a cleavable site, where the complement of the cleavable site is 5' of the (complements of) the sequencing primer binding sites. That is, the complement of the cleavable site is between the insert and the sequencing primer binding site.

Where the adaptor sequence comprises at least two cleavable sites the first and second sites may be between 10 and 40 bases apart, and again, may be 5' of, 3' of, or within one of the sequences selected from the primer-binding site, the index sequence or the sequencing primer binding site.

In one embodiment, the 5' adaptor comprises a first and second cleavable site, where the first site is 3' of the sequencing primer binding site and the second site is 5' of the sequencing primer binding site. That is, the first and second cleavable sites are either side of the sequencing primer binding site. In another embodiment, the 5' adaptor comprises a first and second cleavable site, where the first site is 3' of the sequencing

primer binding site and the second site is 5' of the index sequence. That is, the cleavable sites are either side of the sequence for the sequencing primer and the index sequence.

5 In another embodiment, the 3' adaptor comprises a first and second cleavable site, where the first site is 5' of the sequencing primer binding site and the second site is 3' of the sequencing primer binding site. That is, the first and second cleavable sites are either side of the sequencing primer binding site. In another embodiment, the 3' adaptor comprises a first and second cleavable site, where the first site is 5' of the sequencing primer binding site and the second site is 3' of the index sequence. That is, the cleavable
10 sites are either side of the sequence for the sequencing primer and the index sequence.

In one embodiment, the cleavable site is a restriction site. This allows the cleavage site to be nicked and allows sequencing to occur starting from the nick location (e.g. in conjunction with a polymerase as described below).

15 As used herein, by "restriction site" is meant a sequence of nucleotides recognised by an endonuclease, such as a single-stranded endonuclease. A restriction site may also be referred to as a "recognition site" or "recognition sequence", and such terms may be used interchangeably.

20 In one embodiment, the endonuclease is a single strand restriction endonuclease, a nicking endonuclease or nicking enzyme or nickase (again, such terms may be used interchangeably). By any of these terms is meant an enzyme that can hydrolyze only one strand of the double-stranded polynucleotide (duplex), to produce DNA molecules that
25 are "nicked", rather than fully cleaved on both strands.

Examples of suitable nicking enzymes that may be used include, but are not limited to, Nb.BbvCI, Nb.BsmI, Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nt.BsmAI, Nt.BspQI, Nt.BstNBI, BssSI, Nb.Bpu101 and Nt.CviPII. These nickases can be used either alone or in various
30 combinations. Other suitable nicking endonucleases are available from commercial sources, including New England Biolabs and Fisher Scientific.

The restriction sites vary depending on the nickase used, and are well known in the art. In one example, the restriction site is selected from the following:

35

In one embodiment, the nickase is Nb.BssSI, and the restriction site is CACGAG, wherein Nb.BssSI catalyzes a single strand break within the recognition sequence.

5 In one embodiment, the nickase is Nt.BspQI, and the restriction site is GCTCTTC(1/-7), wherein Nt.BspQI catalyzes a single strand break one base beyond the 3' side of the restriction site.

10 In one embodiment, the nickase is Nt.CviPII and the restriction site is (0/-1)CCD, wherein Nt.CviPII catalyzes a single strand break at the 5' side of the restriction site.

In one embodiment, the nickase is Nt.BstNBI and the restriction site is GAGTC(4/-5), wherein Nt.BstNBI catalyzes a single strand break four bases beyond the 3' side of the restriction site.

15 In one embodiment, the nickase is Nb.BsrDI and the restriction site is GCAATG, wherein Nb.BsrDI catalyzes a single strand break within the restriction site.

20 In one embodiment, the nickase is Nb.BtsI and the restriction site is GCAGTG, wherein Nb.BtsI catalyzes a single strand break within the restriction site.

In one embodiment, the nickase is Nt.AlwI and the restriction site is GGATC(4/-5), wherein Nt.AlwI catalyzes a single strand break four bases beyond the 3' side of the restriction site.

25 In one embodiment, the nickase is Nb.BbvCI and the restriction site is CCTCAGC, wherein Nb.BbvCI catalyzes a single strand break within the restriction site.

30 In one embodiment, the nickase is Nb.BsmI and the restriction site is GAATGC, wherein Nb.BsmI catalyzes a single strand break within the restriction site.

In one embodiment, the nickase is Nt.BsmAI and the restriction site is GTCTC(1/-5), wherein Nt.BsmAI catalyzes a single strand break one base beyond the 3' side of the restriction site.

35 In one embodiment, the nickase is Nb.Bpu10I and the restriction site is CCTNAGC, wherein Nb.Bpu10I catalyzes a single strand break within the restriction site.

Where the restriction site is described in the following format (x/-y), x is the number of nucleotides beyond (i.e. 3' of) the 3' end of the restriction site where cleavage occurs; and y is the number of nucleotides in the restriction site.

5

In an alternative embodiment, the endonuclease is a Cas9 nickase. Examples of a Cas9 nickase include Cas9 D10A and Cas9 H840A. For example, in one embodiment, the Cas9 protein may comprise the D10A or H840A amino acid substitutions. These nickases cleave only the DNA strand that is complementary to and recognized by a gRNA.

10

In one embodiment, the restriction site may be or may comprise a PAM (protospacer adjacent motif) sequence. Examples of suitable PAM sequences include NGG, NGAG, NGCG, NGN, NG, GAA, GAT, NNG, NGN, NRN, YG, NNGRRT, NNNRRT, NNAGAA, NNNNGATT and NNNNCRAA and complements thereof.

15

In a further embodiment, the Cas9 protein may alternatively or additionally comprise the N863A or N854A amino acid substitutions.

20

In a further embodiment, the Cas9 protein has been modified to improve activity. For example, in one embodiment, the Cas9 protein may additionally comprise a D1135E substitution. Alternatively, the Cas9 protein may also be the VQR variant

Cluster generation and amplification

25

Once a double stranded nucleic acid library is formed, typically, the library is subjected to denaturing conditions to provide single stranded nucleic acids. Suitable denaturing conditions will be apparent to the skilled reader with reference to standard molecular biology protocols (Sambrook et al., 2001, Molecular Cloning, A Laboratory Manual, 4th Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory Press, NY; Current Protocols, eds Ausubel et al). In one embodiment, chemical denaturation may be used.

30

Following denaturation, a single-stranded library may be contacted in free solution onto a solid support comprising surface capture moieties (for example P5 and P7 lawn primers).

35

Thus, embodiments of the present invention may be performed on a solid support 200, such as a flowcell. However, in alternative embodiments, seeding and clustering can be conducted off-flowcell using other types of solid support.

5

The solid support 200 may comprise a substrate 204. An example of a solid support is shown in Figure 6. The substrate 204 comprises at least one well 203 (e.g. a nanowell), and typically comprises a plurality of wells 203 (e.g. a plurality of nanowells).

10

In one embodiment, the solid support comprises a plurality of first immobilised primers and a plurality of second immobilised primers.

15

Thus, each well 203 may comprise a plurality of first immobilised primers 201. In addition, each well 203 may comprise a plurality of second immobilised primers 202. Thus, each well 203 may comprise a plurality of first immobilised primers 201 and a plurality of second immobilised primers 202.

20

The first immobilised primer 201 may be attached via a 5'-end of its polynucleotide chain to the solid support 200. When extension occurs from first immobilised primer 201, the extension may be in a direction away from the solid support 200.

25

The second immobilised primer 202 may be attached via a 5'-end of its polynucleotide chain to the solid support 200. When extension occurs from second immobilised primer 202, the extension may be in a direction away from the solid support 200.

30

The first immobilised primer 201 may be different to the second immobilised primer 202 and/or a complement of the second immobilised primer 202. The second immobilised primer 202 may be different to the first immobilised primer 201 and/or a complement of the first immobilised primer 201.

35

The (or each of the) first immobilised primer(s) 201 may comprise a sequence as defined in SEQ ID NO. 1 or 5, or a variant or fragment thereof. The (or each of the) second immobilised primer(s) 202 may comprise a sequence as defined in SEQ ID NO. 2, or a variant or fragment thereof. Whilst first immobilised primer(s) 201 are shown here to correspond to P5 and second immobilised primer(s) 202 are shown here to correspond to P7, the definitions of these may be swapped – in other words, first immobilised

primer(s) 201 may correspond instead to P7, and second immobilised primer(s) 202 may correspond to P5.

5 In one embodiment the immobilised primers – that is the first or second or the first and second immobilised primers are configured to be cleavable under cleavage conditions. In an embodiment, the immobilised primers may be configured to be cleavable by a thermal trigger (e.g. by heating), a light trigger (e.g. by exposure to ultraviolet light), and/or a chemical/biochemical trigger (e.g. an enzyme, periodate).

10 The location at which the immobilised primers are configured to be cleavable under cleavage conditions may also be referred to as a cleavage site. Accordingly, in one embodiment, the immobilised primers comprise at least one cleavable site. The cleavage site may be at the 3' end of within the sequence of the immobilised primer.

15 In one embodiment, the cleavage site comprises a cleavable covalent bond.

As used herein, the term “cleavable covalent bond” refers to a covalent bond that can be cleaved, for example, under the application of heat, light or other (bio)chemical methods (e.g. by exposure to a degradation agent, such as an enzyme or a catalyst), while a “non-cleavable covalent bond” is stable to degradation under such conditions. Non-limiting
20 examples of cleavable covalent bonds include thermally or photolytically cleavable cycloadducts (e.g. furan-maleimide cycloadducts), alkenylene linkages, esters, amides, acetals, hemiaminal ethers, amins, imines, hydrazones, 1,2-diol linkages (e.g. glycols cleavable by periodates), polysulfide linkages (e.g. disulfide linkages), boron-based
25 linkages (e.g. boronic and borinic acids/esters), silicon-based linkages (e.g. silyl ether, siloxane), and phosphorus-based linkages (e.g. phosphite, phosphate) linkages.

In an embodiment, the immobilised primers may be configured to be cleavable by a glycosylase. In other words, the cleavage conditions involve exposure to a glycosylase.
30 In a further embodiment, the immobilised primers may be configured to be cleavable by a glycosylase that recognises any nitrogenous base (e.g. purine or pyrimidine) which is not selected from guanine (G), cytosine (C), adenine (A) and thymine (T) when the first immobilised primer is a DNA sequence; or the immobilised primers may be configured to be cleavable by a glycosylase that recognises any nitrogenous base (e.g. purine or
35 pyrimidine) which is not selected from guanine (G), cytosine (C), adenine (A) and uracil (U) when the immobilised primer is an RNA sequence. In other words, the glycosylase

may recognise an unnatural nucleobase (i.e. one which is not usually present in a typical DNA sequence an RNA sequence). Examples of unnatural nucleobases may include oxoguanine (e.g. 8-oxoguanine), hypoxanthine, xanthine, methylguanines (e.g. O⁶-methylguanine, N⁷-methylguanine), methyladenines (e.g. 3-methyladenine, N⁶-methyladenine), modified cytosines including methylcytosines (e.g. 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxylcytosine), dihydrouracil, and uracil (if the first immobilised primer is a DNA sequence).

In an embodiment, the immobilised primers may be configured to be cleavable by a uracil glycosylase (when the immobilised primer is a DNA sequence) or an oxoguanine glycosylase (e.g. 8-oxoguanine glycosylase); and in a further embodiment, an oxoguanine glycosylase (e.g. 8-oxoguanine glycosylase).

In an embodiment, each immobilised primer that is cleavable may comprise a nucleobase which is not selected from guanine, cytosine, adenine or thymine when the immobilised primer is a DNA sequence, or wherein each immobilised primer that is cleavable may comprise a nucleobase which is not selected from guanine, cytosine, adenine or uracil when the immobilised primer is an RNA sequence. In other words, each immobilised primer that is cleavable may comprise an unnatural nucleobase (i.e. one which is not usually present in a typical DNA sequence an RNA sequence). As mentioned above, examples of unnatural nucleobases may include oxoguanine (e.g. 8-oxoguanine), hypoxanthine, xanthine, methylguanines (e.g. O⁶-methylguanine, N⁷-methylguanine), methyladenines (e.g. 3-methyladenine, N⁶-methyladenine), modified cytosines including methylcytosines (e.g. 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxylcytosine), dihydrouracil, and uracil (if the first immobilised primer is a DNA sequence).

In a further embodiment, each immobilised primer that is cleavable may comprise oxoguanine (e.g. 8-oxoguanine) or uracil when the immobilised primer is a DNA sequence, or wherein each immobilised primer that is cleavable may comprise oxoguanine (e.g. 8-oxoguanine) when the immobilised primer is an RNA sequence; and in an even further embodiment, wherein immobilised primer that is cleavable may comprise oxoguanine (e.g. 8-oxoguanine) when the immobilised primer is a DNA sequence.

In another embodiment the cleavable site is a restriction site. A restriction site is already defined above.

By way of brief example, following attachment of the P5 and P7 primers to the solid support, the solid support may be contacted with the template to be amplified under conditions which permit hybridisation (or annealing – such terms may be used interchangeably) between the template and the immobilised primers. The template is usually added in free solution under suitable hybridisation conditions, which will be apparent to the skilled reader. Typically, hybridisation conditions are, for example, 5xSSC at 40°C. However, other temperatures may be used during hybridisation, for example about 50°C to about 75°C, about 55°C to about 70°C, or about 60°C to about 65°C. Solid-phase amplification can then proceed. The first step of the amplification is a primer extension step in which nucleotides are added to the 3' end of the immobilised primer using the template to produce a fully extended complementary strand. The template is then typically washed off the solid support. The complementary strand will include at its 3' end a primer-binding sequence (i.e. either P5' or P7') which is capable of bridging to the second primer molecule immobilised on the solid support and binding. Further rounds of amplification (analogous to a standard PCR reaction) leads to the formation of clusters or colonies of template molecules bound to the solid support. This is called clustering.

Thus, solid-phase amplification by either a method analogous to that of WO 98/44151 or that of WO 00/18957 (the contents of which are incorporated herein in their entirety by reference) will result in production of a clustered array comprised of colonies of "bridged" amplification products. This process is known as bridge amplification. Both strands of the amplification products will be immobilised on the solid support at or near the 5' end; this attachment being derived from the original attachment of the amplification primers. Typically, the amplification products within each colony will be derived from amplification of a single template molecule. Other amplification procedures may be used, and will be known to the skilled person. For example, amplification may be isothermal amplification using a strand displacement polymerase; or may be exclusion amplification as described in WO 2013/188582. Further information on amplification can be found in WO 02/06456 and WO 07/107710, the contents of which are incorporated herein in their entirety by reference.

Through such approaches, a cluster of template molecules is formed, comprising copies of a template strand and copies of the complement of the template strand.

5 The steps of cluster generation and amplification for templates including a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion, are illustrated below and in Figure 4.

10 In cases where (separate) polynucleotide strands are used, each first polynucleotide sequence may be attached (via the 5'-end of the first polynucleotide sequence) to a first immobilised primer, and wherein each second polynucleotide sequence is attached (via the 5'-end of the second polynucleotide sequence) to a second immobilised primer. Each first polynucleotide sequence may comprise a second adaptor sequence, wherein the second adaptor sequence comprises a portion, which is substantially complementary to the second immobilised primer (or is substantially complementary to the second
15 immobilised primer). The second adaptor sequence may be at a 3'-end of the first polynucleotide sequence. Each second polynucleotide sequence may comprise a first adaptor sequence, wherein the first adaptor sequence comprises a portion, which is substantially complementary to the first immobilised primer (or is substantially complementary to the first immobilised primer). The first adaptor sequence may be at a
20 3'-end of the second polynucleotide sequence.

In an embodiment, a solution comprising a polynucleotide library prepared by ligating adaptor sequences to double-stranded polynucleotide sequences as described above may be flowed across a flowcell.

25

A particular polynucleotide strand from the polynucleotide library to be sequenced comprising, in a 5' to 3' direction, a second primer-binding complement sequence 302 (e.g. P7), a first terminal binding site complement 303' (e.g. SBS12), a forward strand of the sequence 101, a second terminal sequencing primer binding site 304 (e.g. SBS3') and a first primer-binding sequence 301' (e.g. P5'), may anneal (via the first primer-binding sequence 301') to the first immobilised primer 201 (e.g. P5 lawn primer) located within a particular well 203 (Figure 4A).

30

The polynucleotide library may comprise other polynucleotide strands with different forward strands of the sequence 101. Such other polynucleotide strands may anneal to corresponding first immobilised primers 201 (e.g. P5 lawn primers) in different wells 203,

35

thus enabling parallel processing of the various different strands within the polynucleotide library.

5 A new polynucleotide strand may then be synthesised, extending from the first immobilised primer 201 (e.g. P5 lawn primer) in a direction away from the substrate 204. By using complementary base-pairing, this generates a template strand comprising, in a 5' to 3' direction, the first immobilised primer 201 (e.g. P5 lawn primer) which is attached to the solid support 200, a second terminal sequencing primer binding site complement 304' (e.g. SBS3), a forward strand of the template 101' (which represents a type of "first
10 portion"), a first terminal sequencing primer binding site 303 (which represents a type of "first sequencing primer binding site") (e.g. SBS12'), and a second primer-binding sequence 302' (e.g. P7') (Figure 4B). Such a process may utilise an appropriate polymerase, such as a DNA or RNA polymerase.

15 If the polynucleotides in the library comprise index sequences, then corresponding index sequences are also produced in the template.

Furthermore, where the adaptors comprise at least one, or at least two cleavage sites – or complements of cleavage sites - as described above - these will be incorporated into
20 the newly synthesised immobilised strands. Where the adaptors comprise complements of cleavage sites therefore, for example complements of the above restriction sites, these will be translated into cleavage sites (e.g. restriction sites) in the newly synthesised immobilised strands.

25 The polynucleotide strand from the polynucleotide library may then be dehybridised and washed away, leaving a template strand attached to the first immobilised primer 201 (e.g. P5 lawn primer) (Figure 4C).

The second primer-binding sequence 302' (e.g. P7') on the template strand may then
30 anneal to a second immobilised primer 202 (e.g. P7 lawn primer) located within the well 203. This forms a "bridge" or "sequence bridge" (Figure 4D).

A new polynucleotide strand may then be synthesised by bridge amplification, extending
35 from the second immobilised primer 202 (e.g. P7 lawn primer) (initially) in a direction away from the substrate 204. By using complementary base-pairing, this generates a template strand comprising, in a 5' to 3' direction, the second immobilised primer 202

(e.g. P7 lawn primer) which is attached to the solid support 200, a first terminal sequencing primer binding site complement 303' (e.g. SBS12), a forward complement strand of the template 101 (which represents a type of "second portion"), a second terminal sequencing primer binding site 304 (which represents a type of "second sequencing primer binding site") (e.g. SBS3'), and a first primer-binding sequence 301' (e.g. P5') (Figure 4E). Again, such a process may utilise a suitable polymerase, such as a DNA or RNA polymerase.

The strand attached to the second immobilised primer 202 (e.g. P7 lawn primer) may then be dehybridised from the strand attached to the first immobilised primer 201 (e.g. P5 lawn primer) (Figure 4F).

A subsequent bridge amplification cycle can then lead to amplification of the strand attached to the first immobilised primer 201 (e.g. P5 lawn primer) and the strand attached to the second immobilised primer 202 (e.g. P7 lawn primer). Similar to Figure 4D, the second primer-binding sequence 302' (e.g. P7') on the template strand attached to the first immobilised primer 201 (e.g. P5 lawn primer) may then anneal to another second immobilised primer 202 (e.g. P7 lawn primer) located within the well 203. In a similar fashion, the first primer-binding sequence 301' (e.g. P5') on the template strand attached to the second immobilised primer 202 (e.g. P7 lawn primer) may then anneal to another first immobilised primer 201 (e.g. P5 lawn primer) located within the well 203 (Figure 4G).

Completion of bridge amplification and dehybridisation may then provide an amplified (duoclonal) cluster, thus providing a plurality of first polynucleotide sequences comprising the forward strand of the template 101' (i.e. "first portions"), and a plurality of second polynucleotide sequences comprising the forward complement strand of the template 101 (i.e. "second portions") (Figure 4H).

If desired, further bridge amplification cycles may be conducted to increase the number of first polynucleotide sequences and second polynucleotide sequences within the well 203 (Figure 4I).

In this particular example, the "first portion" corresponds with the forward strand of the template 101', and the "second portion" corresponds with the forward complement strand of the template 101.

However, other set-ups may be obtained by changing the library used. For example, by using a loop fork method to prepare a library, a portion at or close to the loop (or the loop complement) may be cleaved (e.g. by nicking). In these cases, the loop may comprise a cleavage site (e.g. a restriction recognition site, a cleavable linker, a modified nucleotide, or the like). By conducting cleavage at the loop, it is possible to produce a well 203, where the “first portion” corresponds with a forward strand of the template, and the “second portion” corresponds with a reverse complement strand of the template. In addition, by using a PCR stitching method to prepare a library, a portion at or close to the overlap region may comprise a cleavage site (e.g. a restriction recognition site, a cleavable linker, a modified nucleotide, or the like). By conducting cleavage at the overlap region, it is possible to produce a well 203, where the “first portion” corresponds with a first insert sequence, and the “second portion” corresponds with a second insert sequence that is genetically unrelated to the first insert sequence. As such, different types of strands for the “first portions” and “second portions” may be prepared for templates including a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion, and as such the forward strand of the template 101' and the forward complement strand of the template 101 may be substituted as appropriate.

Thus, one strand of a concatenated polynucleotide within a polynucleotide library may comprise, in a 5' to 3' direction, a second primer-binding complement sequence 302 (e.g. P7), a first terminal sequencing primer binding site complement 303' (e.g. B15-ME; or if ME is not present, then B15), a first insert sequence 401, a hybridisation complement sequence 403 (e.g. ME'-HYB2-ME; or if ME' and ME are not present, then HYB2), a second insert sequence 402, a second terminal sequencing primer binding site 304 (e.g. ME'-A14'; or if ME' is not present, then A14'), and a first primer-binding sequence 301' (e.g. P5') (Figures 16 and 17 – bottom strand).

Although not shown in Figures 16 and 17, the strand may further comprise one or more index sequences. As such, a first index sequence (e.g. i7) may be provided between the second primer-binding complement sequence 302 (e.g. P7) and the first terminal sequencing primer binding site complement 303' (e.g. B15-ME; or if ME is not present, then B15). Separately, or in addition, a second index complement sequence (e.g. i5') may be provided between the second terminal sequencing primer binding site 304 (e.g. ME'-A14') and the first primer-binding sequence 301' (e.g. P5'). Thus, in some embodiments, one strand of a polynucleotide within a polynucleotide library may

comprise, in a 5' to 3' direction, a second primer-binding complement sequence 302 (e.g. P7), a first index sequence (e.g. i7), a first terminal sequencing primer binding site complement 303' (e.g. B15-ME; or if ME is not present, then B15), a first insert sequence 401, a hybridisation complement sequence 403 (e.g. ME'-HYB2-ME; or if ME' and ME are not present, then HYB2), a second insert sequence 402, a second terminal sequencing primer binding site 304 (e.g. ME'-A14'; or if ME' is not present, then A14'), a second index complement sequence (e.g. i5'), and a first primer-binding sequence 301' (e.g. P5')

Another strand of a concatenated polynucleotide within a polynucleotide library may comprise, in a 5' to 3' direction, a first primer-binding complement sequence 301 (e.g. P5), a second terminal sequencing primer binding site complement 304' (e.g. A14-ME; or if ME is not present, then A14), a second insert complement sequence 402', a hybridisation sequence 403' (e.g. ME'-HYB2'-ME; or if ME' and ME are not present, then HYB2'), a first insert complement sequence 401', a first terminal sequencing primer binding site 303 (e.g. ME'-B15'; or if ME' is not present, then B15'), and a second primer-binding sequence 302' (e.g. P7') (Figures 16 and 17 – top strand).

Although not shown in Figures 16 and 17, the another strand may further comprise one or more index sequences. As such, a second index sequence (e.g. i5) may be provided between the first primer-binding complement sequence 301 (e.g. P5) and the second terminal sequencing primer binding site complement 304' (e.g. A14-ME; or if ME is not present, then A14). Separately, or in addition, a first index complement sequence (e.g. i7') may be provided between the first terminal sequencing primer binding site 303 (e.g. ME'-B15'; or if ME' is not present, then B15') and the second primer-binding sequence 302' (e.g. P7'). Thus, in some embodiments, another strand of a polynucleotide within a polynucleotide library may comprise, in a 5' to 3' direction, a first primer-binding complement sequence 301 (e.g. P5), a second index sequence (e.g. i5), a second terminal sequencing primer binding site complement 304' (e.g. A14-ME; or if ME is not present, then A14).), a second insert complement sequence 402', a hybridisation sequence 403' (e.g. ME'-HYB2'-ME; or if ME' and ME are not present, then HYB2'), a first insert complement sequence 401', a first terminal sequencing primer binding site 303 (e.g. ME'-B15'; or if ME' is not present, then B15'), a first index complement sequence (e.g. i7'), and a second primer-binding sequence 302' (e.g. P7').

As described herein, the first insert sequence 401 and the second insert sequence 402 may comprise different types of library sequences.

5 In one embodiment, the first insert sequence 401 may be different to the second insert sequence 402 (e.g. genetically unrelated, and/or obtained from different sources), for example where the library is prepared using PCR stitching.

10 In another embodiment, the first insert sequence 401 may comprise a forward strand of the sequence 101, and the second insert sequence may comprise a reverse complement strand of the sequence 102' (or the first insert sequence 401 may comprise a reverse strand of the sequence 102, and the second insert sequence 402 may comprise a forward complement strand of the sequence 101'), for example where the library is prepared using a tandem insert method.

15 In another embodiment, the first insert sequence 401 may comprise a forward strand of the sequence 101, and the second insert sequence 402 may comprise a reverse strand of the sequence 102 (or the first insert sequence 401 may comprise a forward complement strand of the sequence 101', and the second insert sequence 402 may comprise a reverse complement strand of the sequence 102'), for example where the
20 library is prepared using a loop fork method.

Methods of SBS sequencing

25 The present invention is directed to methods of SBS sequencing, and in particular double-stranded SBS sequencing. This leads to reductions in error rates – i.e. improvements in accuracy.

30 In one embodiment the method of the invention comprises a first sequencing read that sequences the first polynucleotide strand and a second sequencing read that sequences a second polynucleotide strand, where the first and second reads are temporally separated – that is, the first sequencing read is before the second sequencing read.

35 In another embodiment, the method of the invention comprises altering the ratio of first to second sequencing reads, which in turn allows the sequencing reads to be differentiated, and allows the first and second polynucleotide strands to be sequenced simultaneously.

Advantageously, the methods of the present invention also enable SBS sequencing without the need for a paired-end turn and cluster re-synthesis. This in turn reduces the time taken to sequence a target polynucleotide, thus improving even further the efficiency of the sequencing protocol when using the method of the present invention. A paired-end turn refers to the sequence of stages required to effectively invert the sequence for the second read in paired-end reading, after sequencing read 1 (Figure 5A). The paired-end turn may be facilitated by a cycle of bridge amplification and linearization. The present invention also eliminates the need for cluster re-synthesis of the first or second polynucleotide sequences for read 2 of conventional SBS workflows (Figure 5A). Cluster re-synthesis follows the paired-end turn, and consists of cycles of bridge amplification to restore the clusters of library preparations for the second read (Figure 5A). Thus, as described above, eliminating these steps provide additional time savings. For the simultaneous sequencing with 16 QaM, the sequencing time can be almost halved.

As another advantage, the method of the invention also requires the redesigning of only the amplification primers, which also simplifies subsequent amplification and sequencing steps, as it is possible to rely on normal amplification and sequencing methods (such as those described above), rather than specifically designed amplification or sequencing methods to achieve double-stranded SBS sequencing, and in particular concurrent double-stranded SBS sequencing.

As a yet further advantage, the method of the invention also simplifies the immobilised primers. The immobilised primers of the present invention do not need to be cleavable from the solid support. Accordingly, in one embodiment, the immobilised primers may not comprise a cleavable site. In another embodiment, where the immobilised primers comprise a cleavable site, this cleavable site is different from or cleavable under different cleavable conditions from the cleavable sites present in the first and/or second sequencing primer. In another embodiment, the immobilised primers may not comprise a uracil or oxoguanine (e.g. 8-oxoguanine).

Accordingly, we describe a method of identifying a target polynucleotide, the method comprising

- a. hybridising the first polynucleotide sequence to first immobilised primers on a solid support and hybridising the second polynucleotide sequence to second immobilised primers on a solid support;

- b. synthesising a plurality of first and second polynucleotide sequences by conducting an amplification reaction to extend the first and second immobilised primers;
- c. selectively removing first and second immobilised primers that have not been extended;
- d. applying a plurality of first and second sequencing primers, wherein the first sequencing primers comprise a first cleavable site and wherein the second sequencing primers comprise a second cleavable site and conducting an amplification reaction to extend the first and second sequencing primers; and
- e. selectively cleaving the first cleavable site in the first sequencing primer and/or selectively cleaving the second cleavable site in the second sequencing primer.

5

10

15 Each step is described in further detail below.

In one embodiment the plurality of first and second polynucleotide sequences is synthesised by bridge or exclusion amplification, as described above. This lead to the formation of clusters or colonies of template molecules that are bound to the solid support via the immobilised primers. Accordingly, in one embodiment, synthesising a plurality of first and second polynucleotide sequences by conducting an amplification reaction forms a cluster or plurality of clusters. As described above, bridge amplification may be performed a number of times (i.e. a number of cycles are carried out) to obtain a cluster of first and second polynucleotide sequences.

20

25

In one embodiment, the excess solid support sequence(s) are removed after clustering. By excess solid support sequence(s) it is meant the immobilised primer sequences (e.g. P5, P7') that have not been extended during clustering stages. Removing excess immobilised primers may serve to prevent further bridge amplification in the subsequent amplification stage described below.

30

The excess immobilised primers are preferably removed by an exonuclease. Preferably the exonuclease is a single-strand exonuclease. Non-limiting examples of exonucleases include RecJf, Lambda Exonuclease, T7 exonuclease domain, T5 exonuclease, or the DNA polymerase I-like H3TH domain, Exonuclease V (RecBCD) or Exonuclease VII.

35

Other suitable methods of removing a single-stranded DNA sequence known in the art, may also be used to remove the excess immobilised primers.

5 Accordingly, in one embodiment, the method of preparing a first and second polynucleotide sequence for sequencing comprises removing excess immobilised primers from the solid surface, such as using an exonuclease, and optionally a single-stranded exonuclease.

10 Following removal of the excess immobilised sequences, the library is linearised. Linearisation can occur as only one end of each side of the “bridge” formed during bridge amplification is attached (e.g. covalently bonded) to the solid support. The other side of the bridge is hybridized to its’ complementary immobilized primer. As shown in Figure 13C and 14C, linearization leads to a plurality of single-stranded first and second polynucleotide sequences immobilised on the solid support through the first and second
15 immobilised primers. We may refer to these as first immobilised polynucleotide strands and second immobilised polynucleotide strands.

Once linearised, the first and second polynucleotides may be subjected to further amplification using first and second sequencing primers.

20 Accordingly, in one embodiment, the method comprises applying (i.e. flowing across the surface of the solid support) a plurality of first and second sequencing primers. The sequencing primers must be sufficiently complementary to the sequencing primer-binding sites within the first and second polynucleotide strands to allow hybridisation. In
25 one embodiment, the first and second sequencing primers bind to the first and second polynucleotide strands at first and second sequencing primer-binding sites respectively.

In one embodiment, and as shown in Figure 13C and 14C, the first and second sequencing primers bind (i.e. substantially hybridise) to sequencing primer-binding sites
30 at the 3’ end of the immobilised first and/or second polynucleotide sequences.

As described above, the first and second sequencing primers comprise at least one cleavable site. Cleavage of the cleavable site occurs under cleavage conditions, i.e. reaction conditions that cause cleavage, and results in a nick or gap within the
35 sequencing primer. This cleaved site acts as the initiation site for sequencing by nick-

translation, which in particular may be carried out using a polymerase and a 5' to 3' exonuclease or a polymerase with 5' to 3' activity, as described below.

5 As used herein, the term "first cleavage conditions" refers to reaction conditions that cause cleavage within the first only, or the first and second sequence primer (i.e. at the first or first and second cleavage site). As used herein, the term "second cleavage conditions" refers to reaction conditions that cause cleavage within the second and/or subsequent sequence primers (i.e. the second and/or subsequent cleavage site). In one embodiment, the method comprises exposing the immobilised sequencing primers to
10 first and/or second cleavage conditions.

The location at which the first sequencing primer is configured to be cleavable under first cleavage conditions may also be referred to as a first cleavage site. As such, the first sequencing primer comprises a first cleavage site. The first cleavage site may comprise
15 a cleavable covalent bond. In some cases, when the first cleavage site is nicked, this allows sequencing to occur starting from the nick location (e.g. in conjunction with a polymerase and exonuclease).

The location at which the second sequencing primer is configured to be cleavable under
20 second cleavage conditions may also be referred to as a second cleavage site. As such, the second sequencing primer comprises a second cleavage site. The second cleavage site may comprise a cleavable covalent bond. In some cases, when the second cleavage site is nicked, this allows sequencing to occur starting from the nick location (e.g. in conjunction with a polymerase and exonuclease).

25 Accordingly, in one embodiment, the method comprises selectively cleaving the first cleavable site in the first sequencing primer and/or selectively cleaving the second cleavable site in the second sequencing primer.

30 In an embodiment, the cleavable site may be at the 5' end, the 3' end or within the sequencing primer; in particular, the cleavable site may be at the 3' end of the sequencing primer.

35 As used herein, the term "cleavable covalent bond" refers to a covalent bond that can be cleaved for example under the application of heat, light or other (bio)chemical methods (e.g. by exposure to a degradation agent, such as an enzyme or a catalyst), while a "non-

cleavable covalent bond” is stable to degradation under such conditions. Non-limiting examples of cleavable covalent bonds include thermally or photolytically cleavable cycloadducts (e.g. furan-maleimide cycloadducts), alkenylene linkages, esters, amides, acetals, hemiaminal ethers, amins, imines, hydrazones, 1,2-diol linkages (e.g. glycols
5 cleavable by periodates), polysulfide linkages (e.g. disulfide linkages), boron-based linkages (e.g. boronic and borinic acids/esters), silicon-based linkages (e.g. silyl ether, siloxane), and phosphorus-based linkages (e.g. phosphite, phosphate) linkages.

The first and/or second cleavage conditions may involve exposure to a thermal trigger
10 (e.g. by heating), a light trigger (e.g. by exposure to ultraviolet light), and/or a chemical/biochemical trigger (e.g. an enzyme, periodate).

Accordingly, in an embodiment, a proportion of first and/or second sequencing primers may be configured to be cleavable by a thermal trigger (e.g. by heating), a light trigger
15 (e.g. by exposure to ultraviolet light), and/or a chemical/biochemical trigger (e.g. an enzyme, periodate).

In another embodiment, the first cleavage conditions and the second cleavage conditions are different. This allows control over which sequencing primer is cleaved first, which in
20 turn allows sequential sequencing of the first and second polynucleotide sequence of vice versa.

In an embodiment, the first cleavage conditions and the second cleavage conditions may be the same. This allows the first and second sequencing primer to be simultaneously
25 sequenced, which in turn allows concurrent sequencing of the first and second polynucleotide sequence. In this embodiment, a portion of the first sequencing primers or the second sequencing primers comprise a cleavable site. That is, in one embodiment, the method comprises applying a plurality of first sequencing primers and second sequencing primers, where the first or second sequencing primers comprise a mix of
30 cleavable sequencing primers and un-cleavable primers. By “cleavable” is meant that the sequencing primers comprise at least one cleavable site. By “un-cleavable” or “non-cleavable” is meant that the sequencing primers do not comprise a cleavable site.

In an embodiment, the proportion of first sequencing primers cleavable relative to a
35 proportion of first sequencing primers which are not cleavable may be between 20:80 to

80:20; in a further embodiment, between 1:2 to 2:1; and in an even further embodiment, about 1:1.

5 In an embodiment, the proportion of second sequencing primers cleavable relative to a proportion of second sequencing primers which are not cleavable may be between 20:80 to 80:20; in a further embodiment, between 1:2 to 2:1; and in an even further embodiment, about 1:1.

10 The sequence of the sequencing primers and the sequence primer binding sites are immaterial to the methods of the invention, as long as the sequencing primers are able to bind to the sequence primer-binding site to enable amplification and sequencing of the regions to be identified. However, in an embodiment, each first sequencing primer may comprise a sequence selected from one of SEQ ID NOs. 7 to 10 or a variant or fragment thereof; and each second sequencing primer may comprise a different
15 sequence selected from SEQ ID NOs 7 to 10, or a variant or fragment thereof.

In one embodiment, the cleavable site may comprise one or more unnatural nucleobase (i.e. one which is not usually present in a typical DNA sequence or an RNA sequence). As mentioned above, examples of unnatural nucleobases may include oxoguanine (e.g. 8-oxoguanine), hypoxanthine, xanthine, methylguanines (e.g. O⁶-methylguanine, N⁷-methylguanine), methyladenines (e.g. 3-methyladenine, N⁶-methyladenine), modified cytosines including methylcytosines (e.g. 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxylcytosine), dihydrouracil, and uracil (if the sequence primer is a DNA sequence).

25

In an embodiment, a proportion of first and/or second sequencing primers may be configured to be cleavable by a glycosylase. In other words, the cleavage conditions involve exposure to a glycosylase or a glycosylase and an AP lyase or endonuclease

30 In a further embodiment, a proportion of first and/or second sequencing primers may be configured to be cleavable by a glycosylase that recognises any nitrogenous base (e.g. purine or pyrimidine) which is not selected from guanine (G), cytosine (C), adenine (A) and thymine (T) when the first and/or second sequencing primer is a DNA sequence; or the proportion of sequence primers may be configured to be cleavable by a glycosylase
35 that recognises any nitrogenous base (e.g. purine or pyrimidine) which is not selected from guanine (G), cytosine (C), adenine (A) and uracil (U) when the first and/or second

sequencing primer is an RNA sequence. In other words, the glycosylase may recognise an unnatural nucleobase (i.e. one which is not usually present in a typical DNA sequence or an RNA sequence). Examples of unnatural nucleobases may include oxoguanine (e.g. 8-oxoguanine), hypoxanthine, xanthine, methylguanines (e.g. O6-methylguanine, N7-methylguanine), methyladenines (e.g. 3-methyladenine, N6-methyladenine), modified cytosines including methylcytosines (e.g. 5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, 5-carboxylcytosine), dihydrouracil, and uracil (if the first sequencing primer is a DNA sequence).

The skilled person will understand that, where an unnatural nucleobase is used as a cleavage site, the location of the unnatural nucleobase will be sufficiently within the sequencing primer to allow for both hybridisation of the sequencing primer to the sequencing primer binding site on the immobilised strand and for the resulting nicked sequence primer to be sufficiently long enough to be extended by a polymerase.

In an embodiment, a proportion of sequence primers may be configured to be cleavable by a uracil glycosylase (when the first sequencing primer is a DNA sequence) or an oxoguanine glycosylase (e.g. 8-oxoguanine glycosylase); and in a further embodiment, an oxoguanine glycosylase (e.g. 8-oxoguanine glycosylase).

As used herein, the term “glycosylase” may refer to an enzyme which catalyses the removal of a nitrogenous base from one of the nucleotides in a (poly)nucleotide chain by breaking a N-glycosidic bond, resulting in the formation of an apurinic/aprimidinic site (AP site). An AP lyase or an endonuclease may then be used to digest the remaining phosphodiester 3' bond in the polynucleotide, to create a nick. Alternatively, the phosphodiester backbone will be cleaved in an uncatalysed manner by elimination.

Accordingly, as used herein, a “glycosylase” refers to a nickase when used alone or in conjunction with an AP lyase or an endonuclease.

For DNA chains, the glycosylase may recognise any nitrogenous base (e.g. purine or pyrimidine) which is not selected from guanine (G), cytosine (C), adenine (A) and thymine (T); for RNA chains, the glycosylase may recognise any nitrogenous base (e.g. purine or pyrimidine) which is not selected from guanine (G), cytosine (C), adenine (A) and uracil (U). Examples of typical nitrogenous bases recognised by glycosylases include oxoguanine (e.g. 8-oxoguanine), uracil, inosine and alkylpurines.

Glycosylases may be monofunctional, such that they only possess glycosylase activity (i.e. breaking of the N-glycosidic bond) – cleavage of a phosphodiester bond in the sugar-phosphate backbone may then occur in an uncatalysed manner by elimination.

5 Alternatively, an AP lyase or endonuclease may be used in conjunction with a glycosylase cleave the remaining phosphodiester bond to create a nick. Other glycosylases may be bifunctional, such that they also possess AP lyase activity by catalysing the phosphodiester bond of the (poly)nucleotide chain, for example FpG enzyme. In an embodiment, the glycosylase is bifunctional (i.e. possesses both
10 glycosylase and AP lyase activity).

In another embodiment, the cleavable site may comprise inosine (e.g. deoxyinosine) and is recognised by an endonuclease (e.g. endonuclease V) to nick at the inosine residue.

15 In another embodiment, the cleavable site may comprise uracil (e.g. deoxyuracil), and is recognised by a uracil DNA glycosylase and then endonuclease VIII. This combination is called the USER enzyme mix.

If the USER enzyme mix (comprising uracil glycosylase and endonuclease VIII) or FpG
20 enzyme (a dual-function N-glycosylase and AP-lyase) is used as to introduce the cleavage site, a 3' phosphate group will remain. Therefore, a treatment with a kinase, (e.g. T4 kinase) will be required to release the 3' OH group. Kinases (e.g. T4 kinase) are well-known in the art and are widely available commercially.

25 In one embodiment, the cleavable site may comprise oxoguanine (e.g. 8-oxoguanine). In another embodiment, the cleavable site may comprise uracil, particularly when the sequencing primer is a DNA sequence.

In one embodiment, the cleavable site is an AP site and is recognised by an AP lyase.
30 The AP site may be due to a site modification within a nucleotide where the nitrogenous base has been removed. The AP lyase will recognise the AP site and cleave the phosphodiester bond to create a nick.

In another embodiment, the cleavable site is a restriction site. This allows the cleavage
35 site to be nicked and allows sequencing to occur starting from the nick location (e.g. in conjunction with a polymerase as described below).

As used herein, by "restriction site" is meant a sequence of nucleotides recognised by an endonuclease, such as a single-stranded endonuclease. A restriction site may also be referred to as a "recognition site" or "recognition sequence", and such terms may be used interchangeably.

Accordingly, in one embodiment, the method comprises cleaving the first and/or second sequencing primer by applying (i.e. flowing over the solid support) an endonuclease.

In one embodiment, the endonuclease is a single strand restriction endonuclease, a nicking endonuclease or nicking enzyme or nickase (again, such terms may be used interchangeably). By any of these terms is meant an enzyme that can hydrolyze only one strand of the double-stranded polynucleotide (duplex), to produce DNA molecules that are "nicked", rather than fully cleaved on both strands.

Examples of suitable nicking enzymes that may be used include, but are not limited to, Nb.BbvCI, Nb.BsmI, Nb.BsrDI, Nb.BtsI, Nt.AlwI, Nt.BsmAI, Nt.BspQI, Nt.BstNBI, BssSI, Nb.Bpu101 and Nt.CviPII. These nickases can be used either alone or in various combinations.

The restriction sites vary depending on the nickase used, and are well known in the art. In one example, the restriction site is selected from the following:

In one embodiment, the nickase is Nb.BssSI, and the restriction site is CACGAG, wherein Nb.BssSI catalyzes a single strand break within the recognition sequence.

In one embodiment, the nickase is Nt.BspQI, and the restriction site is GCTCTTC(1/-7), wherein Nt.BspQI catalyzes a single strand break one base beyond the 3' side of the restriction site. To use this nickase, the restriction site must be at least 1 base from the 3' end of the sequencing primer to enable nicking.

In one embodiment, the nickase is Nt.CviPII and the restriction site is (0/-1)CCD, wherein Nt.CviPII catalyzes a single strand break at the 5' side of the restriction site.

In one embodiment, the nickase is Nt.BstNBI and the restriction site is GAGTC(4/-5), wherein Nt.BstNBI catalyzes a single strand break four bases beyond the 3' side of the restriction site.

5 In one embodiment, the nickase is Nb.BsrDI and the restriction site is GCAATG, wherein Nb.BsrDI catalyzes a single strand break within the restriction site.

In one embodiment, the nickase is Nb.BtsI and the restriction site is GCAGTG, wherein Nb.BtsI catalyzes a single strand break within the restriction site.

10

In one embodiment, the nickase is Nt.AlwI and the restriction site is GGATC(4/-5), wherein Nt.AlwI catalyzes a single strand break four bases beyond the 3' side of the restriction site.

15 In one embodiment, the nickase is Nb.BbvCI and the restriction site is CCTCAGC, wherein Nb.BbvCI catalyzes a single strand break within the restriction site.

In one embodiment, the nickase is Nb.BsmI and the restriction site is GAATGC, wherein Nb.BsmI catalyzes a single strand break within the restriction site.

20

In one embodiment, the nickase is Nt.BsmAI and the restriction site is GTCTC(1/-5), wherein Nt.BsmAI catalyzes a single strand break one base beyond the 3' side of the restriction site.

25 In one embodiment, the nickase is Nb.Bpu10I and the restriction site is CCTNAGC, wherein Nb.Bpu10I catalyzes a single strand break within the restriction site.

Where the restriction site is described in the following format (x/-y), x is the number of nucleotides beyond (i.e. 3' of) the 3' end of the restriction site where cleavage occurs;
30 and y is the number of nucleotides in the restriction site.

In an alternative embodiment, the endonuclease is a Cas9 nickase. Examples of a Cas9 nickase include Cas9 D10A and Cas9 H840A. For example, in one embodiment, the Cas9 protein may comprise the D10A or H840A amino acid substitutions. These
35 nickases cleave only the DNA strand that is complementary to and recognized by a gRNA.

In one embodiment, the restriction site may be or may comprise a PAM (protospacer adjacent motif) sequence. Examples of suitable PAM sequences include NGG, NGAG, NGCG, NGN, NG, GAA, GAT, NNG, NGN, NRN, YG, NNGRRT, NNNRRT, NNAGAA, NNNNGATT and NNNNCRAA and complements thereof.

In a further embodiment, the Cas9 protein may alternatively or additionally comprise the N863A or N854A amino acid substitutions.

In a further embodiment, the Cas9 protein has been modified to improve activity. For example, in one embodiment, the Cas9 protein may additionally comprise a D1135E substitution. Alternatively, the Cas9 protein may also be the VQR variant.

Following cleavage of the first and/or second sequencing primers, as described above, the method may comprise the step of conducting an amplification reaction to extend the hybridised first and second sequencing primers. This is shown in Figure 14C and 15C. Specifically, the amplification reaction extends the first and second sequencing primers, using the first and second immobilised strands as a template and a polymerase, to produce an extended polynucleotide strand. The resulting strand will comprise a complement of all or the majority of the first or second polynucleotide immobilised strand, including the insert. The necessary conditions to enable amplification have been discussed above. Unlike the template from which it was produced, the newly synthesised polynucleotide strand remains unattached from the solid surface, and is hybridised to the template strand.

In an embodiment, a number of rounds of amplification are carried out before cleavage. In another embodiment, one round of amplification is carried out before cleavage.

Similarly, where an immobilised strand comprises a first and second portion of a polynucleotide, for example, from fork loop preparation, one or more rounds of amplification may be carried out before cleavage.

In another aspect of the invention, there is provided a method of sequencing a first and second polynucleotide sequence, wherein the method comprises preparing first and second polynucleotide sequences for sequencing using the method described above for preparing a first and second polynucleotide sequence for sequencing; and sequencing

nucleobases in the first polynucleotide sequence and the second polynucleotide sequence.

5 As described herein, the template provides information (e.g. identification of the genetic sequence, identification of epigenetic modifications) on the original target polynucleotide sequence. For example, a sequencing process (e.g. a sequencing-by-synthesis or sequencing-by-ligation process) may reproduce information that was present in the original target polynucleotide sequence, by using complementary base pairing.

10 In one embodiment, the sequencing process comprises a first sequencing read and a second sequencing read. These sequencing reads may be carried out sequentially or concurrently. In one embodiment, the first sequencing read and the second sequencing read are conducted concurrently. In other words, the first sequencing read and the second sequencing read may be conducted at the same time. In one embodiment, the
15 first read sequencing the forward strand of a library strand, and the second read sequences the reverse strand of a library strand.

The first and second sequencing read may comprise cleaving the sequence bridge at the above-described cleavage sites. Cleavage of the above-described cleavage sites
20 may follow after a plurality of rounds of bridge amplification. In one embodiment, the cleavage sites are cleaved following at least 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 cycles, following at least 15, following at least 20 or following at least 25 cycles of bridge amplification.

Accordingly, in one embodiment, sequencing of the first polynucleotide sequence and
25 the second polynucleotide sequence occurs sequentially. That is, read one comprises sequencing the first polynucleotide sequence and read two comprises sequencing the second polynucleotide sequence or vice versa, and read one and two are temporally separated.

30 By way of example, we refer to Figure 13. In steps A to C, the first and second polynucleotide sequences are hybridised to immobilised primers on the solid support (A), and a cluster comprising a plurality of immobilised first and second polynucleotide strands is generated using bridge amplification, as described in detail elsewhere. Excess primers are then removed (B), for example using a single-stranded exonuclease, and
35 sequencing primers flowed across the solid support to hybridise at the 3' ends of the

immobilised first and second polynucleotide sequences (C); again as described in detail above.

In this embodiment, the first and second sequencing primers comprise different cleavage sites. Accordingly, in one embodiment, the first sequencing primers are cleaved by applying first cleavage conditions, such as a first nicking enzyme. Alternatively, the second sequencing primers are cleaved by applying second cleavage conditions, such as a second nicking enzyme. Either leads to the cleavage (i.e. a single-stranded cleavage or nicking) of either the first or second sequencing primer, as shown in Figure 13D. This allows sequencing to be performed starting from the nicked location. In other words, the nicked sites provide sequencing start sites for nick translation or nick sequencing. Accordingly, as shown in Figure 13E, read one comprises cleaving the first or second sequencing primer and sequencing the immobilised strand that is hybridised to the first or second sequencing primer.

Optionally, as shown in Figure 13F, the method may comprise blocking all or substantially all free 3' ends. For example, free 3' ends will be present on the read 1 sequenced strand. The blocking group may be any modification that prevents extension (i.e. elongation) of the free end by a polymerase. This serves to ensure that all polymerases are available to sequence the second sequencing strand. Examples of suitable blocking groups include a hairpin loop (e.g. a polynucleotide attached to the 3'-end, comprising in a 5' to 3' direction, a cleavable site such as a nucleotide comprising uracil, a loop portion, and a complement portion, wherein the complement portion is substantially complementary to all or a portion of the lawn primer), a hydrogen atom instead of a 3'-OH group, a phosphate group, a propyl spacer (e.g. -O-(CH₂)₃-OH instead of a 3'-OH group), a modification blocking the 3'-hydroxyl group (e.g. hydroxyl protecting groups, such as silyl ether groups (e.g. trimethylsilyl, triethylsilyl, triisopropylsilyl, t-butyl(dimethyl)silyl, t-butyl(diphenyl)silyl), ether groups (e.g. benzyl, allyl, t-butyl, methoxymethyl (MOM), 2-methoxyethoxymethyl (MEM), tetrahydropyranyl), or acyl groups (e.g. acetyl, benzoyl)), or an inverted nucleobase.

Next, as shown in Figure 13F, the sequencing primer not cleaved in read 1 is cleaved. In other words, if in read 1, the first sequencing primer was cleaved, in read 2 the second sequencing primer is cleaved and vice versa. Again, the second or first sequencing primers are cleaved by applying second or first cleavage conditions (such as a second or first nicking enzyme). Again, either leads to the cleavage (i.e. a single-stranded

cleavage or nicking) of either the second or first sequencing primer, as shown in Figure 13F. This allows the second sequencing read to be performed starting from the nicked location. Accordingly, as shown in Figure 13G, read one comprises cleaving the first or second sequencing primer and sequencing the immobilised strand that is hybridised to the second or first sequencing primer.

Thus, in one embodiment, there is provided a method of sequencing a first polynucleotide and a second polynucleotide sequentially. Advantageously, the method allows the sequencing of a first and second polynucleotide sequence without needing to re-synthesise the second polynucleotide sequence prior to sequencing (also referred to as cluster resynthesis), which is required for read 2 under conventional SBS workflows (Figure 5A). This in turn reduces the time required to sequence a target polynucleotide compared to conventional single strand sequencing.

In another embodiment, there is provided a method of sequencing a first polynucleotide and a second polynucleotide concurrently. That is read 1 and read 2 are carried out at the same time or substantially the same time.

By way of example, we refer to Figure 14. In steps A to C, the first and second polynucleotide sequences are hybridised to immobilised primers on the solid support (A), and a cluster comprising a plurality of immobilised first and second polynucleotide strands is generated using bridge amplification, as described in detail elsewhere. Excess primers are then removed, for example using a single-stranded exonuclease, and sequencing primers flowed across the solid support to hybridise at the 3' ends of the immobilised first and second polynucleotide sequences; again as described in detail above.

In this embodiment, the first and second sequencing primers comprise the same cleavage sites. In other words, the first and second sequencing primers are cleaved by applying the same cleavage conditions (i.e. cleaved by applying the same nicking enzyme). As such, when this cleavage condition is applied both the first and second sequencing primers are simultaneously cleaved, as shown in Figure 14D. In this embodiment, concurrent sequencing may be achieved using a mixture of cleavable and non-cleavable first sequencing primers and cleavable second sequencing primers. As such, when any one of the above described cleavage conditions is applied (for example, any of the above-described nicking enzymes is applied), a portion of the first sequencing

primers are cleaved and all or substantially all of the second sequencing primers are cleaved. Conversely, concurrent sequencing may be achieved using a mixture of cleavable and non-cleavable second sequencing primers and cleavable first sequencing primers. As such, when any one of the above described cleavage conditions is applied (for example, any of the above-described nicking enzymes is applied), a portion of the second sequencing primers are cleaved and all or substantially all of the first sequencing primers are cleaved. As shown in Figure 14E, either alternative allows sequencing of the first and second polynucleotide sequences to be performed starting from the nicked location.

Any ratio of cleavable:un-cleavable first or second sequencing primers can be used that generates or leads to a sequencing signal that is of a lower intensity than the second or first sequencing signal respectively. In one example, the ratio of cleavable:un-cleavable primers may be: 20:80 to 80:20, or 1:2 to 2:1 or 1:1. In one embodiment, a ratio of 50:50 of cleavable:un-cleavable first or second sequencing primers is used, which in turn generates a sequencing signal that is around 50% of the intensity of the sequencing signal from the second or first sequencing signal respectively.

Advantageously, this leads to a difference in intensity of signals that can be produced by sequencing the first polynucleotide sequences and the second polynucleotide sequences, which then allows both the first polynucleotide sequence and the second polynucleotide sequence to be identified (sequenced) simultaneously.

In turn, this enables at least a doubling of throughput of a sequencing reaction (i.e. increased sequencing efficiency) as well as a decrease in the time taken to sequence target polynucleotide strands. Furthermore, designing this feature into the sequencing primers simplifies the adaptors and the immobilised primers on the solid support, as described above. In addition, the above-described embodiment, would allow spatially separated clusters to be read in a temporally simultaneous manner through the generation of an optically unresolved signal that can be analytically separated using 16QaM, as described below.

In one embodiment, the first polynucleotide sequence is a forward strand of a double-stranded polynucleotide to be identified and the second polynucleotide sequence is a reverse strand of a double-stranded polynucleotide to be identified.

However, in another embodiment, the first polynucleotide sequence comprises a first portion to be identified and a second portion to be identified. Similarly, the second polynucleotide sequence comprises a first portion to be identified and a second portion to be identified. The first and second portions may be genetically unrelated, as described above (Figures 16 and 17). Such sequences can be produced by PCR-stitching, as described above. In another embodiment, the first portion on the first polynucleotide sequence may be a forward strand of a target polynucleotide and the second portion on the first polynucleotide sequence may be the reverse complement of the forward of a target polynucleotide i.e. a copy of the forward strand. Similarly, the first portion on the second polynucleotide sequence may be a reverse strand of a target polynucleotide and the second portion on the second polynucleotide sequence may be the forward complement of the forward of a target polynucleotide i.e. a copy of the reverse strand.

In embodiments where the first and second polynucleotide sequences comprise first and second portions, a further round of sequencing may be performed to sequence the first and second portions respectively.

In one embodiment, the first and second portion of a first polynucleotide and the first and second portion of the second polynucleotide are sequenced one-at-a-time. That is, each portion of each polynucleotide are sequenced sequentially. In this embodiment, the first portion of the first polynucleotide may be sequenced using a first cleavable sequencing primer, the second portion of the first polynucleotide may be sequenced using a second cleavable sequencing primer, the first portion of the second polynucleotide may be sequenced using a third cleavable sequencing primer and the second portion of the second polynucleotide may be sequenced using a fourth cleavable sequencing primer, wherein the first, second, third and fourth cleavable sequencing primers comprise first, second, third and fourth cleavable sites that are all different. As described above, the method may further require blocking any free 3' ends of the sequencing strand following completion of each read sequencing, as described above.

It is immaterial to the invention what order the portions are sequenced, as long as the four portions are sequenced sequentially. In one embodiment, the first and then the second portion of the first polynucleotide will be sequenced, followed by the first and then the second portion of the second polynucleotide. The result will be four sequencing reads. Again, as described above, the method allows the sequencing of four sequences without the step of paired-end resynthesis.

In a further embodiment, an index read (i.e. sequencing of one or more of the index sequences) may be performed. One or more index read may be performed before or after the SBS primer reads. So, in one embodiment, after read 1 and read 2 are completed according to the invention described herein, the extended SBS primer strands are denatured (for example using heat or NaOH) and washed away. Then Index primer 1 and Index primer 2 (if there are dual indexes) are sequentially hybridized to the flow cell and read using standard SBS. In another embodiment, the index reads are performed first. Index 1 is hybridized and read, then denatured and washed off, following which, Index 2 (if needed) is hybridized and read, then denatured and washed off. Then, the amplification with in-solution SBS primers and nick-translation SBS is performed.

In any of the above-described embodiments, sequencing is initiated at the cleavage site within the first and/or second sequencing primer that results from the cleavage event/application of the cleavage conditions. The polymerase enzyme binds to this site and incorporates complementary nucleotides base by base into the growing opposite strand. Accordingly, in one embodiment, the method comprises applying a polymerase. The polymerase may be a strand displacement polymerase (e.g. phi29). Alternatively, a polymerase and a 5' to 3' exonuclease or a polymerase with 5' to 3' exonuclease activity is used. The 5' to 3' exonuclease activity is used to essentially clear the path ahead of the growing strand (i.e. remove the downstream hybridised strand) for the polymerase to follow and incorporate a complementary nucleotide.

For the DNA polymerase to sequence the polynucleotide strand from a nick site, an exonuclease may be used. Exonucleases are enzymes that catalyse the removal of individual nucleotides from an end of a polynucleotide chain, by cleaving the phosphodiester bond via hydrolysis. Exonucleases are described as having a 5' to 3' activity or 3' to 5' activity, according to the direction of the nucleotide that they travel and digest.

In one embodiment, the DNA polymerase has native 5'-3' exonuclease activity. That is, the DNA polymerase naturally has 5' to 3' exonuclease activity. Examples of DNA polymerase with native 5' to 3' exonuclease activity include: Taq DNA Polymerase, T7 DNA polymerase and DNA Polymerase I.

In another embodiment, a DNA polymerase and an exonuclease are separately applied (i.e. flowed across the solid support) either sequentially or concurrently. In one embodiment, the exonuclease is applied prior to incorporation of the polymerase. Examples of suitable exonucleases include RecJf, Lambda Exonuclease, T7
5 exonuclease domain, T5 exonuclease, or the DNA polymerase I-like H3TH domain, Exonuclease V (RecBCD) or Exonuclease VII, with ss- and ds- DNA exonuclease activity.

In one embodiment, the DNA polymerase has been engineered to have 5' to 3' exonuclease activity. In one example, a DNA polymerase may be fused with a protein
10 possessing 5' to 3' exonuclease activity. In this way, the resulting fusion protein will contain an exonuclease domain, a DNA-binding domain and a polymerase domain. Suitable DNA polymerases include Pfu DNA polymerase, DNA Pol δ , Terminator™ DNA Polymerase, DNA Polymerase III. Commercially available fusion proteins include
15 the Phusion DNA polymerase. Alternatively, a fusion protein could be engineered using the DNA polymerases and exonucleases discussed above. The skilled person would understand that fusion proteins can be recombinant fusion proteins, created through genetic engineering of a fusion gene. Most commonly, fusion proteins are created by tandem fusion or linker-mediated fusion.

20 Compared to SBS using strand displacement polymerases, using a polymerase supplemented with 5'-3' exonuclease activity reduces or eliminates issues caused by having a displaced strand in the well. The exonuclease's gradual digestion of the strand ahead of the polymerase activity eliminates the displaced strand and, with it, the issues
25 caused by displaced strand reannealing, which in turn can affect the sequencing signal.

In one embodiment, sequencing may be carried out using any suitable "sequencing-by-synthesis" technique, wherein nucleotides are added successively in cycles to the free
3' hydroxyl group, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction.
30 The nature of the nucleotide added may be determined after each addition. One particular sequencing method relies on the use of modified nucleotides that can act as reversible chain terminators. Such reversible chain terminators comprise removable 3' blocking groups. Once such a modified nucleotide has been incorporated into the growing polynucleotide chain complementary to the region of the template being
35 sequenced there is no free 3'-OH group available to direct further sequence extension and therefore the polymerase cannot add further nucleotides. Once the nature of the

base incorporated into the growing chain has been determined, the 3' block may be removed to allow addition of the next successive nucleotide. By ordering the products derived using these modified nucleotides it is possible to deduce the DNA sequence of the DNA template. Such reactions can be done in a single experiment if each of the modified nucleotides has attached thereto a different label, known to correspond to the particular base, to facilitate discrimination between the bases added at each incorporation step. Suitable labels are described in PCT application PCT/GB2007/001770, the contents of which are incorporated herein by reference in their entirety. Alternatively, a separate reaction may be carried out containing each of the modified nucleotides added individually.

The modified nucleotides may carry a label to facilitate their detection. Such a label may be configured to emit a signal, such as an electromagnetic signal, or a (visible) light signal.

In a particular embodiment, the label is a fluorescent label (e.g. a dye). Thus, such a label may be configured to emit an electromagnetic signal, or a (visible) light signal. One method for detecting the fluorescently labelled nucleotides comprises using laser light of a wavelength specific for the labelled nucleotides, or the use of other suitable sources of illumination. The fluorescence from the label on an incorporated nucleotide may be detected by a CCD camera or other suitable detection means. Suitable detection means are described in PCT/US2007/007991, the contents of which are incorporated herein by reference in their entirety.

However, the detectable label need not be a fluorescent label. Any label can be used which allows the detection of the incorporation of the nucleotide into the DNA sequence.

Each cycle may involve simultaneous delivery of four different nucleotide types to the array of template molecules. Alternatively, different nucleotide types can be added sequentially and an image of the array of template molecules can be obtained between each addition step.

In some embodiments, each nucleotide type may have a (spectrally) distinct label. In other words, four channels may be used to detect four nucleobases (also known as 4-channel chemistry) (Figure 8 – top left). For example, a first nucleotide type (e.g. A) may include a first label (e.g. configured to emit a first wavelength, such as red light), a second

nucleotide type (e.g. G) may include a second label (e.g. configured to emit a second wavelength, such as blue light), a third nucleotide type (e.g. T) may include a third label (e.g. configured to emit a third wavelength, such as green light), and a fourth nucleotide type (e.g. C) may include a fourth label (e.g. configured to emit a fourth wavelength, such as yellow light). Four images can then be obtained, each using a detection channel that is selective for one of the four different labels. For example, the first nucleotide type (e.g. A) may be detected in a first channel (e.g. configured to detect the first wavelength, such as red light), the second nucleotide type (e.g. G) may be detected in a second channel (e.g. configured to detect the second wavelength, such as blue light), the third nucleotide type (e.g. T) may be detected in a third channel (e.g. configured to detect the third wavelength, such as green light), and the fourth nucleotide type (e.g. C) may be detected in a fourth channel (e.g. configured to detect the fourth wavelength, such as yellow light). Although specific pairings of bases to signal types (e.g. wavelengths) are described above, different signal types (e.g. wavelengths) and/or permutations may also be used.

In some embodiments, detection of each nucleotide type may be conducted using fewer than four different labels. For example, sequencing-by-synthesis may be performed using methods and systems described in US 2013/0079232, which is incorporated herein by reference.

Thus, in some embodiments, two channels may be used to detect four nucleobases (also known as 2-channel chemistry) (Figure 8 – top right). For example, a first nucleotide type (e.g. A) may include a first label (e.g. configured to emit a first wavelength, such as green light) and a second label (e.g. configured to emit a second wavelength, such as red light), a second nucleotide type (e.g. G) may not include the first label and may not include the second label, a third nucleotide type (e.g. T) may include the first label (e.g. configured to emit the first wavelength, such as green light) and may not include the second label, and a fourth nucleotide type (e.g. C) may not include the first label and may include the second label (e.g. configured to emit the second wavelength, such as red light). Two images can then be obtained, using detection channels for the first label and the second label. For example, the first nucleotide type (e.g. A) may be detected in both a first channel (e.g. configured to detect the first wavelength, such as red light) and a second channel (e.g. configured to detect the second wavelength, such as green light), the second nucleotide type (e.g. G) may not be detected in the first channel and may not be detected in the second channel, the third nucleotide type (e.g. T) may be detected in the first channel (e.g. configured to detect the first wavelength, such as red light) and may

not be detected in the second channel, and the fourth nucleotide type (e.g. C) may not be detected in the first channel and may be detected in the second channel (e.g. configured to detect the second wavelength, such as green light). Although specific pairings of bases to signal types (e.g. wavelengths) and/or combinations of channels are described above, different signal types (e.g. wavelengths) and/or permutations may also be used.

In some embodiments, one channel may be used to detect four nucleobases (also known as 1-channel chemistry) (Figure 8 – bottom). For example, a first nucleotide type (e.g. A) may include a cleavable label (e.g. configured to emit a wavelength, such as green light), a second nucleotide type (e.g. G) may not include a label, a third nucleotide type (e.g. T) may include a non-cleavable label (e.g. configured to emit the wavelength, such as green light), and a fourth nucleotide type (e.g. C) may include a label-accepting site which does not include the label. A first image can then be obtained, and a subsequent treatment carried out to cleave the label attached to the first nucleotide type, and to attach the label to the label-accepting site on the fourth nucleotide type. A second image may then be obtained. For example, the first nucleotide type (e.g. A) may be detected in a channel (e.g. configured to detect the wavelength, such as green light) in the first image and not detected in the channel in the second image, the second nucleotide type (e.g. G) may not be detected in the channel in the first image and may not be detected in the channel in the second image, the third nucleotide type (e.g. T) may be detected in the channel (e.g. configured to detect the wavelength, such as green light) in the first image and may be detected in the channel (e.g. configured to detect the wavelength, such as green light) in the second image, and the fourth nucleotide type (e.g. C) may not be detected in the channel in the first image and may be detected in the channel in the second image (e.g. configured to detect the wavelength, such as green light). Although specific pairings of bases to signal types (e.g. wavelengths) and/or combinations of images are described above, different signal types (e.g. wavelengths), images and/or permutations may also be used.

30

In another (or additional) embodiment, each adaptor sequence comprises two or at least two cleavage sites (or complements of the cleavage sites), as described above. In this embodiment, the 5' adaptor may comprise at least two cleavage sites, as described above, and the 3' adaptor comprises at least two complements of the same cleavage sites that are in the 5' adaptor, again as described above. Binding of a single-stranded library strand (which includes the insert and a 5' and 3' adaptor sequence) to a first (or

35

second) immobilised primer and extension of the first (or second) immobilised strand, results in a first (or second) immobilised strand comprising two or at least two 5' cleavage sites and at least two or two 3' complements of the same cleavage sites. When this first (or second) immobilised strand bridges, binds to a different (second or first) immobilised primer and the complement strand is resynthesized, as shown in Figure 19, the second (or first) immobilised resynthesized strand will also comprise at least two or two 5' cleavable sites and at least two or two 3' complements of (the same) cleavable sites. As shown in this Figure 19, both 5' ends of the first and second immobilised strands now comprise two or at least two cleavable sites.

In the next step, the cleavable sites at the 5' ends of both immobilised strands are simultaneously cleaved. As shown in Figure 19, this leads to cleavage of just one strand of the sequence bridge, at each 5' end, of the immobilised sequences – in other words, a nick at two sites (e.g. a double nick), at both 5' ends of the sequence bridges.

In another embodiment the each adaptor sequence comprises a single cleavage or complement of a cleavage site and each immobilised primer comprises at least one cleavable site, as described in detail above. Specifically, in one embodiment, the 5' adaptor comprises a single cleavage site and the 3' adaptor comprises a complement of the same cleavage site (that is in the 5' adaptor). Binding of a single-stranded library strand (which includes the insert and a 5' and 3' adaptor sequence) to a first (or second) immobilised primer and extension of the first (or second) immobilised strand, results in a first (or second) immobilised strand comprising a 5' cleavage site and a 3' complements of the same cleavage site. When this immobilised strand bridges, binds to a different (second or first) immobilised primer and the complement strand is resynthesized, as shown in Figure 19, the second (or first) immobilised resynthesized strand will also comprise a 5' cleavable site and a 3' complements of (the same) cleavable site. As shown in this Figure 19, both 5' ends of the first and second immobilised strands now comprise a cleavable site.

In the next step, the cleavable site in the immobilised primer and the cleavable site at the 5' end of both immobilised strands are simultaneously cleaved. As shown in Figure 19, this leads to cleavage of just one strand of the sequence bridge, at each 5' end, of the immobilised sequences – in other words, a nick at two sites (e.g. a double nick), at both 5' ends of the sequence bridges.

Where the cleavage sites are restriction sites, the cleavable sites are cleaved using a endonuclease, such as a single strand restriction endonuclease. Examples of suitable endonucleases are described above. In one embodiment, the cleavage sites can be cleaved simultaneously. For example, where the cleavage sites are restrictions sites (the same restriction site) for a nicking endonuclease, the cleavage sites can be cleaved by the addition of the (a single) nicking endonuclease. Alternatively if different cleavage sites are used – for example, if the 5' and 3' adaptor of one strand of the library strand (e.g. the forward strand) comprises a first cleavage site and complement of the first cleavage site respectively, and the 5' and 3' adaptor of another strand of the library strand (e.g. the reverse strand) comprises a second cleavage site and complement of the second cleavage site respectively, both the first and second cleavage sites can be simultaneously cleaved, for example by the simultaneous addition of a first and second endonuclease. Alternatively, the first and second cleavage sites can be sequentially cleaved by the sequential addition of e.g. a first and second endonuclease.

Accordingly, in one embodiment, the method further comprises applying (i.e. adding/ flowing over the surface of the solid support) a first endonuclease. Alternatively, the method comprises applying (i.e. adding/ flowing over the surface of the solid support) a first and second endonuclease. The first and second endonucleases may be applied sequentially or concurrently.

In one embodiment, the sequence between the two cleavage sites is washed away. The conditions required to wash away the sequence between the two cleavage sites may be determined empirically, and will depend on the sequence makeup of the double nicked insert. In essence, whatever temperature is sufficient to melt the bridge structure is too high. A lower temperature can be determined that is sufficient to denature the short, double-nicked insert. In some embodiments, it could be as low as 50°C, but as explained above, will depend on the sequence context of the insert. As described above, given the position of the cleavage sites the sequence that is washed away will either comprise the sequencing primer binding site or the sequencing primer binding site and an index sequence. However, in all embodiments described, the sequencing start sites are formed simultaneously by e.g. nicking enzymes, which therefore allows both strands of the duplex to be sequenced simultaneously.

In a further embodiment, the method comprises blocking all or substantially all free 3' ends of the immobilised strands. The blocking group may be any modification that

prevents extension (i.e. elongation) of the free end by a polymerase. Examples of suitable blocking groups include a hairpin loop (e.g. a polynucleotide attached to the 3'-end, comprising in a 5' to 3' direction, a cleavable site such as a nucleotide comprising uracil, a loop portion, and a complement portion, wherein the complement portion is substantially complementary to all or a portion of the lawn primer), a hydrogen atom instead of a 3'-OH group, a phosphate group, a propyl spacer (e.g. -O-(CH₂)₃-OH instead of a 3'-OH group), a modification blocking the 3'-hydroxyl group (e.g. hydroxyl protecting groups, such as silyl ether groups (e.g. trimethylsilyl, triethylsilyl, triisopropylsilyl, t-butyl(dimethyl)silyl, t-butyl(diphenyl)silyl), ether groups (e.g. benzyl, allyl, t-butyl, methoxymethyl (MOM), 2-methoxyethoxymethyl (MEM), tetrahydropyranyl), or acyl groups (e.g. acetyl, benzoyl)), or an inverted nucleobase.

In a further embodiment, the method comprises simultaneously sequencing the first and second immobilised strands. Accordingly, the method comprises carrying out a first sequencing read to determine the sequence of the first and second immobilised strands simultaneously, such as by a sequencing-by-synthesis technique or by a sequencing-by ligation technique.

In a further embodiment, the method comprises applying (i.e. adding/ flowing over the surface of the solid support) a plurality of first and second sequencing primers. The first sequencing primers are capable of hybridising to a first sequencing primer binding site on the second immobilised strand. The second sequencing primers are capable of hybridising to a second sequence primer binding site on the first immobilised strand. As described above, and shown in Figure 19(C), washing away the sequence between the two cleavage sites (the two nicked sites) provides a binding site(s) for sequencing primers. In particular, the nick sites provide sequencing start sites for nick translation or nick sequencing.

As used herein, by 'nick translation' is meant a process of extending a polynucleotide from a single-strand break in the polynucleotide (hereafter referred to as a 'nick site' or 'cleavable site') by incorporating labelled nucleotides at one side of the nick site while removing nucleotides from the other side of the nick site, resulting in the progression (or translation) of the nick site along the polynucleotide strand. Through the incorporation of labelled nucleotides, nick translation allows for the polynucleotide to be sequenced from the nick site by the methods described herein. A nick site, and therefore a preferred

sequencing start position, can be designed through the incorporation of a cleavable site into the polynucleotide, as described herein.

5 For a (DNA) polymerase to sequence the polynucleotide strand from a nick site, an exonuclease may be used. Exonucleases are enzymes that catalyse the removal of individual nucleotides from an end of a polynucleotide chain, by cleaving the phosphodiester bond via hydrolysis. Exonucleases are described as having a 5' to 3' activity or 3' to 5' activity, according to the direction of the nucleotide that they travel and digest.

10

Thus, the key stages to nick translation include: the introduction of a nick in a polynucleotide, to act as a start point for the introduction of labelled nucleotides; the activity of an exonuclease enzyme or exonuclease domain to remove nucleotides away from the nick site; and a polymerase that incorporates labelled nucleotides at the nick site. In this way, the nick site is progressively moved or translated along the polynucleotide.

15

Accordingly, the first sequencing read may comprise the binding of a first sequencing primer (also known as a read 1 sequencing primer) to the first sequencing primer binding site (e.g. first terminal sequencing primer binding site 303 in templates including a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion). The second sequencing read may comprise the binding of a second sequencing primer (also known as a read 2 sequencing primer) to the second sequencing primer binding site (e.g. second terminal sequencing primer binding site 304 in templates including a first polynucleotide sequence comprising a first portion and a second polynucleotide sequence comprising a second portion).

20

25

In one embodiment, the method comprises applying a first sequencing primer and a second sequencing primer, where the first sequencing primer comprises a mixed population of blocked and unblocked sequencing primers. Alternatively, the method comprises applying a first sequencing primer and a second sequencing primer, where the second sequencing primer comprises a mixed population of blocked and unblocked sequencing primers. The mixed population of blocked and unblocked first or second sequencing primers can be used to generate a sequence read where the intensity of the first sequencing read is different from the intensity of the second sequencing read, thereby allowing both sequencing reads to be analysed simultaneously. This means, for

30

35

example, that the forward and reverse strand of a library strand can be sequenced simultaneously, and without the need for paired-end resynthesis.

Any ratio of blocked:unblocked first or second sequencing primers can be used that generates or leads to a sequencing signal that is of a lower intensity than the second or first sequencing signal respectively. In one example, the ratio of blocked:unblocked primers may be: 20:80 to 80:20, or 1:2 to 2:1 or 1:1. In one embodiment, a ratio of 50:50 of blocked:unblocked first or second sequencing primers is used, which in turn generates a sequencing signal that is around 50% of the intensity of the sequencing signal from the second or first sequencing signal respectively.

By “blocked” is meant that the sequencing primer comprises a blocking group at a 3' end of the sequencing primer. Suitable blocking groups include a hairpin loop (e.g. a polynucleotide attached to the 3'-end, comprising in a 5' to 3' direction, a cleavable site such as a nucleotide comprising uracil, a loop portion, and a complement portion, wherein the complement portion is substantially complementary to all or a portion of the immobilised primer), a deoxynucleotide, a deoxyribonucleotide, a hydrogen atom instead of a 3'-OH group, a phosphate group, a phosphorothioate group, a propyl spacer (e.g. -O-(CH₂)₃-OH instead of a 3'-OH group), a modification blocking the 3'-hydroxyl group (e.g. hydroxyl protecting groups, such as silyl ether groups (e.g. trimethylsilyl, triethylsilyl, triisopropylsilyl, t-butyl(dimethyl)silyl, t-butyl(diphenyl)silyl), ether groups (e.g. benzyl, allyl, t-butyl, methoxymethyl (MOM), 2-methoxyethoxymethyl (MEM), tetrahydropyranyl), or acyl groups (e.g. acetyl, benzoyl)), or an inverted nucleobase. However, the blocking group may be any modification that prevents extension (i.e. elongation) of the primer by a polymerase.

The sequence of the sequencing primers and the sequence primer binding sites are not material to the methods of the invention, as long as the sequencing primers are able to bind to the sequence primer-binding site to enable amplification and sequencing of the regions to be identified. In one embodiment, the first sequencing primer comprises or consists of a sequence as defined in SEQ ID NO: 7 or a variant thereof. In another embodiment, the second sequencing primer comprises or consists of a sequence as defined in SEQ ID NO: 9 or a variant thereof.

The polymerase enzyme binds to this site and incorporates complementary nucleotides base by base into the growing opposite strand. Accordingly, in one embodiment, the

method comprises applying a polymerase. The polymerase may be a strand displacement polymerase (e.g. phi29). Alternatively, a polymerase and a 5'3' exonuclease or a polymerase with 5' to 3' exonuclease activity is used. The 5' to 3' exonuclease activity is used to essentially clear the path ahead of the growing strand
5 (i.e. remove the downstream hybridised strand)

For the DNA polymerase to sequence the polynucleotide strand from a nick site, an exonuclease may be used. Exonucleases are enzymes that catalyse the removal of individual nucleotides from an end of a polynucleotide chain, by cleaving the
10 phosphodiester bond via hydrolysis. Exonucleases are described as having a 5' to 3' activity or 3' to 5' activity, according to the direction of the nucleotide that they travel and digest.

In one embodiment, the DNA polymerase has native 5'-3' exonuclease activity. That is, the DNA polymerase naturally has 5' to 3' exonuclease activity. Examples of DNA polymerase with native 5' to 3' exonuclease activity include: Taq DNA Polymerase, T7 DNA polymerase and DNA Polymerase I.
15

In another embodiment, a DNA polymerase and an exonuclease are separately applied (i.e. flowed across the solid support) either sequentially or concurrently. In one embodiment, the exonuclease is applied prior to incorporation of the polymerase. Examples of suitable exonucleases include RecJf, Lambda Exonuclease, T7 exonuclease domain, T5 exonuclease, or the DNA polymerase I-like H3TH domain, Exonuclease V (RecBCD) or Exonuclease VII, with ss- and ds- DNA exonuclease
20 activity.
25

In one embodiment, the DNA polymerase has been engineered to have 5' to 3' exonuclease activity. In one example, a DNA polymerase may be fused with a protein possessing 5' to 3' exonuclease activity. In this way, the resulting fusion protein will contain an exonuclease domain, a DNA-binding domain and a polymerase domain. Suitable DNA polymerases include Pfu DNA polymerase, DNA Pol δ , Terminator™ DNA Polymerase, DNA Polymerase III. Commercially available fusion proteins include the Phusion DNA polymerase. Alternatively, a fusion protein could be engineered using the DNA polymerases and exonucleases discussed above. The skilled person would
30 understand that fusion proteins can be recombinant fusion proteins, created through
35

genetic engineering of a fusion gene. Most commonly, fusion proteins are created by tandem fusion or linker-mediated fusion.

Compared to SBS using strand displacement polymerases, using a polymerase supplemented with 5'-3' exonuclease activity reduces or eliminates issues caused by having a displaced strand in the well. The exonuclease's gradual digestion of the strand ahead of the polymerase activity eliminates the displaced strand and with it, the issues caused by displaced strand reannealing, which in turn can affect the sequencing signal. In one embodiment, sequencing may be carried out using any suitable "sequencing-by-synthesis" technique, wherein nucleotides are added successively in cycles to the free 3' hydroxyl group, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction. The nature of the nucleotide added may be determined after each addition. One particular sequencing method relies on the use of modified nucleotides that can act as reversible chain terminators. Such reversible chain terminators comprise removable 3' blocking groups. Once such a modified nucleotide has been incorporated into the growing polynucleotide chain complementary to the region of the template being sequenced there is no free 3'-OH group available to direct further sequence extension and therefore the polymerase cannot add further nucleotides. Once the nature of the base incorporated into the growing chain has been determined, the 3' block may be removed to allow addition of the next successive nucleotide. By ordering the products derived using these modified nucleotides it is possible to deduce the DNA sequence of the DNA template. Such reactions can be done in a single experiment if each of the modified nucleotides has attached thereto a different label, known to correspond to the particular base, to facilitate discrimination between the bases added at each incorporation step. Suitable labels are described in PCT application PCT/GB2007/001770, the contents of which are incorporated herein by reference in their entirety. Alternatively, a separate reaction may be carried out containing each of the modified nucleotides added individually.

The modified nucleotides may carry a label to facilitate their detection. Such a label may be configured to emit a signal, such as an electromagnetic signal, or a (visible) light signal.

In a particular embodiment, the label is a fluorescent label (e.g. a dye). Thus, such a label may be configured to emit an electromagnetic signal, or a (visible) light signal. One method for detecting the fluorescently labelled nucleotides comprises using laser light of

a wavelength specific for the labelled nucleotides, or the use of other suitable sources of illumination. The fluorescence from the label on an incorporated nucleotide may be detected by a CCD camera or other suitable detection means. Suitable detection means are described in PCT/US2007/007991, the contents of which are incorporated herein by reference in their entirety.

However, the detectable label need not be a fluorescent label. Any label can be used which allows the detection of the incorporation of the nucleotide into the DNA sequence.

Each cycle may involve simultaneous delivery of four different nucleotide types to the array of template molecules. Alternatively, different nucleotide types can be added sequentially and an image of the array of template molecules can be obtained between each addition step.

In some embodiments, each nucleotide type may have a (spectrally) distinct label. In other words, four channels may be used to detect four nucleobases (also known as 4-channel chemistry) (Figure 8 – left). For example, a first nucleotide type (e.g. A) may include a first label (e.g. configured to emit a first wavelength, such as red light), a second nucleotide type (e.g. G) may include a second label (e.g. configured to emit a second wavelength, such as blue light), a third nucleotide type (e.g. T) may include a third label (e.g. configured to emit a third wavelength, such as green light), and a fourth nucleotide type (e.g. C) may include a fourth label (e.g. configured to emit a fourth wavelength, such as yellow light). Four images can then be obtained, each using a detection channel that is selective for one of the four different labels. For example, the first nucleotide type (e.g. A) may be detected in a first channel (e.g. configured to detect the first wavelength, such as red light), the second nucleotide type (e.g. G) may be detected in a second channel (e.g. configured to detect the second wavelength, such as blue light), the third nucleotide type (e.g. T) may be detected in a third channel (e.g. configured to detect the third wavelength, such as green light), and the fourth nucleotide type (e.g. C) may be detected in a fourth channel (e.g. configured to detect the fourth wavelength, such as yellow light). Although specific pairings of bases to signal types (e.g. wavelengths) are described above, different signal types (e.g. wavelengths) and/or permutations may also be used.

In some embodiments, detection of each nucleotide type may be conducted using fewer than four different labels. For example, sequencing-by-synthesis may be performed

using methods and systems described in US 2013/0079232, which is incorporated herein by reference.

Thus, in some embodiments, two channels may be used to detect four nucleobases (also known as 2-channel chemistry) (Figure 8 – middle). For example, a first nucleotide type (e.g. A) may include a first label (e.g. configured to emit a first wavelength, such as green light) and a second label (e.g. configured to emit a second wavelength, such as red light), a second nucleotide type (e.g. G) may not include the first label and may not include the second label, a third nucleotide type (e.g. T) may include the first label (e.g. configured to emit the first wavelength, such as green light) and may not include the second label, and a fourth nucleotide type (e.g. C) may not include the first label and may include the second label (e.g. configured to emit the second wavelength, such as red light). Two images can then be obtained, using detection channels for the first label and the second label. For example, the first nucleotide type (e.g. A) may be detected in both a first channel (e.g. configured to detect the first wavelength, such as red light) and a second channel (e.g. configured to detect the second wavelength, such as green light), the second nucleotide type (e.g. G) may not be detected in the first channel and may not be detected in the second channel, the third nucleotide type (e.g. T) may be detected in the first channel (e.g. configured to detect the first wavelength, such as red light) and may not be detected in the second channel, and the fourth nucleotide type (e.g. C) may not be detected in the first channel and may be detected in the second channel (e.g. configured to detect the second wavelength, such as green light). Although specific pairings of bases to signal types (e.g. wavelengths) and/or combinations of channels are described above, different signal types (e.g. wavelengths) and/or permutations may also be used.

In some embodiments, one channel may be used to detect four nucleobases (also known as 1-channel chemistry) (Figure 8 – right). For example, a first nucleotide type (e.g. A) may include a cleavable label (e.g. configured to emit a wavelength, such as green light), a second nucleotide type (e.g. G) may not include a label, a third nucleotide type (e.g. T) may include a non-cleavable label (e.g. configured to emit the wavelength, such as green light), and a fourth nucleotide type (e.g. C) may include a label-accepting site which does not include the label. A first image can then be obtained, and a subsequent treatment carried out to cleave the label attached to the first nucleotide type, and to attach the label to the label-accepting site on the fourth nucleotide type. A second image may then be obtained. For example, the first nucleotide type (e.g. A) may be detected in a

channel (e.g. configured to detect the wavelength, such as green light) in the first image and not detected in the channel in the second image, the second nucleotide type (e.g. G) may not be detected in the channel in the first image and may not be detected in the channel in the second image, the third nucleotide type (e.g. T) may be detected in the channel (e.g. configured to detect the wavelength, such as green light) in the first image and may be detected in the channel (e.g. configured to detect the wavelength, such as green light) in the second image, and the fourth nucleotide type (e.g. C) may not be detected in the channel in the first image and may be detected in the channel in the second image (e.g. configured to detect the wavelength, such as green light). Although specific pairings of bases to signal types (e.g. wavelengths) and/or combinations of images are described above, different signal types (e.g. wavelengths), images and/or permutations may also be used.

In summary, the above-described example would allow spatially separated clusters to be read in a temporally simultaneous manner through the generation of an optically unresolved signal that can be analytically separated using 16QaM.

Data analysis using 16 QaM

Figure 9 is a scatter plot showing an example of sixteen distributions of signals generated by polynucleotide sequences disclosed herein.

The scatter plot of Figure 9 shows sixteen distributions (or bins) of intensity values from the combination of a brighter signal (i.e. a first signal as described herein) and a dimmer signal (i.e. a second signal as described herein); the two signals may be co-localized and may not be optically resolved as described above. The intensity values shown in Figure 9 may be up to a scale or normalisation factor; the units of the intensity values may be arbitrary or relative (i.e., representing the ratio of the actual intensity to a reference intensity). The sum of the brighter signal generated by the first portions and the dimmer signal generated by the second portions results in a combined signal. The combined signal may be captured by a first optical channel and a second optical channel. Since the brighter signal may be A, T, C or G, and the dimmer signal may be A, T, C or G, there are sixteen possibilities for the combined signal, corresponding to sixteen distinguishable patterns when optically captured. That is, each of the sixteen possibilities corresponds to a bin shown in Figure 9. The computer system can map the combined signal generated into one of the sixteen bins, and thus determine the added nucleobase at the first portion and the added nucleobase at the second portion, respectively.

For example, when the combined signal is mapped to bin 1612 for a base calling cycle, the computer processor base calls both the added nucleobase at the first portion and the added nucleobase at the second portion as C. When the combined signal is mapped to bin 1614 for the base calling cycle, the processor base calls the added nucleobase at the first portion as C and the added nucleobase at the second portion as T. When the combined signal is mapped to bin 1616 for the base calling cycle, the processor base calls the added nucleobase at the first portion as C and the added nucleobase at the second portion as G. When the combined signal is mapped to bin 1618 for the base calling cycle, the processor base calls the added nucleobase at the first portion as C and the added nucleobase at the second portion as A.

When the combined signal is mapped to bin 1622 for the base calling cycle, the processor base calls the added nucleobase at the first portion as T and the added nucleobase at the second portion as C. When the combined signal is mapped to bin 1624 for the base calling cycle, the processor base calls both the added nucleobase at the first portion and the added nucleobase at the second portion as T. When the combined signal is mapped to bin 1626 for the base calling cycle, the processor base calls the added nucleobase at the first portion as T and the added nucleobase at the second portion as G. When the combined signal is mapped to bin 1628 for the base calling cycle, the processor base calls the added nucleobase at the first portion as T and the added nucleobase at the second portion as A.

When the combined signal is mapped to bin 1632 for the base calling cycle, the processor base calls the added nucleobase at the first portion as G and the added nucleobase at the second portion as C. When the combined signal is mapped to bin 1634 for the base calling cycle, the processor base calls the added nucleobase at the first portion as G and the added nucleobase at the second portion as T. When the combined signal is mapped to bin 1636 for the base calling cycle, the processor base calls both the added nucleobase at the first portion and the added nucleobase at the second portion as G. When the combined signal is mapped to bin 1638 for the base calling cycle, the processor base calls the added nucleobase at the first portion as G and the added nucleobase at the second portion as A.

When the combined signal is mapped to bin 1642 for the base calling cycle, the processor base calls the added nucleobase at the first portion as A and the added

nucleobase at the second portion as C. When the combined signal is mapped to bin 1644 for the base calling cycle, the processor base calls the added nucleobase at the first portion as A and the added nucleobase at the second portion as T. When the combined signal is mapped to bin 1646 for the base calling cycle, the processor base calls the added nucleobase at the first portion as A and the added nucleobase at the second portion as G. When the combined signal is mapped to bin 1648 for the base calling cycle, the processor base calls both the added nucleobase at the first portion and the added nucleobase at the second portion as A.

10 In this particular example, T is configured to emit a signal in both the IMAGE 1 channel and the IMAGE 2 channel, A is configured to emit a signal in the IMAGE 1 channel only, C is configured to emit a signal in the IMAGE 2 channel only, and G does not emit a signal in either channel. However, different permutations of nucleobases can be used to achieve the same effect by performing dye swaps. For example, A may be configured to
15 emit a signal in both the IMAGE 1 channel and the IMAGE 2 channel, T may be configured to emit a signal in the IMAGE 1 channel only, C may be configured to emit a signal in the IMAGE 2 channel only, and G may be configured to not emit a signal in either channel.

20 Further details regarding performing base-calling based on a scatter plot having sixteen bins may be found in U.S. Patent Application Publication No. 2019/0212294, the disclosure of which is incorporated herein by reference.

Figure 10 is a flow diagram showing a method 1700 of base calling according to the
25 present disclosure. The described method allows for simultaneous sequencing of two (or more) portions (e.g. the first portion and the second portion) in a single sequencing run from a single combined signal obtained from the first portion and the second portion, thus requiring less sequencing reagent consumption and faster generation of data from both the first portion and the second portion. Further, the simplified method may reduce
30 the number of workflow steps while producing the same yield as compared to existing next-generation sequencing methods. Thus, the simplified method may result in reduced sequencing runtime.

As shown in Figure 10, the disclosed method 1700 may start from block 1701. The
35 method may then move to block 1710.

At block 1710, intensity data is obtained. The intensity data includes first intensity data and second intensity data. The first intensity data comprises a combined intensity of a first signal component obtained based upon a respective first nucleobase of the first portion and a second signal component obtained based upon a respective second nucleobase of the second portion. Similarly, the second intensity data comprises a combined intensity of a third signal component obtained based upon the respective first nucleobase of the first portion and a fourth signal component obtained based upon the respective second nucleobase of the second portion.

As such, the first portion is capable of generating a first signal comprising a first signal component and a third signal component. The second portion is capable of generating a second signal comprising a second signal component and a fourth signal component.

As described above, the first portion and the second portion may be arranged on the solid support such that signals from the first portion and the second portion are detected by a single sensing portion and/or may comprise a single cluster such that first signals and second signals from each of the respective first portions and second portions cannot be spatially resolved.

In one example, obtaining the intensity data comprises selecting intensity data that corresponds to two (or more) different portions (e.g. the first portion and the second portion). In one example, intensity data is selected based upon a chastity score. A chastity score may be calculated as the ratio of the brightest base intensity divided by the sum of the brightest and second brightest base intensities. The desired chastity score may be different depending upon the expected intensity ratio of the light emissions associated with the different portions. As described above, it may be desired to produce clusters comprising the first portion and the second portion, which give rise to signals in a ratio of 2:1. In one example, high-quality data corresponding to two portions with an intensity ratio of 2:1 may have a chastity score of around 0.8 to 0.9.

After the intensity data has been obtained, the method may proceed to block 1720. In this step, one of a plurality of classifications is selected based on the intensity data. Each classification represents a possible combination of respective first and second nucleobases. In one example, the plurality of classifications comprises sixteen classifications as shown in Figure 9, each representing a unique combination of first and second nucleobases. Where there are two portions, there are sixteen possible

combinations of first and second nucleobases. Selecting the classification based on the first and second intensity data comprises selecting the classification based on the combined intensity of the first and second signal components and the combined intensity of the third and fourth signal components.

5

The method may then proceed to block 1730, where the respective first and second nucleobases are base called based on the classification selected in block 1720. The signals generated during a cycle of a sequencing are indicative of the identity of the nucleobase(s) added during sequencing (e.g. using sequencing-by-synthesis). It will be appreciated that there is a direct correspondence between the identity of the nucleobases that are incorporated and the identity of the complementary base at the corresponding position of the template sequence bound to the solid support. Therefore, any references herein to the base calling of respective nucleobases at the two portions encompasses the base calling of nucleobases hybridised to the template sequences and, alternatively or additionally, the identification of the corresponding nucleobases of the template sequences. The method may then end at block 1740.

10

15

Kits

Methods as described herein may be performed by a user physically. In other words, a user may themselves conduct the methods of preparing polynucleotide sequences for identification as described herein, and as such the methods as described herein may not need to be computer-implemented.

20

In one aspect of the invention there is provided a sequencing kit comprising a plurality of first and second sequencing primers, wherein the first sequencing primers comprise a first cleavage site and wherein the second sequencing primers comprise a second cleavage site, where the first and second cleavage sites are different.

25

In one aspect of the invention there is provided a sequencing kit comprising a plurality of first and second sequencing primers, wherein the first sequencing primers comprise a first cleavage site and wherein the second sequencing primers comprise a second cleavage site, where the first and second cleavage sites are the same. In one embodiment, the kit comprises a mixture of cleavable and non-cleavable second sequencing primers and cleavable first sequencing primers. Alternatively, the kit comprises a mixture of cleavable and non-cleavable first sequencing primers and

30

35

cleavable second sequencing primers. In one embodiment, the ratio of cleavable: non/un-cleavable first or second sequencing primers is 1:1.

5 The kit may also further comprise a polymerase and a 5' to 3' exonuclease or a polymerase with 5' to 3' exonuclease activity.

10 The kit may also comprise cleavage effectors. By "cleavage effectors" is meant effectors that mediate cleavage of the first and/or second sequencing primers. As described above, examples of suitable cleavage effectors include AP lyases, glycosylases with or without an AP lyase or endonuclease, and endonucleases, such as endonuclease V and the USER mix (which is a cocktail of uracil glycosylase and endonuclease VIII).

The kit may also comprise a kinase, such as T4 kinase.

15 In another aspect of the invention, there is provided a library preparation kit, wherein the kit comprises a plurality of first 5' adaptors, a plurality of second 3' adaptors, as described above. In one embodiment, the kits further comprise instructions for use. In a further embodiment, the kit may further comprise at least one single-stranded endonuclease or restriction endonuclease. In one aspect, the endonuclease is selected from Nt. BspQI, 20 Cas9 D10A and Cas9 H840A. In another embodiment the kit may additionally comprise a uracil glycosylase or USER enzyme mix (which is a cocktail of uracil glycosylase and endonuclease VIII).

25 In another aspect of the invention there is provided a solid support comprising a plurality of a third and/or fourth primer immobilised thereon, as described above.

30 In another aspect of the invention, there is provided a kit comprising instructions for preparing polynucleotide sequences for identification according to the methods described herein and/or sequencing polynucleotide sequences according to the methods described herein.

Computer programs and products

35 In other embodiments, methods as described herein may be performed by a computer. In other words, a computer may contain instructions to conduct the methods of preparing

polynucleotide sequences for identification as described herein, and as such the methods as described herein may be computer-implemented.

5 Accordingly, in another aspect of the invention, there is provided a data processing device comprising means for carrying out the methods as described herein.

The data processing device may be a polynucleotide sequencer.

10 The data processing device may comprise reagents used for methods as described herein.

The data processing device may comprise a solid support as described herein, such as a flow cell.

15 In another aspect of the invention, there is provided a computer program product comprising instructions which, when the program is executed by a processor, cause the processor to carry out the methods as described herein.

20 In another aspect of the invention, there is provided a computer-readable storage medium comprising instructions which, when executed by a processor, cause the processor to carry out the methods as described herein.

In another aspect of the invention, there is provided a computer-readable data carrier having stored thereon the computer program product as described herein.

25 In another aspect of the invention, there is provided a data carrier signal carrying the computer program product as described herein.

30 The various illustrative imaging or data processing techniques described in connection with the embodiments disclosed herein can be implemented as electronic hardware, computer software, or combinations of both. To illustrate this interchangeability of hardware and software, various illustrative components, blocks, modules, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. The described
35 functionality can be implemented in varying ways for each particular application, but such

implementation decisions should not be interpreted as causing a departure from the scope of the disclosure.

5 The various illustrative detection systems described in connection with the embodiments disclosed herein can be implemented or performed by a machine, such as a processor configured with specific instructions, a digital signal processor (DSP), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) or other programmable logic device, discrete gate or transistor logic, discrete hardware components, or any combination thereof designed to perform the functions described
10 herein. A processor can be a microprocessor, but in the alternative, the processor can be a controller, microcontroller, or state machine, combinations of the same, or the like. A processor can also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration. For
15 example, systems described herein may be implemented using a discrete memory chip, a portion of memory in a microprocessor, flash, EPROM, or other types of memory.

The elements of a method, process, or algorithm described in connection with the embodiments disclosed herein can be embodied directly in hardware, in a software
20 module executed by a processor, or in a combination of the two. A software module can reside in RAM memory, flash memory, ROM memory, EPROM memory, EEPROM memory, registers, hard disk, a removable disk, a CD-ROM, or any other form of computer-readable storage medium known in the art. An exemplary storage medium can be coupled to the processor such that the processor can read information from, and write
25 information to, the storage medium. In the alternative, the storage medium can be integral to the processor. The processor and the storage medium can reside in an ASIC. A software module can comprise computer-executable instructions which cause a hardware processor to execute the computer-executable instructions.

30 Computer-executable instructions may be stored in a (transitory or non-transitory) computer readable storage medium (e.g., memory, storage system, etc.) storing code, or computer readable instructions.

The invention is also defined in the following clauses:

1. A method of preparing a first polynucleotide sequence and a second polynucleotide sequence for sequencing, wherein the method comprises
 - a. hybridising the first polynucleotide sequence to first immobilised primers on a solid support and hybridising the second polynucleotide sequence to second immobilised primers on a solid support;
 - b. synthesising a plurality of first and second polynucleotide sequences by conducting an amplification reaction to extend the first and second immobilised primers;
 - c. selectively removing first and second immobilised primers that have not been extended;
 - d. applying a plurality of first and second sequencing primers, wherein the first sequencing primers comprise a first cleavable site and wherein the second sequencing primers comprise a second cleavable site and conducting an amplification reaction to extend the first and second sequencing primers; and
 - e. selectively cleaving the first cleavable site in the first sequencing primer and/or selectively cleaving the second cleavable site in the second sequencing primer.
2. A method according to clause 1, wherein the first and second immobilised primers that have not been extended are removed using a single-stranded exonuclease.
3. A method according to any of clauses 1 to 2, wherein the first and second cleavable sites is an unnatural nucleotide, wherein preferably the unnatural nucleotide is selected from oxoguanine, even more preferably 8-oxoguanine, a hypoxanthine, even more preferably inosine, xanthine, a methylguanine, a methyladenines, a modified cytosine, dihydrouracil, and uracil.
4. A method according to any of clauses 1 to 3, wherein the first and second cleavable sites are the same.
5. A method according to any of clauses 1 to 3, wherein the first and second cleavable sites are different.
6. A method according to any one of clauses 1 to 5 wherein the first or second sequencing primers comprise a mixture of cleavable sequencing primers and non-cleavable sequencing primers.

7. A method according to any one of clauses 1 to 6 wherein the method further comprises a step of linearising the first polynucleotide sequence and the second polynucleotide sequence following step (b).
- 5 8. A method according to any one of clauses 1 to 7, wherein the first polynucleotide sequence is a forward strand of a double-stranded polynucleotide to be identified and the second polynucleotide sequence is a reverse strand of a double-stranded polynucleotide to be identified.
9. A method of sequencing a first and second polynucleotide sequence, wherein the method comprises:
 - 10 preparing first and second polynucleotide sequences for sequencing using a method according to any one of clauses 1 to 8; and
 - sequencing nucleobases in the first polynucleotide sequence and the second polynucleotide sequence.
- 15 10. A method according to clause 9, wherein the method comprises cleaving the first cleavable site in the first sequencing primer and sequencing the nucleobases in the first polynucleotide sequence.
11. A method according to clause 9 or 10, wherein the method further comprises subsequently cleaving the second cleavable site in the second sequencing primer and sequencing the nucleobases in the first polynucleotide sequence.
- 20 12. A method according to any of clauses 9 to 11, wherein the method further comprises blocking the 3' end of the first polynucleotide sequence prior to sequencing the second polynucleotide.
- 25 13. A method according to clause 9, wherein the method comprises concurrently cleaving the first cleavable site in the first sequencing primer and cleaving the second cleavable site in the second sequencing primer and sequencing nucleobases in the first polynucleotide sequence and the second polynucleotide sequence, wherein preferably the method comprises applying a plurality of sequencing primers, wherein the first or second sequencing primers comprise a mixture of cleavable sequencing primers and non-cleavable sequencing primers.
- 30 14. A method according to clause 13, wherein the first or second sequencing primers comprise a 1:1 mixture of cleavable sequencing primers and non-cleavable sequencing primers.

15. A method according to any one of clauses 13 to 14, wherein the step of concurrently sequencing nucleobases comprises performing sequencing-by-synthesis or sequencing-by-ligation.
- 5 16. A method according to any one of clauses 13 to 15, wherein the step of concurrently sequencing nucleobases comprises treatment with a polymerase and a 5'-3' exonuclease or a polymerase with 5' to 3' exonuclease activity.
17. A method according to any one of clauses 13 to 16, wherein the step of concurrently sequencing nucleobases comprises treatment with a strand displacement polymerase (e.g. phi29).
- 10 18. A sequencing kit comprising a plurality of first and second sequencing primers, wherein the first and second sequencing primers comprise at least one cleavage site.
19. The sequencing kit of clause 18, wherein the kit further comprises first and second sequencing primers and wherein the first and second sequencing primers
15 comprise a mixture of cleavable sequencing primers and un-cleavable sequencing primers.
20. The sequencing kit of clause 18 or 19, wherein the kit further comprises a polymerase and a 5' to 3' exonuclease or a polymerase with 5' to 3' exonuclease activity.
- 20 21. The sequencing kit of clause 18 to 20, wherein the kit further comprises a single-stranded exonuclease and cleavable effectors, such as nickases.
22. A data processing device comprising means for carrying out a method according to any one of clauses 1 to 17.
23. A data processing device according to clause 22, wherein the data processing
25 device is a polynucleotide sequencer.
24. A computer program product comprising instructions which, when the program is executed by a processor, cause the processor to carry out a method according to any one of clauses 1 to 17.
25. A computer-readable storage medium comprising instructions which, when
30 executed by a processor, cause the processor to carry out a method according to any one of clauses 1 to 17.
26. A computer-readable data carrier having stored thereon a computer program product according to clauses 24.
27. A data carrier signal carrying a computer program product according to clauses
35 24.

The invention may also be defined by the following clauses:

1. A method of preparing a first polynucleotide sequence and a second polynucleotide sequence for simultaneous sequencing, wherein the method comprises
 - a. attaching a first adaptor to a 5' end of the first polynucleotide sequence and attaching a first adaptor to a 5' end of the second polynucleotide sequence;
 - b. attaching a second adaptor to a 3' end of the first polynucleotide sequence and attaching a second adaptor to a 3' end of the second polynucleotide sequence; wherein the first adaptor comprises at least one cleavable site and the second adaptor comprises a complement of at least one cleavable site, and wherein the first and second adaptors comprises an immobilised primer-binding sequence and a sequencing primer binding sequence or complements thereof;
 - c. hybridising the first and second polynucleotide sequences to first and second immobilised primers on a solid support, wherein optionally, the first and second immobilised primers comprise at least one cleavable site;
 - d. forming a cluster of amplified first and second polynucleotide sequences; and
 - e. cleaving at least one cleavable sites to provide sequencing sites for nick translation.
2. A method according to clause 1, wherein the first adaptor comprises two or at least two cleavable sites and the second adaptor comprises two or at least two cleavable sites.
3. A method according to clause 1, wherein the first adaptor comprises one cleavable site and the second adaptor comprises one complement of a cleavable site and wherein the first and second immobilised primers comprise one cleavable site.
4. A method according to any one of clauses 1 to 3, further comprising a step of removing a sequence between the cleavable sites.
5. A method according to any one of clauses 1 to 4, further comprising a step of blocking 3'-ends of the first polynucleotide sequences and the second polynucleotide sequences.

6. A method according to any one of clauses 1 to 5, further comprising the step of applying a plurality of first and second sequencing primers, wherein the first or second sequencing primers comprise a mixture of blocked sequencing primers and unblocked sequencing primers.
- 5 7. A method according to clause 6, wherein the blocked sequencing primer comprises a blocking group at a 3' end of the sequencing primer.
8. A method according to clause 7, wherein the blocking group is selected from the group consisting of: a hairpin loop, a deoxynucleotide, a deoxyribonucleotide, a hydrogen atom instead of a 3'-OH group, a phosphate group, a phosphorothioate group, a propyl spacer, a modification blocking the 3'-hydroxyl group, or an
10 inverted nucleobase.
9. A method according to any one of clauses 6 to 8, wherein the first or second sequencing primers comprise a mixture of blocked sequencing primers and unblocked sequencing primers, and wherein the ratio of blocked to unblocked
15 sequencing primers is 1:1.
10. A method according to any one of clauses 1 to 9, wherein the cleavable sites are restriction sites for a nicking endonuclease.
11. A method according to any one of clauses 1 to 10, wherein the cluster is formed by bridge amplification.
- 20 12. A method according to any one of clauses 1 to 11, wherein the step of cleaving at the cleavage sites is carried out following bridge amplification and before linearization of the extended sequences.
13. A method according to any one of clauses 1 to 12, wherein the first polynucleotide sequence is a forward strand of a double-stranded polynucleotide to be identified
25 and the second polynucleotide sequence is a reverse strand of a double-stranded polynucleotide to be identified.
14. A method of simultaneously sequencing a first and second polynucleotide sequence, comprising:
preparing polynucleotide sequences for simultaneous sequencing using
30 a method according to any one of clauses 1 to 13; and
concurrently sequencing nucleobases in the first portion and the second portion.
15. A method according to clause 14, wherein the step of concurrently sequencing nucleobases comprises performing sequencing-by-synthesis or sequencing-by-ligation.
35

16. A method according to any one of clauses 14 to 15, wherein the step of concurrently sequencing nucleobases comprises treatment with a polymerase and a 5'-3' exonuclease or a polymerase with 5' to 3' exonuclease activity.
- 5 17. A library preparation kit comprising of a plurality of first adaptors and a plurality of second adaptors, wherein the first adaptors comprise at least one cleavable site and the second adaptors comprise a complement of at least one cleavable site, and wherein the first and second adaptors comprises an immobilised primer-binding sequence and a sequencing primer binding sequence or complements thereof.
- 10 18. The library preparation kit of clause 17, wherein the kit further comprises first and second sequencing primers and wherein the first and second sequencing primers comprise a mixture of blocked sequencing primers and unblocked sequencing primers, and wherein the ratio of blocked to unblocked sequencing primers is 1:1.
- 15 19. The library preparation kit of clause 17 to 18, wherein the kit further comprises a polymerase and a 5' to 3' exonuclease or a polymerase with 5' to 3' exonuclease activity.
- 20 20. A data processing device comprising means for carrying out a method according to any one of clauses 1 to 16.
21. A data processing device according to clause 20, wherein the data processing device is a polynucleotide sequencer.
- 20 22. A computer program product comprising instructions, which when the program is executed by a processor cause the processor to carry out a method according to any one of clauses 1 to 16.
- 25 23. A computer-readable storage medium comprising instructions which, when executed by a processor, cause the processor to carry out a method according to any one of clauses 1 to 16.
24. A computer-readable data carrier having stored thereon a computer program product according to clause 22.
- 30 25. A data carrier signal carrying a computer program product according to clause 22.

Additional Notes

The embodiments described herein are exemplary. Modifications, rearrangements, substitute processes, etc. may be made to these embodiments and still be encompassed within the teachings set forth herein. One or more of the steps, processes, or methods

described herein may be carried out by one or more processing and/or digital devices, suitably programmed.

5 Conditional language used herein, such as, among others, “can,” “might,” “may,” “e.g.,” and the like, unless specifically stated otherwise, or otherwise understood within the context as used, is generally intended to convey that certain embodiments include, while other embodiments do not include, certain features, elements and/or states. Thus, such conditional language is not generally intended to imply that features, elements and/or states are in any way required for one or more embodiments or that one or more
10 embodiments necessarily include logic for deciding, with or without author input or prompting, whether these features, elements and/or states are included or are to be performed in any particular embodiment. The terms “comprising,” “including,” “having,” “involving,” and the like are synonymous and are used inclusively, in an open-ended fashion, and do not exclude additional elements, features, acts, operations, and so forth. Also, the term “or” is used in its inclusive sense (and not in its exclusive sense) so that
15 when used, for example, to connect a list of elements, the term “or” means one, some, or all of the elements in the list. The term “comprising” may be considered to encompass “consisting”.

20 Disjunctive language such as the phrase “at least one of X, Y or Z,” unless specifically stated otherwise, is otherwise understood with the context as used in general to present that an item, term, etc., may be either X, Y or Z, or any combination thereof (e.g., X, Y and/or Z). Thus, such disjunctive language is not generally intended to, and should not, imply that certain embodiments require at least one of X, at least one of Y or at least one
25 of Z to each be present.

The terms “about” or “approximate” and the like are synonymous and are used to indicate that the value modified by the term has an understood range associated with it, where the range can be $\pm 20\%$, $\pm 15\%$, $\pm 10\%$, $\pm 5\%$, or $\pm 1\%$. The term “substantially” is used to
30 indicate that a result (e.g., measurement value) is close to a targeted value, where close can mean, for example, the result is within 80% of the value, within 90% of the value, within 95% of the value, or within 99% of the value. The term “partially” is used to indicate that an effect is only in part or to a limited extent.

35 Unless otherwise explicitly stated, articles such as “a” or “an” should generally be interpreted to include one or more described items. Accordingly, phrases such as “a

device configured to” or “a device to” are intended to include one or more recited devices. Such one or more recited devices can also be collectively configured to carry out the stated recitations. For example, “a processor to carry out recitations A, B and C” can include a first processor configured to carry out recitation A working in conjunction with
5 a second processor configured to carry out recitations B and C.

While the above detailed description has shown, described, and pointed out novel features as applied to illustrative embodiments, it will be understood that various omissions, substitutions, and changes in the form and details of the devices or algorithms
10 illustrated can be made without departing from the spirit of the disclosure. As will be recognized, certain embodiments described herein can be embodied within a form that does not provide all of the features and benefits set forth herein, as some features can be used or practiced separately from others. All changes which come within the meaning and range of equivalency of the claims are to be embraced within their scope.

15 It should be appreciated that all combinations of the foregoing concepts (provided such concepts are not mutually inconsistent) are contemplated as being part of the inventive subject matter disclosed herein. In particular, all combinations of claimed subject matter appearing at the end of this disclosure are contemplated as being part of the inventive
20 subject matter disclosed herein.

The present invention will now be described by way of the following non-limiting examples.

Example 1: Evaluation of dsSBS to accurately read sequences that include high G
25 content

Sequencing performance at a known G-quadruplex region was compared with ssSBS and dsSBS. Assay was performed on an iSeq with ssSBS performed according to standard manufacturer protocol and dsDNA SBS performed as shown in Figure 13. SBS (both dsDNA and ssDNA) was performed using various methods such as described in
30 US Patent No. 11,293,061B2 and US Patent Pub. No. US 2021/0403500A1, with incorporation time extended to 2 minutes. For dsSBS, Nicking with GAN_TaqL-Pol(A) fusion was performed as a separate step after incorporation step, as described in US 18/089,867 and published as US 2023-0212667, the content of which is incorporated by reference in its entirety. Nicking time was 30s per SBS cycle. Sequencing performance
35 at a region containing a known G-quadruplex is shown in Figures 11 and 12.

Figure 11 is a plot prepared using Integrative Genomics Viewer from the Broad Institute, and shows significant improvement in coverage through a region containing a known G-quadruplex, indicated by the lowercase base letters at the bottom of the plot.

Figure 12 shows that dsSBS using the present invention reduces sequencing errors at G-quad regions. Read 1-5 represents dsSBS and read 6, standard ssSBS. Read 6 (standard SBS) had a greater error rate in the region indicated between the vertical lines, which contains a G-quadruplex. dsSBS (Reads 1-5) had much lower error rate, demonstrating the ability of double stranded sequencing to remove the effect of G-quadruplexes on sequencing performance.

10 **SEQUENCE LISTING**

SEQ ID NO: 1: P5 sequence

AATGATACGGCGACCAACGAGATCTACAC

SEQ ID NO: 2: P7 sequence

CAAGCAGAAGACGGCATACGAGAT

15 **SEQ ID NO: 3:** P5' sequence (complementary to P5)

GTGTAGATCTCGGTGGTCCCGTATCATT

SEQ ID NO: 4: P7' sequence (complementary to P7)

ATCTCGTATGCCGTCTTCTGCTTG

SEQ ID NO: 5: Alternative P5 sequence

20 AATGATACGGCGACCGA

SEQ ID NO: 6: Alternative P5' sequence (complementary to alternative P5 sequence)

TCGGTCGCCGTATCATT

SEQ ID NO. 7: SBS3

ACACTCTTCCCTACACGACGCTCTTCCGATCT

25

SEQ ID NO. 8: SBS3'

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

SEQ ID NO. 9: SBS12

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

30

SEQ ID NO. 10: SBS12'

AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC

CLAIMS:

1. A method of sequencing a first and second polynucleotide sequence, wherein the method comprises preparing the first and second polynucleotide sequence for sequencing and sequencing the nucleobases in the first and second polynucleotide sequence, wherein the first and second polynucleotide sequences are sequenced using nick translation; wherein
- preparing a first polynucleotide sequence and a second polynucleotide sequence for sequencing comprises
 - a. hybridising the first polynucleotide sequence to first immobilised primers on a solid support and hybridising the second polynucleotide sequence to second immobilised primers on a solid support;
 - b. synthesising a plurality of first and second polynucleotide sequences by conducting an amplification reaction to extend the first and second immobilised primers;
 - c. selectively removing first and second immobilised primers that have not been extended;
 - d. applying a plurality of first and second sequencing primers, wherein the first sequencing primers comprise a first cleavable site and wherein the second sequencing primers comprise a second cleavable site and conducting an amplification reaction to extend the first and second sequencing primers; and
 - e. selectively cleaving the first cleavable site in the first sequencing primer and/or selectively cleaving the second cleavable site in the second sequencing primer; or wherein
 - preparing a first polynucleotide sequence and a second polynucleotide sequence for sequencing comprises:
 - f. attaching a first adaptor to a 5' end of the first polynucleotide sequence and attaching a first adaptor to a 5' end of the second polynucleotide sequence;
 - g. attaching a second adaptor to a 3' end of the first polynucleotide sequence and attaching a second adaptor to a 3' end of the second polynucleotide sequence; wherein the first adaptor comprises at least one cleavable site and the second adaptor comprises a complement of at least one cleavable site, and wherein the first and second adaptors comprises

an immobilised primer-binding sequence and a sequencing primer binding sequence or complements thereof;

- h. hybridising the first and second polynucleotide sequences to first and second immobilised primers on a solid support, wherein optionally, the first and second immobilised primers comprise at least one cleavable site;
- i. forming a cluster of amplified first and second polynucleotide sequences; and
- j. cleaving at least one cleavable sites to provide sequencing sites for nick translation.

2. A method according to claim 1, wherein the first and second immobilised primers that have not been extended are removed using a single-stranded exonuclease following step (e).

3. A method according to any of claims 1 to 2, wherein the first and second cleavable site is an unnatural nucleotide, wherein preferably the unnatural nucleotide is selected from oxoguanine, even more preferably 8-oxoguanine, a hypoxanthine, even more preferably inosine, xanthine, a methylguanine, a methyladenines, a modified cytosine, dihydrouracil, and uracil.

4. A method according to any of claims 1 to 3, wherein the first and second cleavable sites are the same or different, and wherein preferably, the first or second sequencing primers comprise a mixture of cleavable sequencing primers and non-cleavable sequencing primers.

5. A method according to any one of claims 1 to 4, wherein the method further comprises a step of linearising the first polynucleotide sequence and the second polynucleotide sequence following step (b).

6. A method according to claim 1, wherein the first adaptor of step (g) comprises two or at least two cleavable sites and the second adaptor of step (g) comprises two or at least two cleavable sites.

7. A method according to claim 6, wherein the first adaptor comprises one cleavable site and the second adaptor comprises one complement of a cleavable site and wherein the first and second immobilised primers comprise one cleavable site.

8. A method according to any one of claims 6 to 7, further comprising a step of removing a sequence between the cleavable sites.

5 9. A method according to any one of claims 1 to 8, further comprising a step of blocking 3'-ends of the first polynucleotide sequences and the second polynucleotide sequences.

10 10. A method according to any one of claims 1 and 6 to 9, wherein following step (j) the method further comprises the step of applying a plurality of first and second sequencing primers, wherein the first or second sequencing primers comprise a mixture of blocked sequencing primers and unblocked sequencing primers, wherein preferably the blocked sequencing primer comprises a blocking group at a 3' end of the sequencing primer, wherein preferably the blocking group is
15 selected from the group consisting of: a hairpin loop, a deoxynucleotide, a deoxyribonucleotide, a hydrogen atom instead of a 3'-OH group, a phosphate group, a phosphorothioate group, a propyl spacer, a modification blocking the 3'-hydroxyl group, or an inverted nucleobase, and wherein preferably the ratio of blocked to unblocked sequencing primers is 1:1.

20 11. A method according to any one of claims 1 to 10, wherein the cleavable sites are restriction sites for a nicking endonuclease.

25 12. A method according to any one of claims 1 to 11, wherein the cluster is formed by bridge amplification.

30 13. A method according to claim 1, wherein the step of (j) of cleaving at the cleavage sites is carried out following bridge amplification and before linearization of the extended sequences.

35 14. A method according to any one of claims 1 to 13, wherein the first polynucleotide sequence is a forward strand of a double-stranded polynucleotide to be identified and the second polynucleotide sequence is a reverse strand of a double-stranded polynucleotide to be identified.

15. A method according to any of claims 1 to 5, wherein the method comprises cleaving the first cleavable site in the first sequencing primer and sequencing the nucleobases in the first polynucleotide sequence.
- 5 16. A method according to any of claims 1 to 5 or 15, wherein the method further comprises subsequently cleaving the second cleavable site in the second sequencing primer and sequencing the nucleobases in the first polynucleotide sequence, wherein preferably, the method further comprises blocking the 3' end of the first polynucleotide sequence prior to sequencing the second
10 polynucleotide.
17. A method according to claims 1 to 5, wherein the method comprises concurrently cleaving the first cleavable site in the first sequencing primer and cleaving the second cleavable site in the second sequencing primer and sequencing
15 nucleobases in the first polynucleotide sequence and the second polynucleotide sequence, wherein preferably the method comprises applying a plurality of sequencing primers, wherein the first or second sequencing primers comprise a mixture of cleavable sequencing primers and non-cleavable sequencing primers, wherein preferably, the first or second sequencing primers comprise a 1:1 mixture
20 of cleavable sequencing primers and non-cleavable sequencing primers.
18. A method according to any one of claims 1 to 17, wherein the method comprises concurrently sequencing the first and second polynucleotide sequences, wherein preferably, the step of concurrently sequencing nucleobases comprises
25 performing sequencing-by-synthesis or sequencing-by-ligation; or wherein the method comprises treatment with a polymerase and a 5'-3' exonuclease or a polymerase with 5' to 3' exonuclease activity; or wherein the method comprises treatment with a strand displacement polymerase (e.g. phi29).
- 30 19. A sequencing kit comprising a plurality of first and second sequencing primers, wherein the first and second sequencing primers comprise at least one cleavage site, wherein preferably, the first and second sequencing primers comprise a mixture of cleavable sequencing primers and un-cleavable sequencing primers.
- 35 20. A library preparation kit comprising of a plurality of first adaptors and a plurality of second adaptors, wherein the first adaptors comprise at least one cleavable

site and the second adaptors comprise a complement of at least one cleavable site, and wherein the first and second adaptors comprises an immobilised primer-binding sequence and a sequencing primer binding sequence or complements thereof.

5

21. The library preparation kit of claim 20, wherein the kit further comprises first and second sequencing primers and wherein the first and second sequencing primers comprise a mixture of blocked sequencing primers and unblocked sequencing primers, and wherein the ratio of blocked to unblocked sequencing primers is 1:1.

10

22. The sequencing kit of claim 19 or the library preparation kit of claim 20 or 21, wherein the kit further comprises a polymerase and a 5' to 3' exonuclease or a polymerase with 5' to 3' exonuclease activity.

15

23. The sequencing kit of claim 19, or the library preparation kit of claim 20 or 21, wherein the kit further comprises a single-stranded exonuclease and cleavable effectors, such as nickases.

20

24. A data processing device comprising means for carrying out a method according to any one of claims 1 to 18, wherein preferably the data processing device is a polynucleotide sequencer.

25

25. A computer program product comprising instructions which, when the program is executed by a processor, cause the processor to carry out a method according to any one of claims 1 to 18.

30

26. A computer-readable storage medium comprising instructions which, when executed by a processor, cause the processor to carry out a method according to any one of claims 1 to 18.

27. A computer-readable data carrier having stored thereon a computer program product according to claim 25.

35

28. A data carrier signal carrying a computer program product according to claim 25.

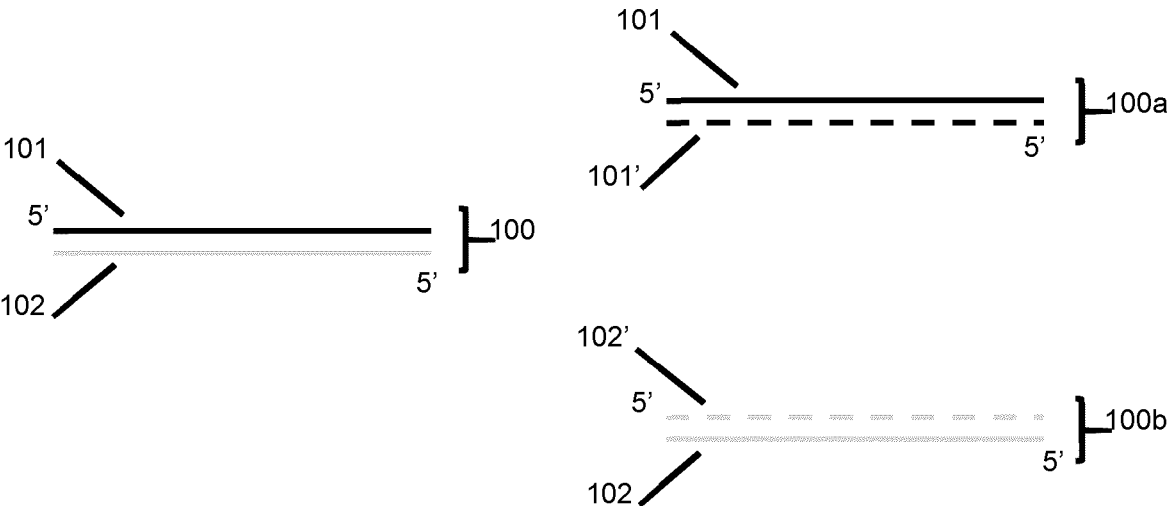


Figure 1

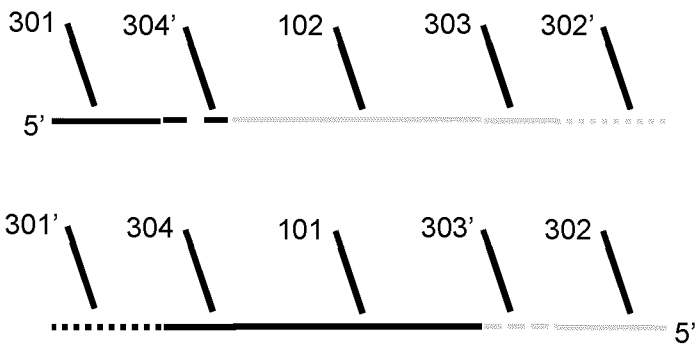


Figure 2



Figure 3

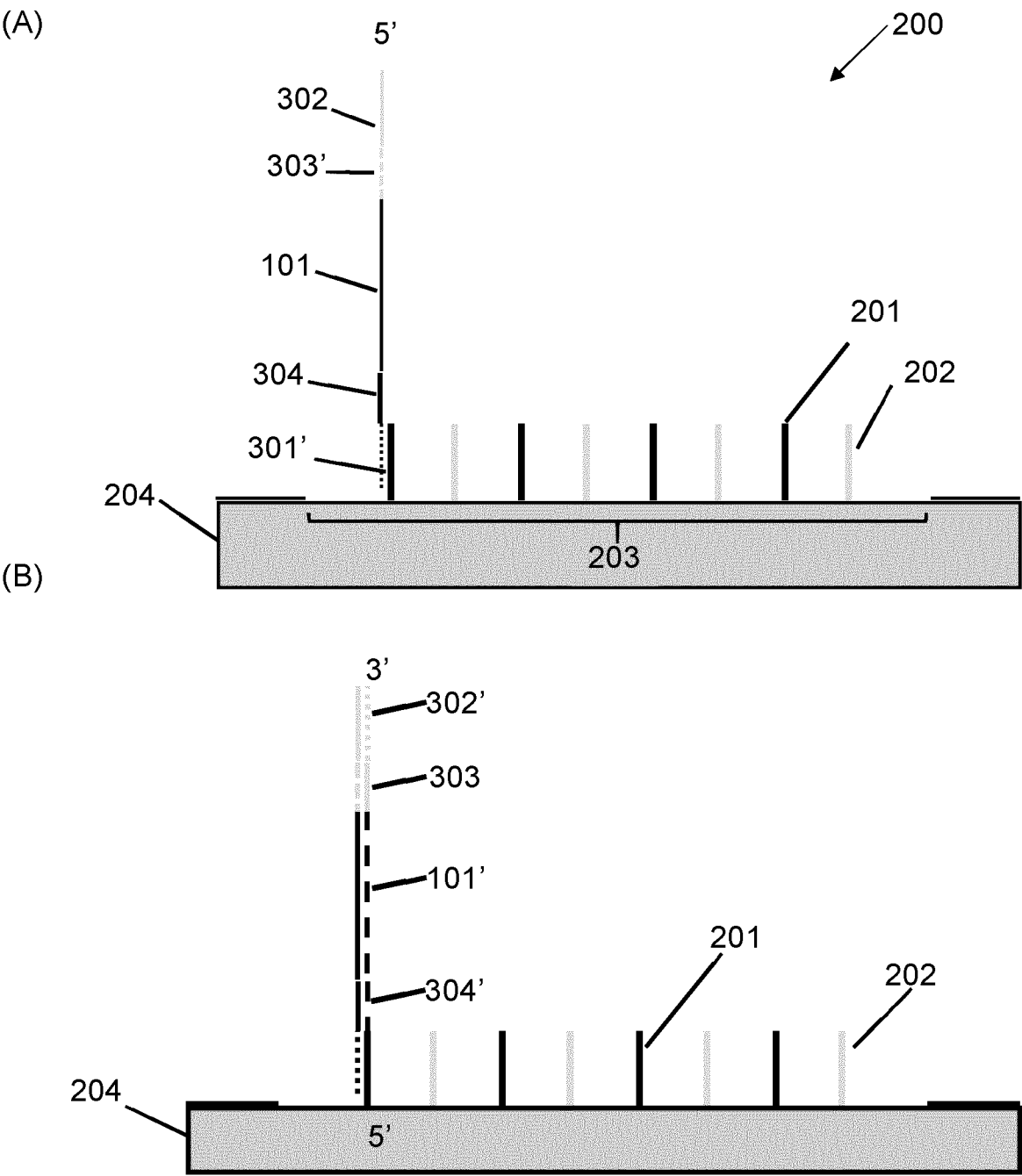


Figure 4

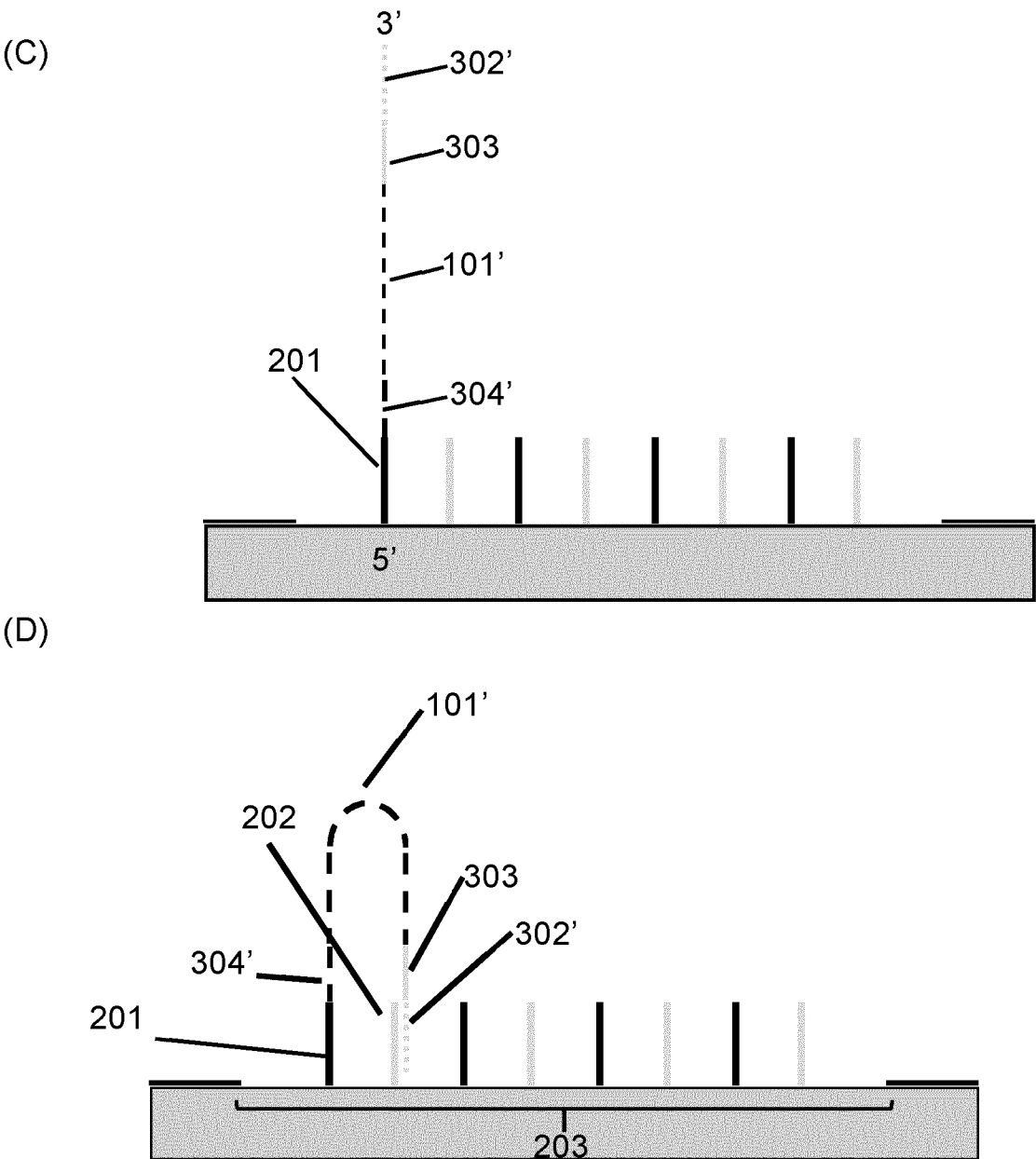


Figure 4 (cont.)

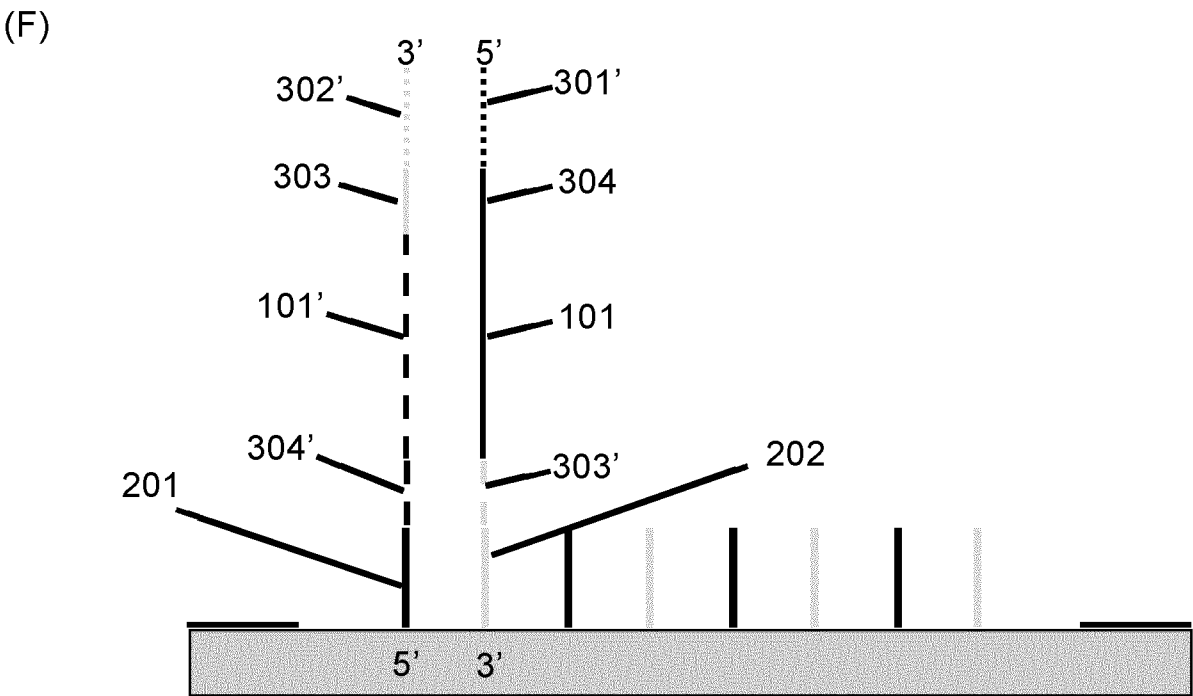
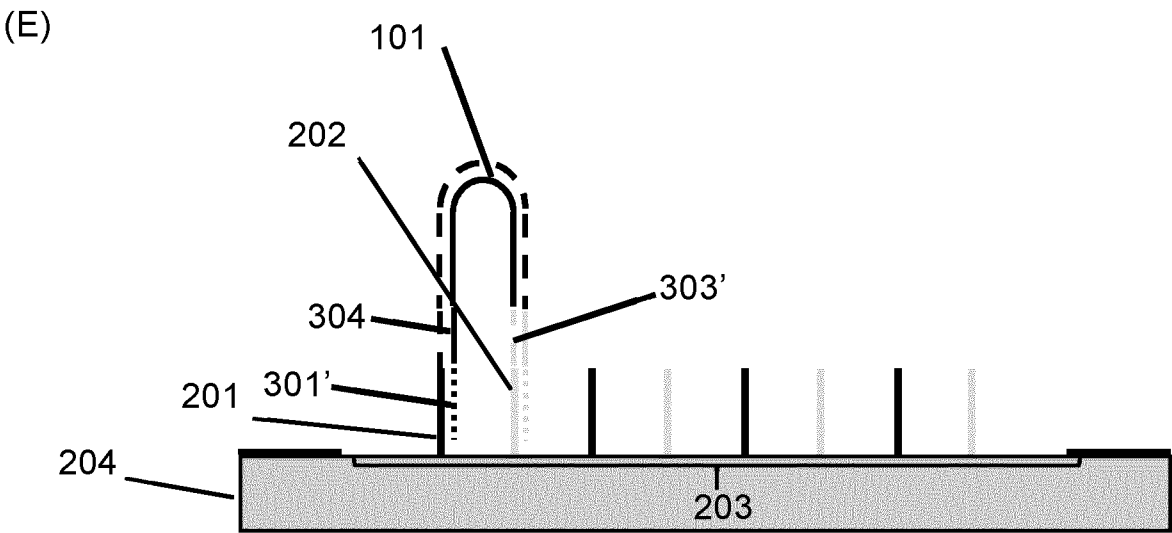
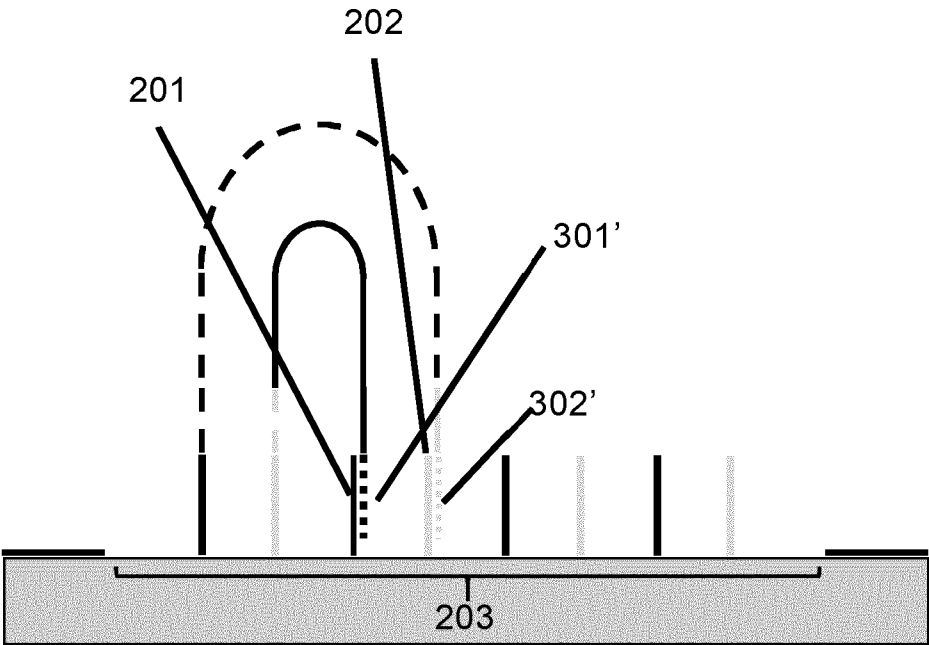


Figure 4 (cont.)

(G)



(H)

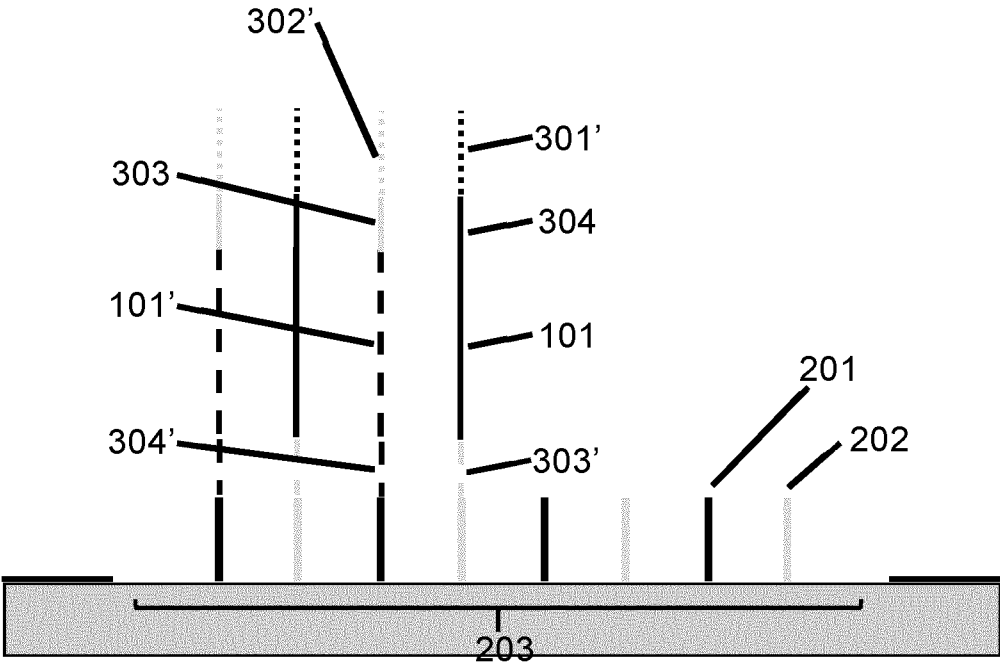


Figure 4 (cont.)

(I)

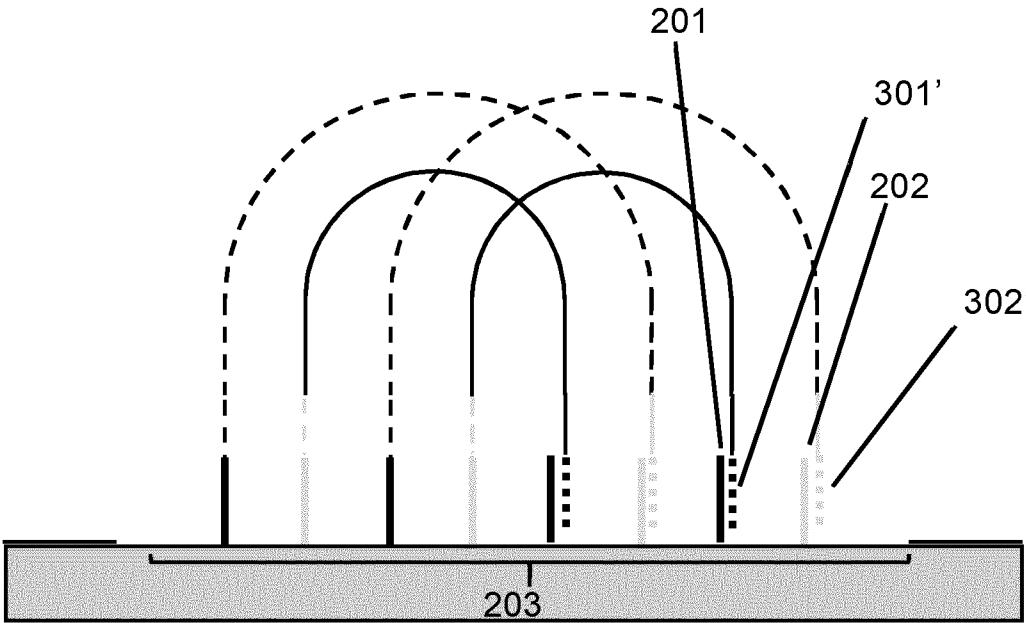


Figure 4 (cont.)

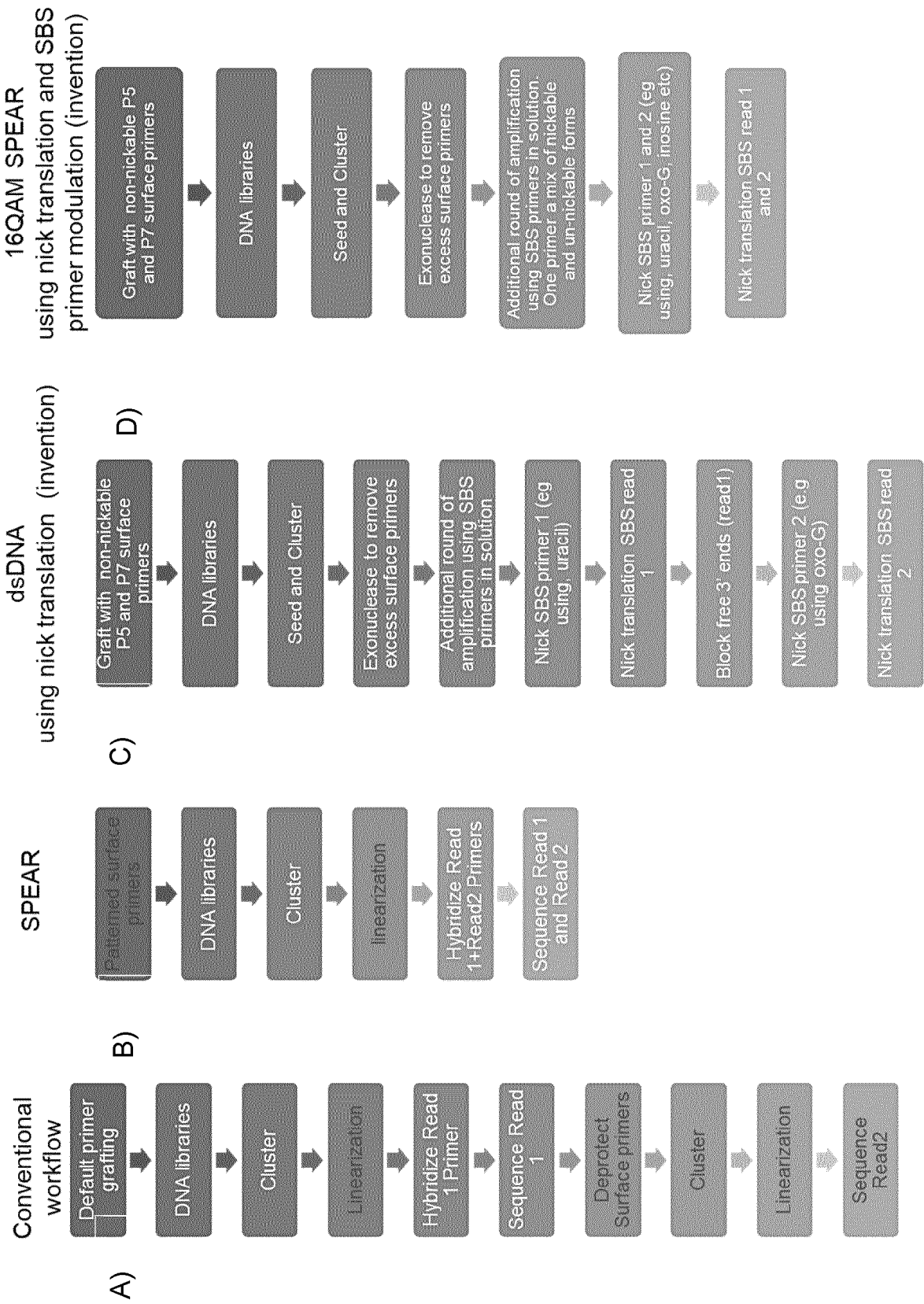


Figure 5

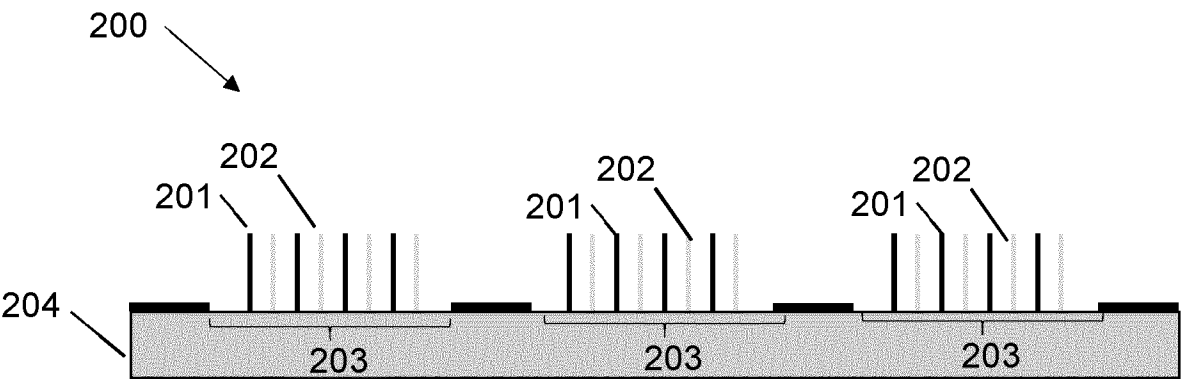


Figure 6

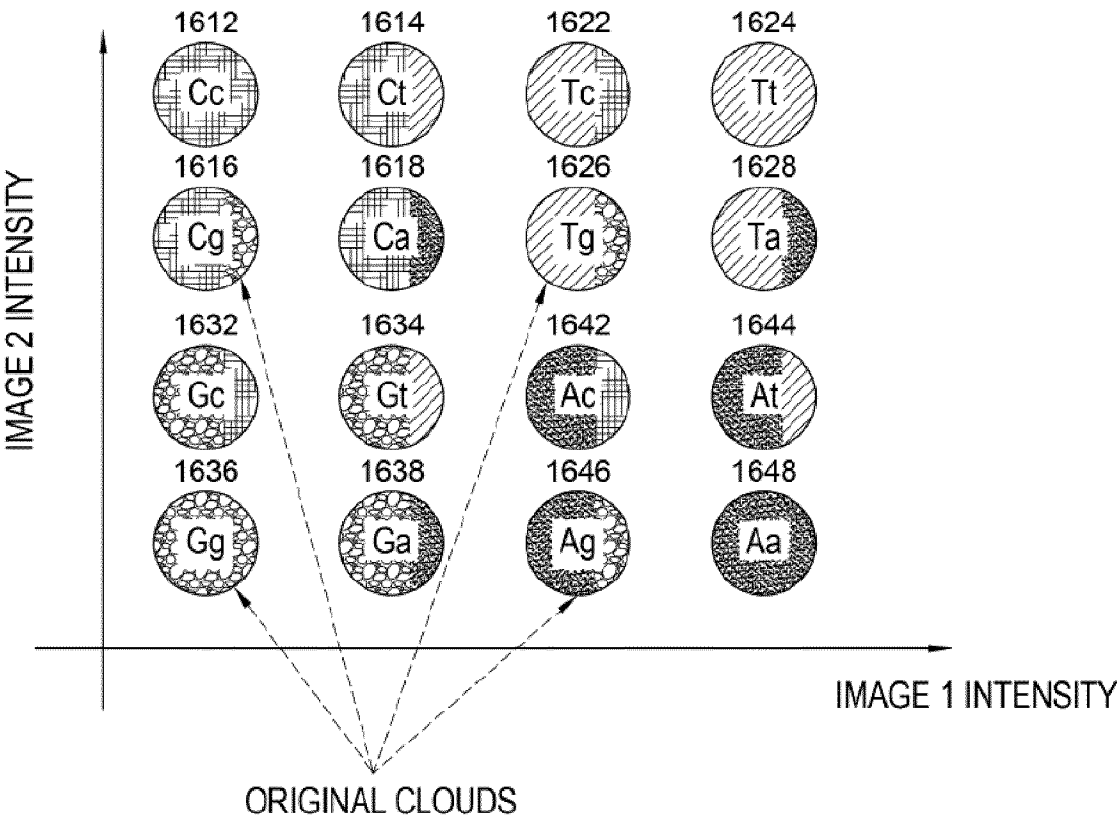


Figure 7

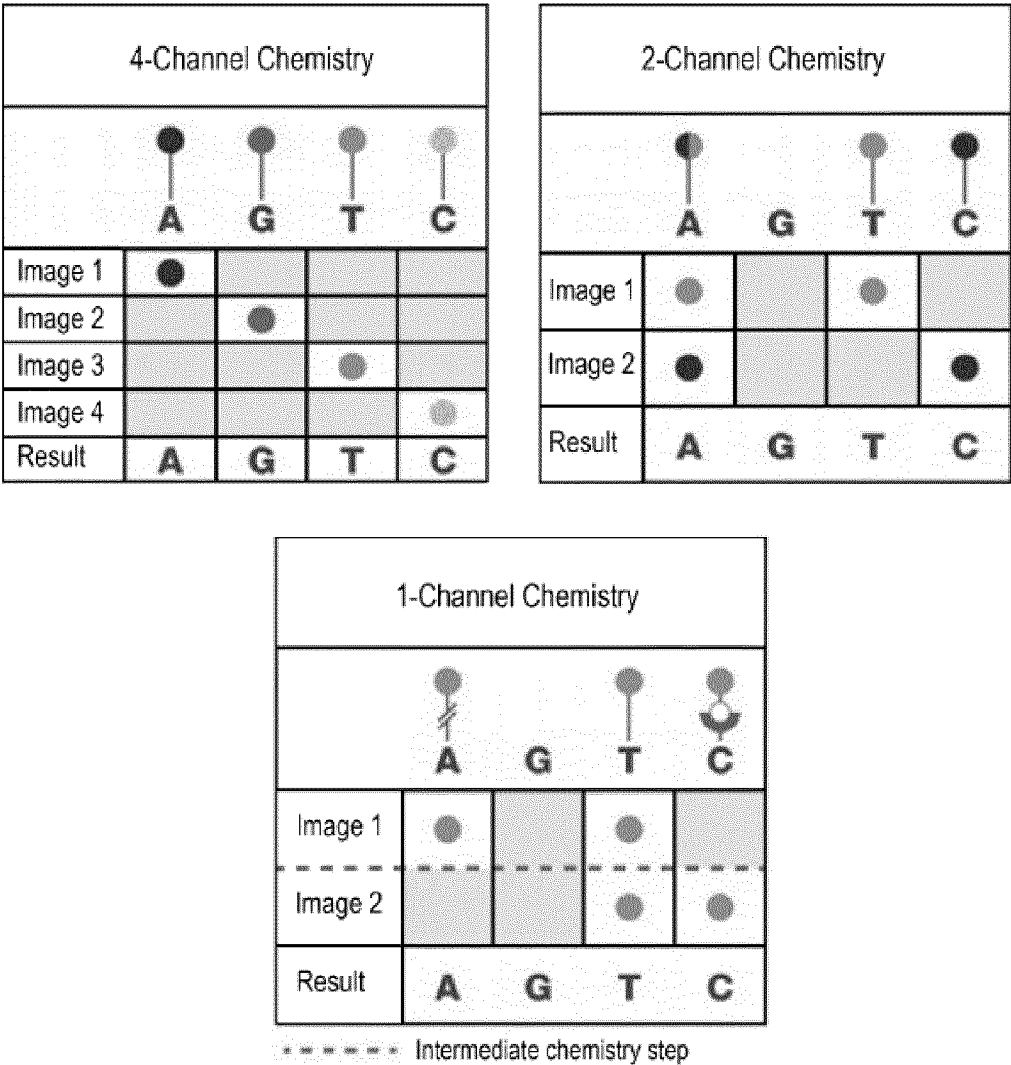
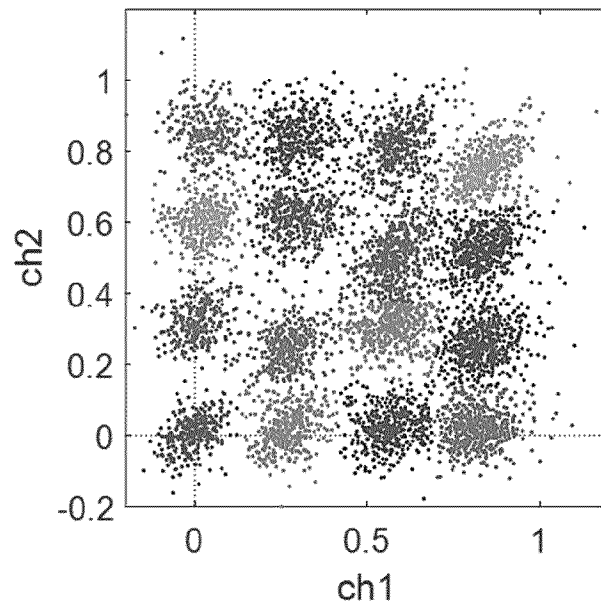
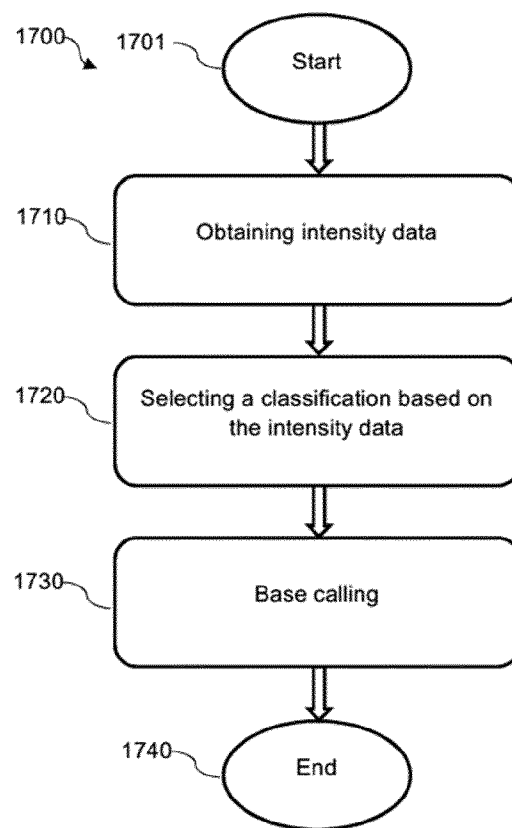


Figure 8

**Figure 9****Figure 10**

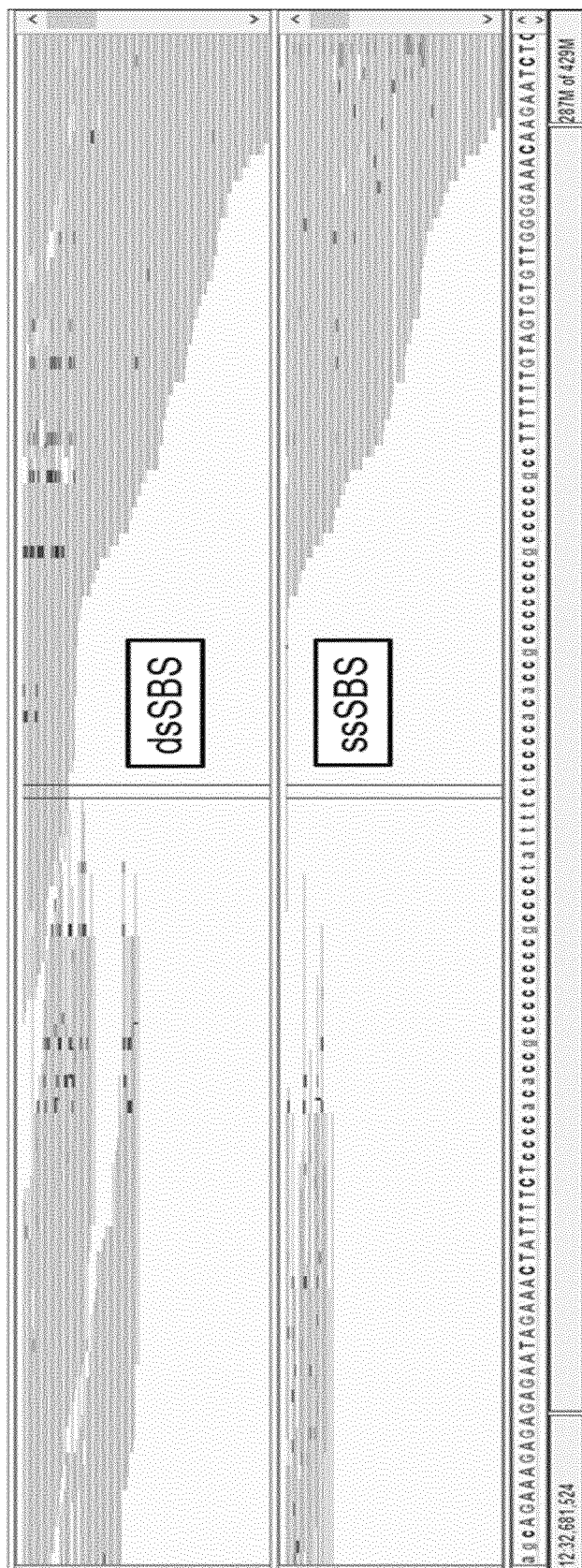


Figure 11

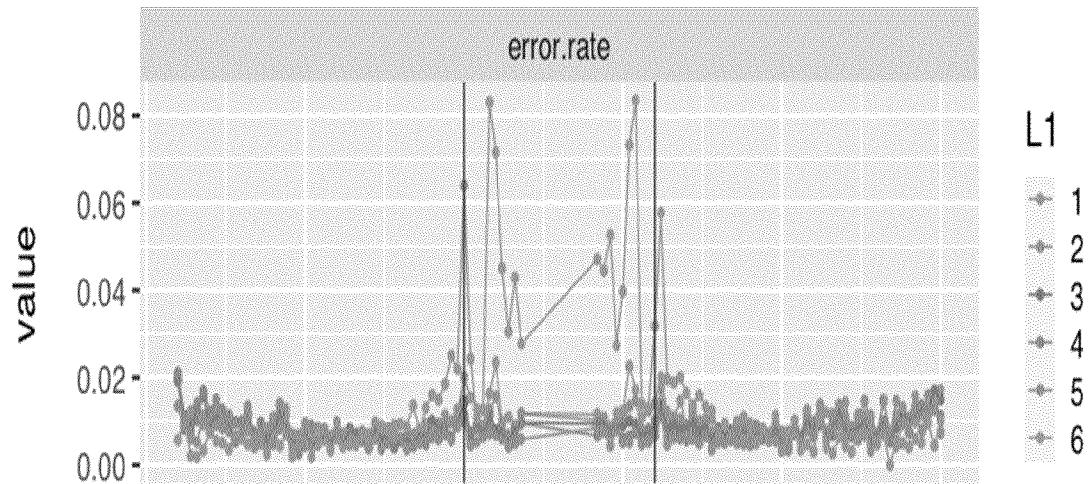


Figure 12

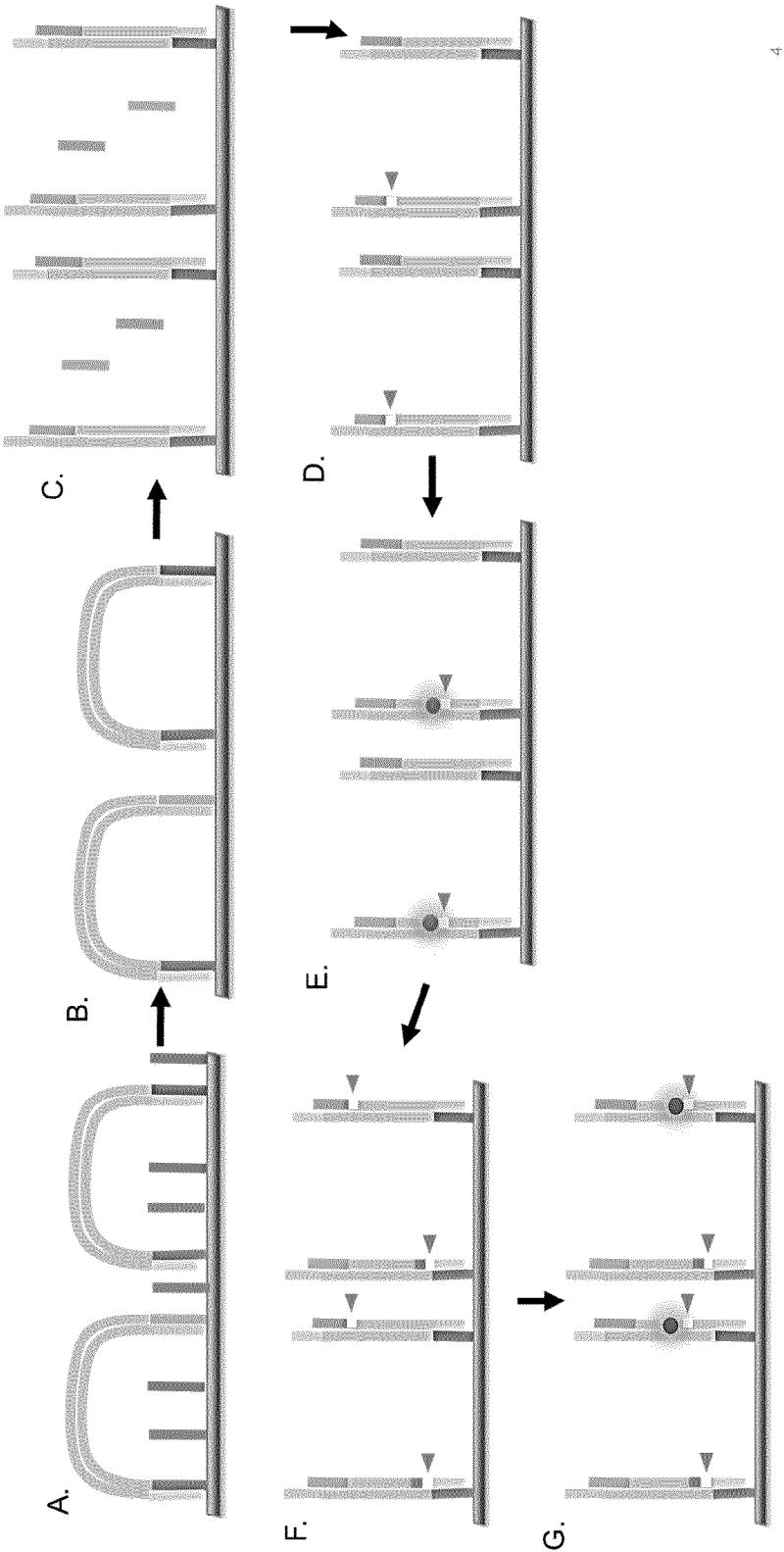


Figure 13

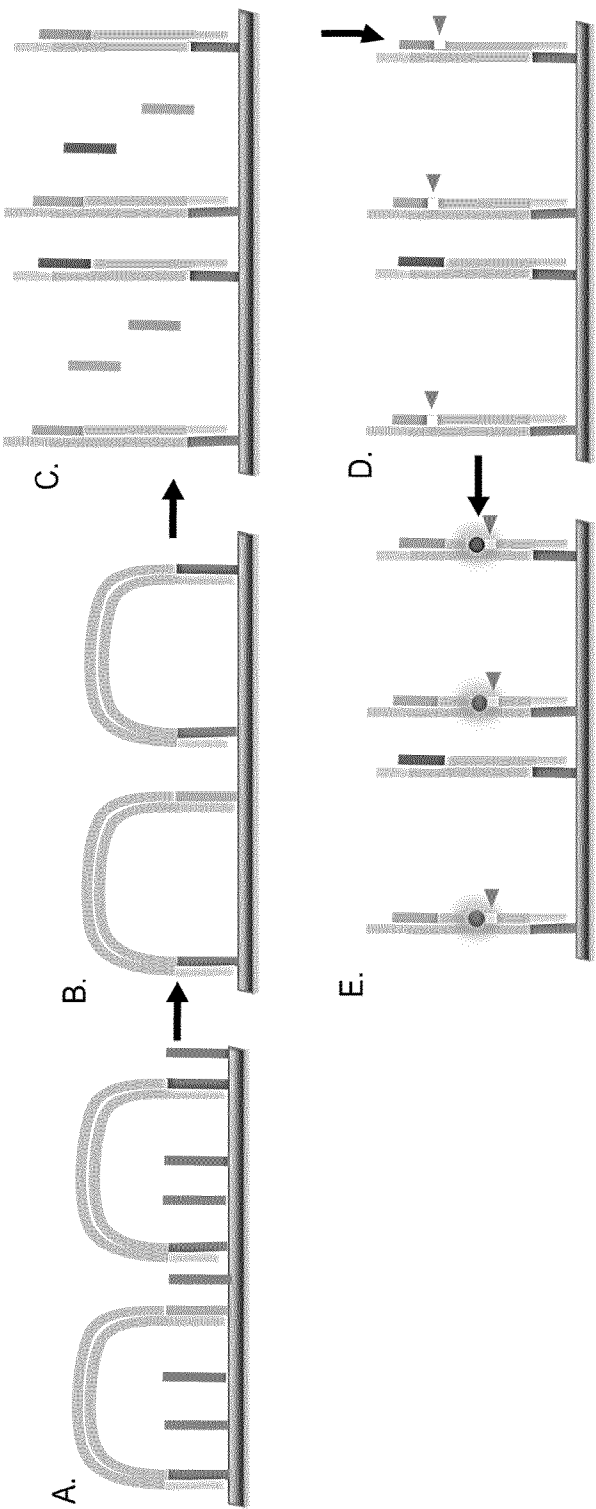


Figure 14

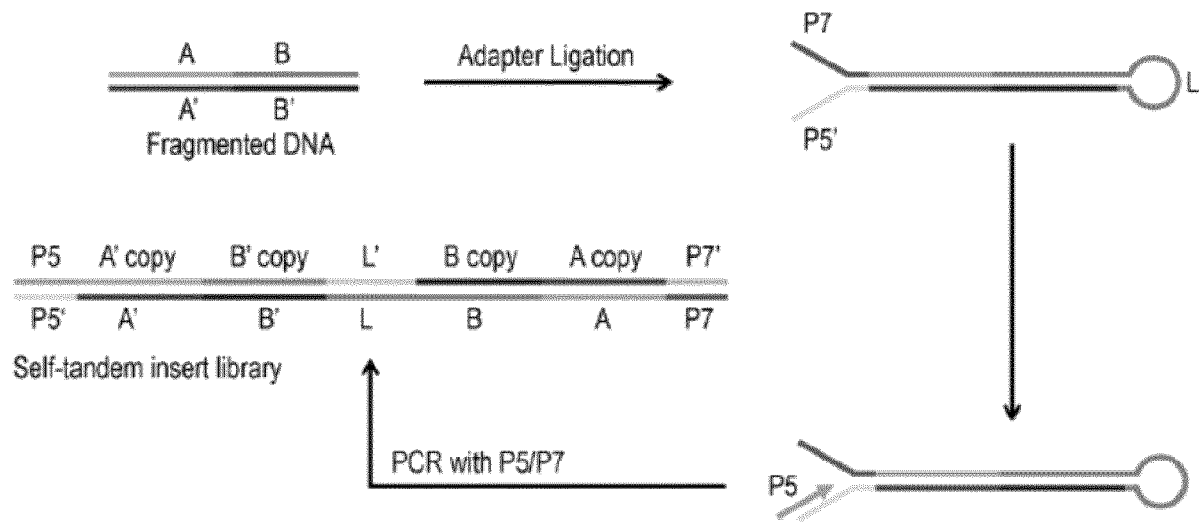


Figure 15

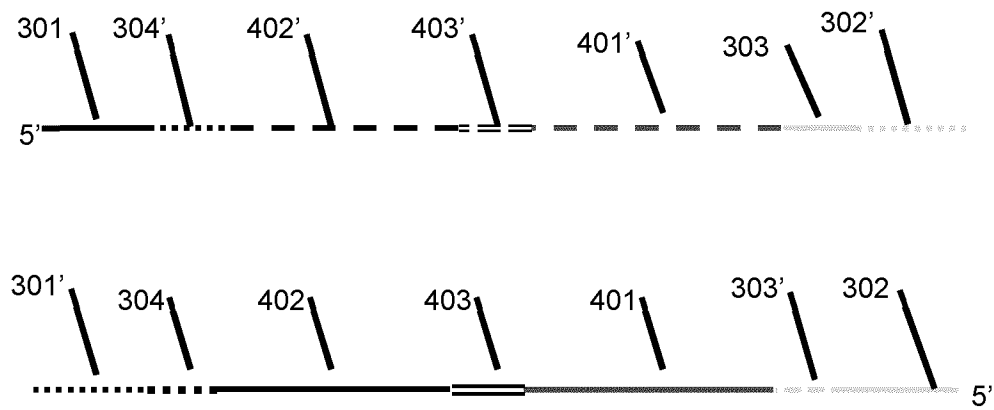


Figure 16

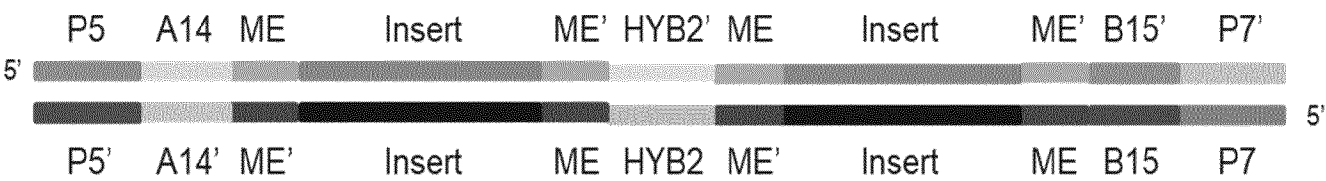


Figure 17

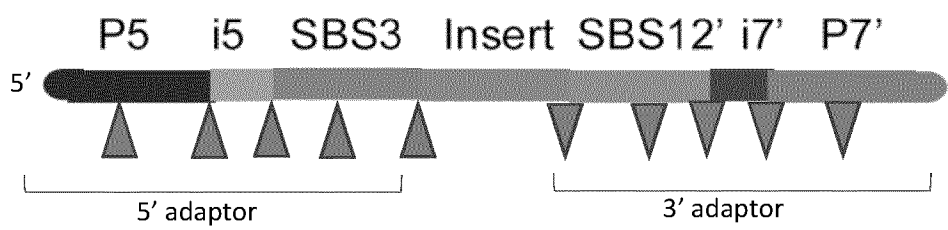


Figure 18

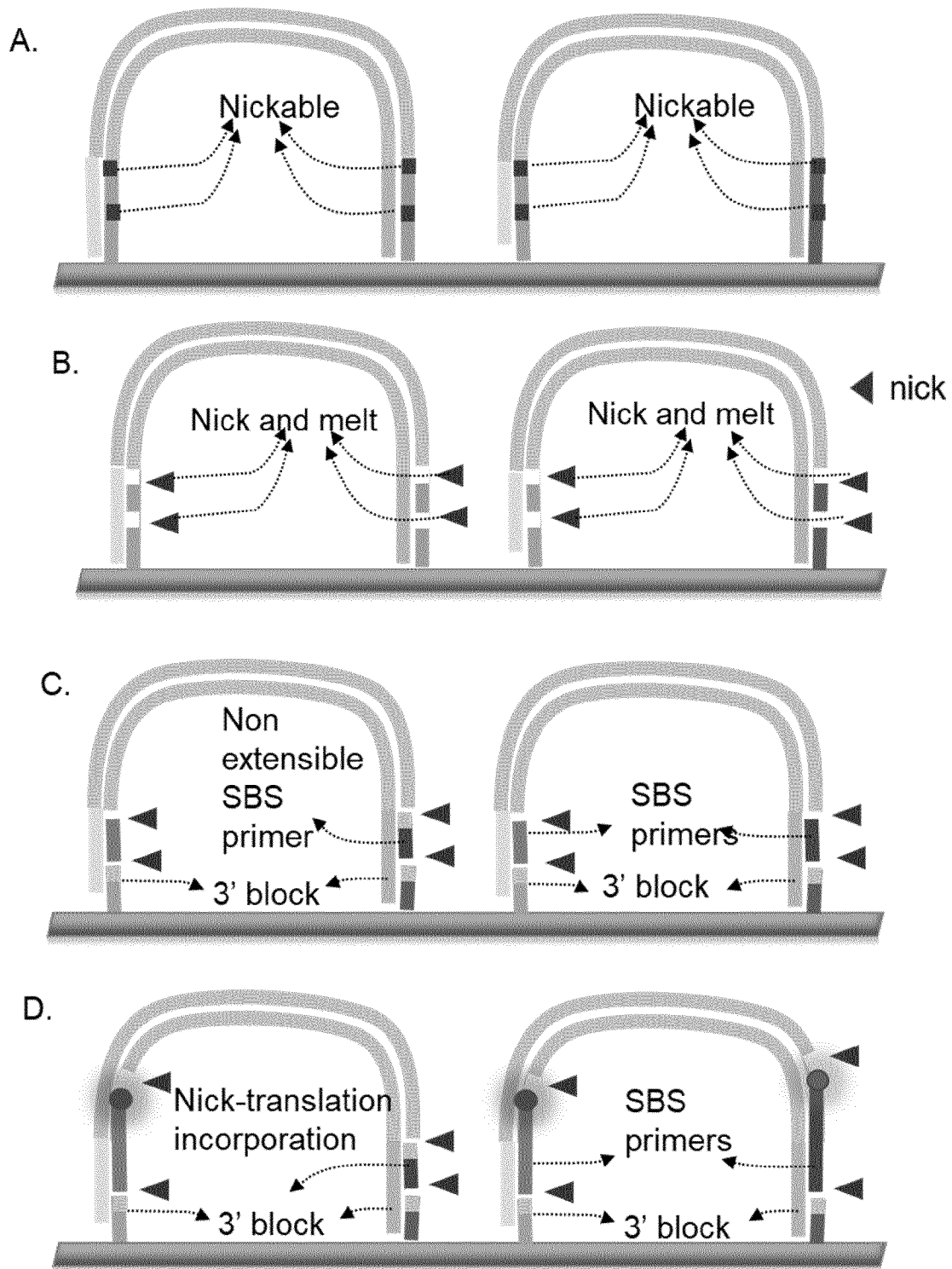


Figure 19

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2024/076525

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, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2023/126457 A1 (ILLUMINA CAMBRIDGE LTD [GB]; ILLUMINA INC [US]) 6 July 2023 (2023-07-06) the whole document -----	1-5,9, 11, 14-19, 22-28
A	WO 2008/041002 A2 (SOLEXA LTD [GB]; RIGATTI ROBERTO [GB] ET AL.) 10 April 2008 (2008-04-10) the whole document -----	1-5,9, 11, 14-19, 22-28
A	WO 2007/010251 A2 (SOLEXA LTD [GB]; LIU XIAOHAI [GB] ET AL.) 25 January 2007 (2007-01-25) the whole document ----- -/-	1-5,9, 11, 14-19, 22-28



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

11 December 2024

Date of mailing of the international search report

12/02/2025

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

Bradbrook, Derek

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2024/076525

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2007/010252 A1 (SOLEXA LTD [GB]; SMITH GEOFFREY PAUL [GB]) 25 January 2007 (2007-01-25) the whole document -----	1-5,9, 11, 14-19, 22-28
A	WO 2017/117235 A1 (OMNIOME INC [US]) 6 July 2017 (2017-07-06) the whole document -----	1-5,9, 11, 14-19, 22-28

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2024/076525

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2024/076525

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims;; it is covered by claims Nos.:
2 - 5, 15 - 17, 19 (completely); 1, 9, 11, 14, 18, 22 - 28 (partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 2-5, 15-17, 19 (completely); 1, 9, 11, 14, 18, 22-28 (partially)

A method of sequencing a first and second polynucleotide sequence, wherein the method comprises preparing the first and second polynucleotide sequence for sequencing and sequencing the nucleobases in the first and second polynucleotide sequence, wherein the first and second polynucleotide sequences are sequenced using nick translation; wherein preparing a first polynucleotide sequence and a second polynucleotide sequence for sequencing comprises steps a to e of claim 1; a sequencing kit comprising first and second sequencing primers; a data processing device and associated computer-related products adapted to carry out said method.

2. claims: 6-8, 10, 12, 13, 20, 21 (completely); 1, 9, 11, 14, 18, 22-28 (partially)

A method of sequencing a first and second polynucleotide sequence, wherein the method comprises preparing the first and second polynucleotide sequence for sequencing and sequencing the nucleobases in the first and second polynucleotide sequence, wherein the first and second polynucleotide sequences are sequenced using nick translation; wherein preparing a first polynucleotide sequence and a second polynucleotide sequence for sequencing comprises steps f to j or claim 1; a library preparation kit comprising a plurality of first and second adapters; a data processing device and associated computer-related products adapted to carry out said method.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2024/076525

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2023126457 A1	06-07-2023	AU 2022424380 A1	18-01-2024
		CA 3222937 A1	06-07-2023
		EP 4457364 A1	06-11-2024
		US 2023212667 A1	06-07-2023
		WO 2023126457 A1	06-07-2023

WO 2008041002 A2	10-04-2008	CN 101663405 A	03-03-2010
		DK 2084295 T3	12-08-2013
		EP 2084295 A2	05-08-2009
		EP 2453023 A2	16-05-2012
		EP 3670672 A1	24-06-2020
		ES 2424155 T3	27-09-2013
		US 2009088327 A1	02-04-2009
		US 2011014657 A1	20-01-2011
		US 2011223601 A1	15-09-2011
		US 2012094848 A1	19-04-2012
		US 2012309634 A1	06-12-2012
		US 2013274115 A1	17-10-2013
		US 2014336057 A1	13-11-2014
		US 2016090623 A1	31-03-2016
		WO 2008041002 A2	10-04-2008

WO 2007010251 A2	25-01-2007	EP 1910559 A2	16-04-2008
		EP 2298930 A1	23-03-2011
		ES 2398808 T3	21-03-2013
		HK 1154636 A1	27-04-2012
		US 2009118128 A1	07-05-2009
		US 2012208194 A1	16-08-2012
		US 2014349891 A1	27-11-2014
		US 2015343410 A1	03-12-2015
		US 2017021325 A1	26-01-2017
		US 2018353926 A1	13-12-2018
		WO 2007010251 A2	25-01-2007

WO 2007010252 A1	25-01-2007	AT E493513 T1	15-01-2011
		EP 1910560 A1	16-04-2008
		US 2009181370 A1	16-07-2009
		US 2012053074 A1	01-03-2012
		US 2013171648 A1	04-07-2013
		US 2015203911 A1	23-07-2015
		US 2016160277 A1	09-06-2016
		US 2017327885 A1	16-11-2017
		US 2020216896 A1	09-07-2020
		US 2023193384 A1	22-06-2023
		WO 2007010252 A1	25-01-2007

WO 2017117235 A1	06-07-2017	NONE	
