



(51) International Patent Classification:

C12Q 1/6869 (2018.01) C12Q 1/6876 (2018.01)

(21) International Application Number:

PCT/SG2021/050426

(22) International Filing Date:

21 July 2021 (21.07.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

10202006969U 21 July 2020 (21.07.2020) SG

(71) Applicants: **AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH** [SG/SG]; 1 Fusionopolis Way, #20-10 Connexis North Tower, Singapore 138632 (SG). **PYROBETT PTE LTD.** [SG/SG]; 8 Commonwealth Lane, #02-02, Singapore 149555 (SG).

(72) Inventors: **LEZHAHA, Alexander**; c/o Genome Institute of Singapore, 60 Biopolis Street, #02-01, Genome, Singapore 138672 (SG). **SARMA, Suman**; c/o Genome Institute of Singapore, 60 Biopolis Street, #02-01, Genome, Singapore 138672 (SG). **KONG, Say Li**; c/o Genome Institute of Singapore, 60 Biopolis Street, #02-01, Genome, Singapore 138672 (SG). **CORBETT, John**; c/o Pyrobett Pte Ltd., 8 Commonwealth Lane, #02-02, Singapore 149555 (SG).

(74) Agent: **DONALDSON & BURKINSHAW LLP**; 24 Raffles Place, #15-00 Clifford Centre, Singapore 048621 (SG).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO,

(54) Title: A PYROSEQUENCING METHOD

Dispensation 1: Measure A peak height
Dispensation 2: Test for G peak

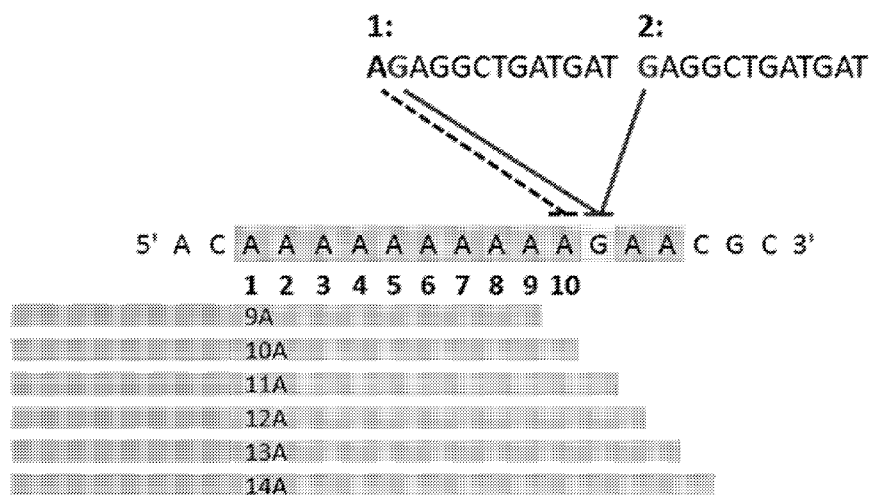


FIG. 2

(57) Abstract: There is provided a method, such as a pyrosequencing method, for sequencing a polynucleotide sample comprising a homopolymeric region R_s , the method comprising: providing template polynucleotides comprising the polynucleotide sample and copies thereof; providing a plurality of R_p -containing primers comprising, at each of their 3' end, a homopolymeric region R_p comprising nucleotides that are complementary to the nucleotides in the homopolymeric region R_s , and contacting the plurality of R_p -containing primers with the template polynucleotides under conditions suitable for hybridization and synthesis of complementary polynucleotides, wherein the homopolymeric regions R_p in the plurality of R_p -containing primers are of different lengths. Also provided are related primers and kits.



NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,
SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

- *of inventorship (Rule 4.17(iv))*

Published:

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

A PYROSEQUENCING METHOD

TECHNICAL FIELD

The present disclosure relates broadly to a method for sequencing a sample comprising a repeat region, such as a homopolymeric region, and related primers and kits.

BACKGROUND

The determination of target sequences with high accuracy is important for annotating the target sequences of tested sample either for verification purpose or de novo assembly for discovery purpose. However, this is often complicated at regions of high polymorphisms or repeated sequences. Sanger sequencing could only provide semi-quantitative readout, it could not give accurate readout for regions with high polymorphism of various frequency or repeated sequences. Though next-generation sequencing technology could provide qualitative and quantitative information for sequences in regions of high polymorphism, it could not provide accurate readout in regions of repeated sequences with routine mapping and analysis pipeline. Pyrosequencing provides accurate and quantitative readout without expensive equipment and complicated workflow, thereby addressing these limitations of Sanger and next-generation sequencing. However, pyrosequencing has low accuracy in determining repeated sequences in repeat regions.

Thus, there is a need to provide a method for sequencing a sample comprising a repeat region, such as a homopolymeric region, and related primers and kits that address or at least ameliorate the above-mentioned problems.

SUMMARY

In one aspect, there is provided a pyrosequencing method for sequencing a polynucleotide sample comprising a homopolymeric region R_s , the method comprising: providing template polynucleotides comprising the polynucleotide sample and copies thereof; providing a plurality of R_p -containing primers comprising, at each of their 3' end, a homopolymeric region R_p comprising nucleotides that are

complementary to the nucleotides in the homopolymeric region R_s , and contacting the plurality of R_p -containing primers with the template polynucleotides under conditions suitable for hybridization and synthesis of complementary polynucleotides, wherein the homopolymeric regions R_p in the plurality of R_p -containing primers are of different lengths.

In one embodiment, the homopolymeric regions R_p in the plurality of R_p -containing primers comprise a length selected from: X nucleotides (nt), $X+1$ nt, $X+2$ nt, $X+3$ nt, $X+4$ nt, $X+5$ nt, $X+6$ nt, $X+7$ nt, $X+8$ nt, $X+9$ nt, $X+10$ nt and combinations thereof, wherein X is an integer no less than 3.

In one embodiment, the plurality of R_p -containing primers comprise at least three R_p -containing primers.

In one embodiment, contacting the plurality of R_p -containing primers with the template polynucleotides comprises partitioning the template polynucleotides into a plurality of partitions of template polynucleotides and contacting each partition with a R_p -containing primer with a homopolymeric region R_p of a different length.

In one embodiment, contacting the plurality of R_p -containing primers with the template polynucleotides comprises applying the plurality of R_p -containing primers sequentially to the template polynucleotides.

In one embodiment, the method further comprises adding a nucleotide N_1 to the template polynucleotides after the template polynucleotides are contacted with a R_p -containing primer, the nucleotide N_1 corresponding to the nucleotides in the homopolymeric region R_p ; and determining whether the nucleotide N_1 is incorporated in the complementary polynucleotides.

In one embodiment, the method further comprises adding a nucleotide N_2 to the template polynucleotides after the template polynucleotides are contacted with a R_p -containing primer, wherein nucleotide N_2 is different from the nucleotides in the homopolymeric region R_p , optionally wherein nucleotide N_2 is complementary to an adjacent nucleotide that is directly next to the homopolymeric region R_s on a 5' side in the polynucleotide sample; and determining whether the nucleotide N_2 is incorporated in the complementary polynucleotides.

In one embodiment, the method further comprises identifying the R_p -containing primer contacted with the template polynucleotides when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides; and concluding that the homopolymeric region R_s has a length

corresponding to the length of the homopolymeric region R_p of the identified R_p -containing primer.

In one embodiment, the method further comprises identifying the R_p -containing primer contacted with the template polynucleotides when nucleotide N_1 is incorporated in the complementary polynucleotides; determining the number of nucleotide N_1 incorporated; and concluding that the homopolymeric region R_s has a length corresponding to the sum of the number of nucleotide N_1 incorporated and the length of the homopolymeric region R_p of the identified R_p -containing primer.

In one embodiment, the method further comprises verifying that nucleotide N_1 is incorporated in the complementary polynucleotides of template polynucleotides contacted with a R_p -containing primer comprising a shorter homopolymeric region R_p than that of the identified R_p -containing primer before the concluding step.

In one embodiment, wherein where the nucleotide N_1 is not incorporated in the complementary polynucleotides of the template polynucleotides contacted with the plurality of R_p -containing primers, the method further comprises:

- a) providing a further R_p -containing primer comprising a shorter homopolymeric region R_p than those of the plurality of R_p -containing primers;
- b) contacting the template polynucleotides with the further R_p -containing primer;
- c) determining whether the nucleotide N_1 and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of the template polynucleotides of step b); and
- d) concluding that the homopolymeric region R_s has a length corresponding to the length of the homopolymeric region R_p of the further R_p -containing primer when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of step c).

In one embodiment, wherein the nucleotide N_1 is incorporated in the complementary polynucleotides of the template polynucleotides contacted with the plurality of R_p -containing primers, the method further comprises:

- a) providing a further R_p -containing primer comprising a longer homopolymeric region R_p than those of the plurality of R_p -containing primers;
- b) contacting the template polynucleotides with the further R_p -containing primer;
- c) determining whether the nucleotide N_1 and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of the template polynucleotides of step b); and

- d) concluding that the homopolymeric region R_s has a length corresponding to the length of the homopolymeric region R_p of the further R_p -containing primer when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of step c).

In one embodiment, the method further comprises verifying that nucleotide N_1 is incorporated in the complementary polynucleotides of the template polynucleotides contacted with a R_p -containing primer comprising a shorter homopolymeric region R_p than that of the further R_p -containing primer before the concluding step.

In one embodiment, the method further comprises amplifying the polynucleotide sample and/or cloning the polynucleotide sample into a vector to obtain the template polynucleotides.

In one embodiment, the method further comprises determining a length of a homopolymeric region in a reference sequence of the polynucleotide sample prior to providing the plurality of R_p -containing primers.

In one embodiment, the plurality of R_p -containing primers comprises a R_p -containing primer which homopolymeric region R_p has a length corresponding to the length of the homopolymeric region in the reference sequence; optionally, a R_p -containing primer which homopolymeric region R_p has a length shorter than the length of the homopolymeric region in the reference sequence; and further optionally, a R_p -containing primer which homopolymeric region R_p has a length longer than the length of the homopolymeric region in the reference sequence.

In one aspect, there is provided a plurality of R_p -containing primers for use in embodiments of the method described herein, wherein the plurality of R_p -containing primers comprises, at each of their 3' end, a homopolymeric region R_p , further wherein the homopolymeric regions R_p in the plurality of R_p -containing primers are of different lengths.

In one embodiment of the plurality of R_p -containing primers, the homopolymeric regions R_p in the plurality of R_p -containing primers comprise a length selected from: X nucleotides (nt), X+1 nt, X+2 nt, X+3 nt, X+4 nt, X+5 nt, X+6 nt, X+7 nt, X+8 nt, X+9 nt, X+10 nt and combinations thereof, wherein X is an integer no less than 3.

In one embodiment of the plurality of R_p -containing primers, the plurality of R_p -containing primers comprise at least three R_p -containing primers.

In one aspect, there is provided a kit for use in embodiments of the method described herein, the kit comprising: embodiments of the plurality of R_p -containing

primers described herein; and at least one more component selected from the group consisting of: nucleotides, polymerase, adenosine phosphosulfate (APS), ATP sulfurylase, luciferin, luciferase and apyrase.

DEFINITIONS

As used herein, the term "nucleotide" refers to any natural or non-natural nucleotide, including modified nucleotides (e.g., methylated or biotinylated nucleotides), nucleotide analogs and nucleotide mimics that can be incorporated into a polynucleotide by a polymerase. Examples of a nucleotide include, but are not limited to, deoxyribonucleoside mono-, di-, and triphosphate; deoxyadenosine mono-, di- and triphosphate; deoxyguanosine mono-, di- and triphosphate; deoxythymidine mono-, di- and triphosphate; deoxycytidine mono-, di- and triphosphate; and 2'-deoxyadenosine-5'-(α -thio) mono-, di- and triphosphate. The term "polynucleotide" encompasses polymeric forms of nucleotides of any length.

As used herein, the term "sequencing" refers to the determination of an order of nucleotides/base sequences in a polynucleotide/nucleic acid sample. The term "sequencing by synthesis" refers to the sequencing of a polynucleotide/nucleic acid sample by synthesis of a complementary strand. The term "pyrosequencing" refers to any sequencing by synthesis method that is based on the detection of a pyrophosphate group that is generated when a nucleotide is incorporated in a growing polynucleotide/nucleic acid, e.g., a growing complementary strand. The term "pyrosequencing" is not limited to the traditional pyrosequencing methods such as those described in Nyren, P. (2007). "The History of Pyrosequencing". *Methods Mol Biology* 373: 1-14, and also includes modified pyrosequencing methods.

"Contacting" and its variants, when used in reference to any set of components, includes any process whereby the components to be contacted are mixed into same mixture (for example, are added into the same compartment or solution), and does not necessarily require actual physical contact between the recited components. The recited components can be contacted in any order or any combination (or sub-combination), and can include situations where one or some of the recited components are subsequently removed from the mixture, optionally prior to addition of other recited components. For example, "contacting A with B and C" includes any and all of the

following situations: (i) A is mixed with C, then B is added to the mixture; (ii) A and B are mixed into a mixture; B is removed from the mixture, and then C is added to the mixture; and (iii) A is added to a mixture of B and C. "Contacting" a target polynucleotide with one or more reaction components, such as a primer or a polymerase, includes any or all of the following situations: (i) the target polynucleotide is contacted with a first component of a reaction mixture to create a mixture; then other components of the reaction mixture are added in any order or combination to the mixture; and (ii) the reaction mixture is fully formed prior to mixture with the target polynucleotide.

"Primer" refers to a polynucleotide capable of hybridizing to a template polynucleotide and acting as the initiation point for incorporating extension nucleotides according to the sequence of the template polynucleotide for synthesis of a complementary polynucleotide.

The term "treatment", "treat" and "therapy", and synonyms thereof as used herein refer to both therapeutic treatment and prophylactic or preventative measures, wherein the object is to prevent or slow down (lessen) a medical condition, which includes but is not limited to diseases, symptoms and disorders. A medical condition also includes a body's response to a disease or disorder, e.g., inflammation. Those in need of such treatment include those already with a medical condition as well as those prone to getting the medical condition or those in whom a medical condition is to be prevented.

The term "subject" as used herein includes patients and non-patients. The term "patient" refers to individuals suffering or are likely to suffer from a medical condition such as cancer, while "non-patients" refer to individuals not suffering and are likely to not suffer from the medical condition. "Non-patients" include healthy individuals, non-diseased individuals and/or an individual free from the medical condition. The term "subject" includes humans and animals. Animals include murine and the like. "Murine" refers to any mammal from the family Muridae, such as mouse, rat, and the like.

The term "micro" as used herein is to be interpreted broadly to include dimensions from about 1 micron to about 1000 microns.

The term "nano" as used herein is to be interpreted broadly to include dimensions less than about 1000 nm.

The term "particle" as used herein broadly refers to a discrete entity or a discrete body. The particle described herein can include an organic, an inorganic or a biological

particle. The particle used described herein may also be a macro-particle that is formed by an aggregate of a plurality of sub-particles or a fragment of a small object. The particle of the present disclosure may be spherical, substantially spherical, or non-spherical, such as irregularly shaped particles or ellipsoidally shaped particles. The term "size" when used to refer to the particle broadly refers to the largest dimension of the particle. For example, when the particle is substantially spherical, the term "size" can refer to the diameter of the particle; or when the particle is substantially non-spherical, the term "size" can refer to the largest length of the particle.

The terms "coupled" or "connected" as used in this description are intended to cover both directly connected or connected through one or more intermediate means, unless otherwise stated.

The term "associated with", used herein when referring to two elements refers to a broad relationship between the two elements. The relationship includes, but is not limited to a physical, a chemical or a biological relationship. For example, when element A is associated with element B, elements A and B may be directly or indirectly attached to each other or element A may contain element B or vice versa.

The term "adjacent" used herein when referring to two elements refers to one element being in close proximity to another element and may be but is not limited to the elements contacting each other or may further include the elements being separated by one or more further elements disposed therebetween.

The term "and/or", e.g., "X and/or Y" is understood to mean either "X and Y" or "X or Y" and should be taken to provide explicit support for both meanings or for either meaning.

Further, in the description herein, the word "substantially" whenever used is understood to include, but not restricted to, "entirely" or "completely" and the like. In addition, terms such as "comprising", "comprise", and the like whenever used, are intended to be non-restricting descriptive language in that they broadly include elements/components recited after such terms, in addition to other components not explicitly recited. For example, when "comprising" is used, reference to a "one" feature is also intended to be a reference to "at least one" of that feature. Terms such as "consisting", "consist", and the like, may in the appropriate context, be considered as a subset of terms such as "comprising", "comprise", and the like. Therefore, in embodiments disclosed herein using the terms such as "comprising", "comprise", and the like, it will be appreciated that these embodiments provide teaching for

corresponding embodiments using terms such as “consisting”, “consist”, and the like. Further, terms such as “about”, “approximately” and the like whenever used, typically means a reasonable variation, for example a variation of +/- 5% of the disclosed value, or a variance of 4% of the disclosed value, or a variance of 3% of the disclosed value, a variance of 2% of the disclosed value or a variance of 1% of the disclosed value.

Furthermore, in the description herein, certain values may be disclosed in a range. The values showing the end points of a range are intended to illustrate a preferred range. Whenever a range has been described, it is intended that the range covers and teaches all possible sub-ranges as well as individual numerical values within that range. That is, the end points of a range should not be interpreted as inflexible limitations. For example, a description of a range of 1% to 5% is intended to have specifically disclosed sub-ranges 1% to 2%, 1% to 3%, 1% to 4%, 2% to 3% etc., as well as individually, values within that range such as 1%, 2%, 3%, 4% and 5%. It is to be appreciated that the individual numerical values within the range also include integers, fractions and decimals. Furthermore, whenever a range has been described, it is also intended that the range covers and teaches values of up to 2 additional decimal places or significant figures (where appropriate) from the shown numerical end points. For example, a description of a range of 1% to 5% is intended to have specifically disclosed the ranges 1.00% to 5.00% and also 1.0% to 5.0% and all their intermediate values (such as 1.01%, 1.02% ... 4.98%, 4.99%, 5.00% and 1.1%, 1.2% ... 4.8%, 4.9%, 5.0% etc.,) spanning the ranges. The intention of the above specific disclosure is applicable to any depth/breadth of a range.

Additionally, when describing some embodiments, the disclosure may have disclosed a method and/or process as a particular sequence of steps. However, unless otherwise required, it will be appreciated that the method or process should not be limited to the particular sequence of steps disclosed. Other sequences of steps may be possible. The particular order of the steps disclosed herein should not be construed as undue limitations. Unless otherwise required, a method and/or process disclosed herein should not be limited to the steps being carried out in the order written. The sequence of steps may be varied and still remain within the scope of the disclosure.

Furthermore, it will be appreciated that while the present disclosure provides embodiments having one or more of the features/characteristics discussed herein, one or more of these features/characteristics may also be disclaimed in other alternative

embodiments and the present disclosure provides support for such disclaimers and these associated alternative embodiments.

DESCRIPTION OF EMBODIMENTS

Exemplary, non-limiting embodiments of a method for sequencing a sample comprising a repeat region, such as a homopolymeric region, and related primers and kits are disclosed hereinafter.

In various embodiments, there is provided a sequencing method. Particularly, the sequencing method is capable of sequencing a sample comprising a repeat region. In various embodiments, the method comprises contacting the sample with one or more primers comprising a repeat region. In various embodiments, the method comprises contacting the sample with a first primer comprising a repeat region of a first length, and optionally, further contacting the sample or clones/copies thereof with a second or further primer(s) comprising a repeat region of a second or further length(s) that is different from the first length. Advantageously, embodiments of the method are capable of accurately determining the length/sequence of a repeat region in a sample. Embodiments of the method are also capable of producing a quantitative readout.

The repeat region may be composed of tandem repeats, optionally short tandem repeats. The repeat region may be composed of repeat units of single nucleotides or multi nucleotides (e.g., dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, pentanucleotide repeats, hexanucleotide repeats etc.). For example, the repeat region may be a homopolymeric region composed of single nucleotide repeats e.g., a repeat region having the sequence AAAA, or it may be a repeat region composed of multi-nucleotide repeats e.g., a repeat region having the sequence GAAAAGAAAA, with GAAAA being the repeat unit. In one embodiment, the repeat unit comprises a single nucleotide repeat unit selected from the group consisting of: A, T, C and G. For example, where the single nucleotide repeat unit is A, the repeat region may be AA, AAA or AAAA etc. For example, where the single nucleotide repeat unit is T, the repeat region may be TT, TTT or TTTT etc. For example, where the single nucleotide repeat unit is C, the repeat region may be CC,

CCC or CCCC etc. For example, where the single nucleotide repeat unit is G, the repeat region may be GG, GGG or GGGG etc.

In various embodiments, the repeat region comprises at least about two, at least about three, at least about four, at least about five, at least about six, at least about seven, at least about eight, at least about nine, at least about ten, at least about 11, at least about 12, at least about 13, at least about 14, at least about 15, at least about 16, at least about 17, at least about 18, at least about 19, at least about 20, at least about 21, at least about 22, at least about 23, at least about 24, at least about 25, at least about 26, at least about 27, at least about 28, at least about 29 or at least about 30 repeat units. In some embodiments, the repeat region comprise at least about 10, at least about 20, at least about 30, at least about 40, at least about 50, at least about 60, at least about 70, at least about 80, at least about 90, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190, at least about 200, at least about 210, at least about 220, at least about 230, at least about 240, at least about 250, at least about 260, at least about 270, at least about 280, at least about 290 or at least about 300 repeat units.

In various embodiments, the length of the repeat region or the number of single nucleotide repeats in the repeat region is at least about two, at least about three, at least about four, at least about five, at least about six, at least about seven, at least about eight, at least about nine, at least about ten, at least about 11, at least about 12, at least about 13, at least about 14, at least about 15, at least about 16, at least about 17, at least about 18, at least about 19, at least about 20, at least about 21, at least about 22, at least about 23, at least about 24, at least about 25, at least about 26, at least about 27, at least about 28, at least about 29 or at least about 30 bases/nucleotides. In some embodiments, the length of the repeat region or the number of single nucleotide repeats in the repeat region is at least about 10 bases/nucleotides, at least about 20, at least about 30, at least about 40, at least about 50, at least about 60, at least about 70, at least about 80, at least about 90, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190, at least about 200, at least about 210, at least about 220, at least about 230, at least about 240, at least about 250, at least about 260, at least about 270, at least about 280, at least about 290 or at least about 300 bases/nucleotides.

In various embodiments, the repeat region of the first primer comprises the same repeat unit as the repeat region of the second primer or further primer. In other words, the repeat unit may be the same in the repeat region of the first primer and in the repeat region of the second primer or further primer. For example, if the repeat region in the first primer is GAAAAGAAAAGAAAAA, with GAAAA being the repeat unit, the repeat region of the second or further primer may be GAAAAGAAAA, GAAAAGAAAAGAAAAAGAAAA, GAAAAGAAAAGAAAAAGAAAAGAAAA etc. i.e., having the same repeat unit that is GAAAA. For example, if the repeat region of the first primer is AAAA, with A being the repeat unit, the repeat region of the second or further primer may be AAA, AAAAA, AAAAAA etc. i.e., having the same repeat unit that is A.

In some embodiments, the repeat region of the primer comprises the same repeat unit as the repeat region in the sample, a reference sequence of the sample or copies/clones thereof. For example, where the repeat region the sample, a reference sequence of the sample or copies/clones thereof comprises AAA, the repeat region of the primer may comprise AAA. In some embodiments, the repeat region of the primer comprises a repeat unit that is complementary to the repeat unit in the repeat region in the sample, a reference sequence of the sample or copies/clones thereof. For example, where the repeat region in the sample, a reference sequence of the sample or copies/clones thereof comprises TTT, the repeat region of the primer may comprise AAA.

In various embodiments, the length of the repeat region or the number of repeat units in the primer substantially corresponds to a predicted/estimated or a verified/validated length of a repeat region or number of repeat units in the sample. In various embodiments, the length of the repeat region or the number of repeat units in the primer substantially corresponds to a predicted/estimated or a verified/validated length of a repeat region or number of repeat units in a reference sequence of the sample. The reference sequence may be derived/obtained from a sample of a healthy individual/population, a sample of a diseased patient/population or it may be derived/obtained from a database. For example, the reference sequence derived/obtained from a database may be a wild-type sequence. In some embodiments, the sample is derived/obtained from a diseased patient, and the corresponding reference sequence is derived/obtained from a sample of a healthy individual/population. In some embodiments, the sample is derived/obtained from a

diseased cell/tissue of a patient, and the corresponding reference sequence is derived/obtained from a paired healthy cell/tissue of the patient. In some embodiments, the sample and the reference sequence comprise the same target region or region of interest e.g., a gene. In some embodiments, where a sequence (such as a sequence of a repeat region) of the reference sequence or portion thereof is not known/confirmed, the method comprises a step of sequencing the reference sequence or portion thereof e.g., via Sanger sequencing, next generation sequencing or other similar methods that has substantially high accuracy in sequencing a repeated region.

The primer may further comprise a sequence that is complementary to parts of the reference sequence ascending and/or descending the repeat region. In some embodiments, the primer further comprises a sequence that is complementary to a sequence attached to the sample, optionally attached to the 3' end of the sample. In some embodiments therefore, the method further comprises attaching a sequence to the sample, optionally attaching a sequence to the 3' end of the sample. In some embodiments, the method further comprises cloning the sample into a vector to obtain template polynucleotides. In various embodiments, the vector comprises any carrier nucleic acid molecule/polypeptide/oligonucleotide into which a polynucleotide sequence can be inserted. In various embodiments, the vector comprises a sequencing vector.

The reference sequence or a repeat region of the reference sequence may have one or more of the following features: (i) comprises at least about two, at least about three, at least about four, at least about five, at least about six, at least about seven, at least about eight, at least about nine, at least about ten, at least about 11, at least about 12, at least about 13, at least about 14, at least about 15, at least about 16, at least about 17, at least about 18, at least about 19, at least about 20, at least about 21, at least about 22, at least about 23, at least about 24, at least about 25, at least about 26, at least about 27, at least about 28, at least about 29 or at least about 30 base pairs; (ii) comprises at least about two, at least about three, at least about four, at least about five, at least about six, at least about seven, at least about eight, at least about nine, at least about ten, at least about 11, at least about 12, at least about 13, at least about 14, at least about 15, at least about 16, at least about 17, at least about 18, at least about 19, at least about 20, at least about 21, at least about 22, at least about 23, at least about 24, at least about 25, at least about 26, at least

about 27, at least about 28, at least about 29 or at least about 30 repeat units, optionally wherein the repeat unit comprises a single nucleotide repeat unit, further optionally wherein the single nucleotide repeat unit is selected from the group consisting of: A, T, C and G; and (iii) comprises tandem repeats, optionally short tandem repeats.

In some embodiments, where a reference sequence is not present, the method may comprise repeated dispensation/cycling of different nucleotides, e.g., repeated dispensation/cycling of the four nucleotides TCGA in order, and determining which nucleotide appears in the read out. Embodiments of the method may therefore be used for de novo sequencing.

In various embodiments, the method comprises contacting the sample with at least about two, at least about three, at least about four, at least about five, at least about six, at least about seven, at least about eight, at least about nine or at least about ten primers comprising a repeat region of different lengths or comprising a different number of repeat units in their repeat regions. A greater number of primers, e.g., more than ten primers, may be used to determine long repeats. Depending on the number of repeated nucleotides or the length of the repeat region in the sample, and whether a reference sequence is available, different number of primers may be used. In some embodiments, it may be sufficient to contact the sample with one primer, for example, if prior knowledge/information of the expected sequence of the sample/target region is known from a reference sequence. In some embodiments, multiple primers of different lengths or multiple graduated census primers may be used, for example, for de novo sequencing to determine the number of repeats present in a sample.

In various embodiments, there is provided a sequencing method comprising contacting a sample and/or copies/clones thereof with a plurality of primers comprising a repeat region of different lengths or a plurality of graduated census primers.

In various embodiments, the plurality of primers having repeat regions of different lengths are added in succession or added separately i.e., not as a cocktail mix. In some examples, a first primer comprising a repeat region of a first length is added to the sample, followed by a second primer comprising a repeat region of a second length, followed by a third primer comprising a repeat region of a third length etc. Additional steps, such as a step of washing away the added primer or a step of determining whether primers are annealed/hybridized to the sample and/or whether

the primers are elongated, may be performed following the addition of a primer and before the addition of the next primer. In some examples, a first primer comprising a repeat region of a first length is added to the sample in a first partition/tube/well/containment body, a second primer comprising a repeat region of a second length is added to a copy/clone of the sample in a second partition/tube/well/containment body, a third primer comprising a repeat region of a third length is added to a copy/clone of the sample in a third partition/tube/well/containment body. In some embodiments therefore, in one partition/tube/well/containment body comprising a sample, there is only one primer type having a repeat region of a single length; there are no two or more primer types each having a repeat region of different lengths.

In various embodiments, the second primer has a greater number of repeat units in its repeat region than the first primer. For example, the first primer may have n repeat units, and the second primer may have $n+m$ repeat units, wherein n and m are both positive integers. In some embodiments, n is a positive integer of from 2 to 30. In some embodiments, m is a positive integer of from 1 to 9. In some embodiments, the first primer has n repeat units in its repeat region, the second primer has $n+1$ repeat units in its repeat region, and optionally the further primer may comprise a third primer having $n+2$ repeat units in its repeat region, a fourth primer having $n+3$ repeat units in its repeat region, a fifth primer having $n+4$ repeat units in its repeat region, a sixth primer having $n+5$ repeat units in its repeat region, a seventh primer having $n+6$ repeat units in its repeat region, an eighth primer having $n+7$ repeat units in its repeat region, a ninth primer having $n+8$ repeat units in its repeat region, a tenth primer having $n+9$ repeat units in its repeat region etc., wherein n is an integer of 2 to 30. In one embodiment, the method comprises contacting the sample and/or an identical clone/copy thereto with at least six primers having a different number of repeat units in their repeat region, the different length of repeat units being selected from the group consisting of: n repeat units, $n+1$ repeat units, $n+1$ repeat units, $n+3$ repeat units, $n+4$ repeat units and $n+5$ repeat units.

In various embodiments, the repeat region of the second length is longer than the repeat region of the first length. For example, the first length may be n bases/nucleotides, and the second length may be $n+m$ bases/nucleotides, wherein n and m are both positive integers. In some embodiments, n is a positive integer of from 2 to 30. In some embodiments, m is a positive integer of from 1 to 9. In some

embodiments, the first length is n bases/nucleotides, the second length is $n+1$ bases/nucleotides, and optionally the further length comprises a third length of $n+2$ bases/nucleotides, a fourth length of $n+3$ bases/nucleotides, a fifth length of $n+4$ bases/nucleotides, a sixth length of $n+5$ bases/nucleotides, a seventh length of $n+6$ bases/nucleotides, an eighth length of $n+7$ bases/nucleotides, a ninth length of $n+8$ bases/nucleotides, a tenth length of $n+9$ bases/nucleotides etc., wherein n is an integer of 2 to 30. In one embodiment, the method comprises contacting the sample and/or an identical clone/copy thereto with at least six primers having a repeat region of different lengths, the different lengths being selected from the group consisting of: n bases/nucleotides, $n+1$ bases/nucleotides, $n+1$ bases/nucleotides, $n+3$ bases/nucleotides, $n+4$ bases/nucleotides and $n+5$ bases/nucleotides.

In various embodiments, the repeat region is located at a terminal end (e.g., 3' end or 5' end) of the primer(s). In various embodiments, the repeat region is at a 3' end of the primer(s).

The sample may comprise a nucleic acid or a polynucleotide. Examples of a nucleic acid or a polynucleotide include deoxyribonucleic acid (DNA) (including cDNA) and/or ribonucleic acid (RNA). In one embodiment, the sample comprises DNA. In one embodiment, the sample comprises a single-stranded polynucleotide. In one embodiment, the sample comprises a single-stranded DNA.

In various embodiments therefore, there is provided a method of sequencing a polynucleotide sample comprising a homopolymeric region R_s , the method comprising: providing template polynucleotides comprising the polynucleotide sample and copies thereof; providing a plurality of R_p -containing primers comprising, at each of their 3' end, a homopolymeric region R_p comprising nucleotides that are complementary to the nucleotides in the homopolymeric region R_s , and contacting the plurality of R_p -containing primers with the template polynucleotides under conditions suitable for hybridization and synthesis of complementary polynucleotides, wherein the homopolymeric regions R_p in the plurality of R_p -containing primers are of different lengths. The polynucleotide sample may be cloned into a vector to obtain the template polynucleotides.

In various embodiments, the homopolymeric regions R_p in the plurality of R_p -containing primers comprise a length selected from: X nucleotides (nt), $X+Y$ nt, $X+2Y$ nt, $X+3Y$ nt, $X+4Y$ nt, $X+5Y$ nt, $X+6Y$ nt, $X+7Y$ nt, $X+8Y$ nt, $X+9Y$ nt, $X+10Y$ nt and combinations thereof, wherein X and Y are both positive integers. In various

embodiments, X is at least about 2. In various embodiments, X is a positive integer of from about 2 to about 300, from about 2 to about 200, from about 2 to about 100, from about 2 to about 50 or from about 2 to about 30. In various embodiments, Y is at least about 1. In various embodiments, Y is from about 1 to about 50, from about 1 to about 40, from about 1 to about 30, from about 1 to about 20, from about 1 to about 10, from about 1 to about 9, from about 1 to about 8, from about 1 to about 7, from about 1 to about 6, from about 1 to about 5, from about 1 to about 4, from about 1 to about 3 or from about 1 to about 2.

In various embodiments, X is more than the length of a homopolymeric region that can be accurately determined by a conventional pyrosequencing method. In some examples, conventional pyrosequencing is deficient in that it may get saturated even at 3 or 4 repeats. In various embodiments therefore, X is more than about 2, more than about 3, more than about 4, more than about 5, more than about 6, more than about 7, more than about 8, more than about 9, more than about 10, more than about 11, more than about 12, more than about 13, more than about 14, more than about 15, more than about 16, more than about 17, more than about 18, more than about 19 or more than about 20. In various embodiments, X is at least about 2, at least about 3, at least about 4, at least about 5, at least about 6, at least about 7, at least about 8, at least about 9, at least about 10, at least about 11, at least about 12, at least about 13, at least about 14, at least about 15, at least about 16, at least about 17, at least about 18, at least about 19 or at least about 20. In various embodiments, X is no less than about 2, no less than about 3, no less than about 4, no less than about 5, no less than about 6, no less than about 7, no less than about 8, no less than about 9, no less than about 10, no less than about 11, no less than about 12, no less than about 13, no less than about 14, no less than about 15, no less than about 16, no less than about 17, no less than about 18, no less than about 19 or no less than about 20. In one embodiment, the homopolymeric regions R_p in the plurality of R_p -containing primers comprise a length selected from: X nucleotides (nt), X+1 nt, X+2 nt, X+3 nt, X+4 nt, X+5 nt, X+6 nt, X+7 nt, X+8 nt, X+9 nt, X+10 nt and combinations thereof, wherein X is an integer no less than 3.

In various embodiments, the plurality of R_p -containing primers comprise at least at least about two, at least about three, at least about four, at least about five, at least about six, at least about seven, at least about eight, at least about nine, at least about ten, at least about 11, at least about 12, at least about 13, at least about 14, at least

about 15, at least about 16, at least about 17, at least about 18, at least about 19 or at least about 20 R_p -containing primers. In one embodiment, the plurality of R_p -containing primers comprise at least three R_p -containing primers.

Each of the plurality of R_p -containing primers having homopolymeric regions R_p of different lengths may be contacted with each template polynucleotide, or the plurality of R_p -containing primers may be contacted separately or sequentially with a template polynucleotide. In some embodiments therefore, contacting the plurality of R_p -containing primers with the template polynucleotides comprises partitioning the template polynucleotides into a plurality of partitions of template polynucleotides and contacting each partition with a R_p -containing primer with a homopolymeric region R_p of a different length. In some embodiments, contacting the plurality of R_p -containing primers with the template polynucleotides comprises applying the plurality of R_p -containing primers sequentially to the template polynucleotides. For example, a first R_p -containing primer having a homopolymeric region R_p of a first length may be added to the template polynucleotides, followed by a second R_p -containing primer having a homopolymeric region R_p of a second length, followed by a third R_p -containing primer having a homopolymeric region R_p of a third length etc. After each R_p -containing primer is added to the template polynucleotides, it may be removed or washed away before the next R_p -containing primer is added to the template polynucleotides.

When the plurality of R_p -containing primers are contacted with the template polynucleotides, a primer which homopolymeric region R_p is shorter than or equal in length to the homopolymeric region R_s of the sample/template polynucleotides may hybridize/anneal to a template polynucleotide, while a primer which homopolymeric region R_p is longer than the homopolymeric region R_s of the sample/template polynucleotides may not be capable of hybridizing/annealing to a template polynucleotide. Thus, the length of the homopolymeric region R_s of the sample may be determined by determining whether the primers are hybridized to the template nucleotides and/or whether the primers are elongated or extended by the addition of one or more nucleotides. It will be appreciated that, following the hybridization/annealing of a primer to a template polynucleotide (if hybridization/annealing of the primer to the template polynucleotide takes place), a nucleotide may only be incorporated into the primer if it is complementary to the next corresponding nucleotide in the template polynucleotide/sample. For example, a primer which homopolymeric region R_p is equal in length to the homopolymeric region

R_s of the sample is expected to hybridize/anneal to a template polynucleotide, and the addition of a nucleotide corresponding to the nucleotides present in the homopolymeric region R_p is not expected to be incorporated, while the addition of a different nucleotide may be incorporated. A primer which homopolymeric region R_p is shorter than the homopolymeric region R_s of the sample (e.g., by one nucleotide) is expected to hybridize/anneal to a template polynucleotide, and the addition of a nucleotide corresponding to the nucleotides present in the homopolymeric region R_p is expected to be incorporated, with the intensity of the signal generated from the incorporation being proportional to the number/amount of nucleotides added. The addition of a different nucleotide may not be incorporated until the homopolymeric region in the growing complementary strand (from the incorporation of nucleotide(s) corresponding to the nucleotides present in the homopolymeric region R_p) equals the length of the homopolymeric region R_s of the sample. By counting the number/amount of nucleotides added from the intensity of the signal generated, the length of the homopolymeric region R_s of the sample may be determined. A primer which homopolymeric region R_p is longer than the homopolymeric region R_s of the sample (e.g., by one nucleotide) is not expected to hybridize/anneal to a template polynucleotide, and the addition of a nucleotide corresponding to the nucleotides present in the homopolymeric region R_p is not expected to be incorporated. The addition of a different nucleotide is also not expected to be incorporated. The observation made from the contact with single primer, or collective observations from contact with a plurality of primers may therefore provide information on the length of the homopolymeric region R_s in the sample. In various embodiments, the method comprises tabulating the counts of incorporated nucleotides(s) for each template polynucleotide for each primer comprising a homopolymeric region R_p of a different length.

In various embodiments therefore, the method further comprises determining whether the primer(s) is hybridized to the sample and/or whether the primer(s) is elongated. The determining step may comprise adding one or more nucleotides to the sample and determining whether the one or more nucleotides is incorporated. Methods for determining whether a nucleotide is incorporated in a growing polynucleotide chain are known in the art. For example, inorganic pyrophosphate is released when a nucleotide is incorporated in a growing polynucleotide chain by a polymerase. This release of pyrophosphate may be converted into a

light/fluorescence/chemiluminescent signal, for example, by converting pyrophosphate into adenosine triphosphate (ATP) by the use of ATP sulfurylase, and then allowing the ATP to react with luciferin in the presence of luciferase to emit chemiluminescence. Hence, the incorporation of a nucleotide may be detected by detecting a signal such as a light/fluorescence/chemiluminescent signal. In some embodiments therefore, determining whether the one or more nucleotides is incorporated comprises detecting for the presence or absence of a light/fluorescence/chemiluminescent signal. In various embodiments, a nucleotide is incorporated when a light/fluorescence/chemiluminescent signal is detected and the nucleotide is not incorporated when a light/fluorescence/chemiluminescent signal is not detected. In various embodiments, the amount/intensity of the light/fluorescence/chemiluminescent signal is proportional, optionally directly proportional to the amount/number of nucleotides incorporated. For example, if 2As are incorporated, the intensity of the light/fluorescence/chemiluminescent emitted is two times more than the intensity of light/fluorescence/chemiluminescent emitted when 1A is incorporated. Without being bound by theory, it is believed that, in the synthesis of a complementary polynucleotide chain, each incorporation event is accompanied by the release of pyrophosphate (or PPi) in a quantity equimolar to the amount/number of incorporated nucleotides. ATP sulfurylase may be used to convert the PPi to ATP in the presence of adenosine 5' phosphosulfate (APS), and this ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin which generates light in amounts that are proportional to the amount of ATP. In various embodiments therefore, the method comprises detecting whether a light/fluorescence/chemiluminescent signal is generated/emitted/produced upon addition of a nucleotide to the template polynucleotide(s).

Detecting whether a light/fluorescence/chemiluminescent signal is generated/emitted/produced upon addition of a nucleotide may comprise use of a detector for detection, optionally wherein the detector comprises a photodiode, a photomultiplier partition, or a charge-coupled device (CCD) camera. Determining whether a nucleotide is incorporated may further comprise determining whether a peak is generated in a sequencing readout such as in a pyrosequencing flow-gram. In various embodiments, a nucleotide is incorporated when a peak is generated and the nucleotide is not incorporated when a peak is not generated. In various embodiments,

the height of the peak is proportional, optionally directly proportional to the amount/number of nucleotides incorporated.

The one or more nucleotides may comprise (i) a nucleotide corresponding to a nucleotide in the repeat region of the primer or the sample/reference sequence; (ii) a nucleotide immediately adjacent to, ascending or descending (e.g. a 3' nucleotide or a 5' nucleotide immediately neighboring) the repeat region of the sample/reference sequence; (iii) a nucleotide immediately adjacent to, ascending or descending (e.g. a 3' nucleotide or a 5' nucleotide immediately neighboring) the nucleotide in (ii) in the sample reference sequence; and/or (iv) further nucleotide(s) further adjacent to, ascending or descending the nucleotide in (ii) in the sample/reference sequence.

The one or more nucleotides may be deoxyribonucleoside triphosphate (dNTP), although not limited as such. In various embodiments, the dNTP is selected from the group consisting of: deoxythymidine triphosphate (dTTP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP) and 2'-deoxyadenosine-5'-(α -thio)-triphosphate (dATP α S). In various embodiments, the dNTP does not comprise dATP. dATP may be recognised by luciferase and result in false signals of pyrophosphate detection in pyrosequencing.

In various embodiments, the method comprises adding a nucleotide N₁ to the template polynucleotides after the template polynucleotides are contacted with a R_p-containing primer, the nucleotide N₁ corresponding to the nucleotides in the homopolymeric region R_p; and determining whether the nucleotide N₁ is incorporated in the complementary polynucleotides. In various embodiments, the method comprises adding a nucleotide N₂ to the template polynucleotides after the template polynucleotides are contacted with a R_p-containing primer, wherein nucleotide N₂ is different from the nucleotides in the homopolymeric region R_p; and determining whether the nucleotide N₂ is incorporated in the complementary polynucleotides. In various embodiments, nucleotide N₂ corresponds to a nucleotide that is complementary to the nucleotide immediately adjacent to (e.g., immediately ascending/descending) the homopolymeric region R_s in the polynucleotide sample and/or immediately adjacent to (e.g., immediately ascending/descending) a homopolymeric region (e.g., the corresponding homopolymeric region) in a reference sequence of the polynucleotide sample. In various embodiments, nucleotide N₂ corresponds to a nucleotide that is complementary to an adjacent nucleotide that is directly next to the homopolymeric region R_s on a 3' side in the polynucleotide sample

and/or corresponds to a nucleotide that is complementary to an adjacent nucleotide that is directly next to the homopolymeric region (e.g., the corresponding homopolymeric region) on a 3' side in a reference sequence of the polynucleotide sample. In various embodiments, nucleotide N_2 corresponds to a nucleotide that is complementary to an adjacent nucleotide that is directly next to the homopolymeric region R_s on a 5' side in the polynucleotide sample and/or corresponds to a nucleotide that is complementary to an adjacent nucleotide that is directly next to the homopolymeric region (e.g., the corresponding homopolymeric region) on a 5' side in a reference sequence of the polynucleotide sample.

In various embodiments, the method further comprises identifying the R_p -containing primer contacted with the template polynucleotides when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides; and concluding that the homopolymeric region R_s has a length corresponding to the length of the homopolymeric region R_p of the identified R_p -containing primer. In various embodiments, the method further comprises identifying the R_p -containing primer contacted with the template polynucleotides when nucleotide N_1 is incorporated in the complementary polynucleotides; determining the number of nucleotide N_1 incorporated; and concluding that the homopolymeric region R_s has a length corresponding to the sum of the number of nucleotide N_1 incorporated and the length of the homopolymeric region R_p of the identified R_p -containing primer. In various embodiments, the signal emitted/produced/generated from the incorporation is not a saturated signal. In various embodiments, the number of nucleotide N_1 incorporated/ determined to be incorporated is no more than the length of a homopolymeric region that can be accurately determined by a pyrosequencing method e.g., a conventional pyrosequencing method. In various embodiments, the number of nucleotide N_1 incorporated/ determined to be incorporated is no more than about five, no more than about four, no more than about three, no more than about two or no more than about one. In various embodiments, the number of nucleotide N_1 incorporated/ determined to be incorporated is from about one to about five, from about one to about four or from one to about three. In various embodiments, when the signal emitted/produced/generated from the incorporation is a saturated signal, the method comprises contacting the template polynucleotides with a R_p -containing primer comprising a longer homopolymeric region R_p , and where N_1 is incorporated in the complementary polynucleotides and the signal emitted/produced/generated from

the incorporation is not a saturated signal, then concluding that the homopolymeric region R_s has a length corresponding to the sum of the number of nucleotide N_1 incorporated and the length of the homopolymeric region R_p of the R_p -containing primer comprising the longer homopolymeric region R_p . The method may also further comprise a step of verifying that nucleotide N_1 is incorporated in the complementary polynucleotides of template polynucleotides contacted with a R_p -containing primer comprising a shorter homopolymeric region R_p than that of the identified R_p -containing primer before the concluding step. The verifying step may further comprise verifying that a correct number/an expected number of nucleotide N_1 is incorporated in the complementary polynucleotides of template polynucleotides contacted with the R_p -containing primer comprising a shorter homopolymeric region R_p . The correct number/the expected number of nucleotide N_1 incorporated may be obtained by calculating the difference between (i) the length of the homopolymeric R_p region in the identified R_p -containing primer; or the sum of the length of the homopolymeric R_p region in the identified R_p -containing primer and the number of nucleotide N_1 incorporated after contact with the identified R_p -containing primer (as the case may be) and (ii) the length of the shorter homopolymeric region R_p . In various embodiments therefore, the correct number/an expected number of nucleotide N_1 incorporated corresponds to the difference between (i) the length of the homopolymeric R_p region in the identified R_p -containing primer; or the sum of the length of the homopolymeric R_p region in the identified R_p -containing primer and the number of nucleotide N_1 incorporated after contact with the identified R_p -containing primer (as the case may be) and (ii) the length of the shorter homopolymeric region R_p .

Wherein where the nucleotide N_1 is not incorporated in the complementary polynucleotides of the template polynucleotides contacted with the plurality of R_p -containing primers, the method may further comprise: a) providing a further R_p -containing primer comprising a shorter homopolymeric region R_p than those of the plurality of R_p -containing primers; b) contacting the template polynucleotides with the further R_p -containing primer; c) determining whether the nucleotide N_1 and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of the template polynucleotides of step b); and d) concluding that the homopolymeric region R_s has a length corresponding to the length of the homopolymeric region R_p of the further R_p -containing primer when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of step c).

Wherein the nucleotide N_1 is incorporated in the complementary polynucleotides of the template polynucleotides contacted with the plurality of R_p -containing primers, the method may further comprise: a) providing a further R_p -containing primer comprising a longer homopolymeric region R_p than those of the plurality of R_p -containing primers; b) contacting the template polynucleotides with the further R_p -containing primer; c) determining whether the nucleotide N_1 and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of the template polynucleotides of step b); and d) concluding that the homopolymeric region R_s has a length corresponding to the length of the homopolymeric region R_p of the further R_p -containing primer when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of step c).

The method may further comprise verifying that nucleotide N_1 is incorporated in the complementary polynucleotides of the template polynucleotides contacted with a R_p -containing primer comprising a shorter homopolymeric region R_p than that of the further R_p -containing primer before the concluding step. The verifying step may further comprise verifying that a correct number/an expected number of nucleotide N_1 is incorporated in the complementary polynucleotides of template polynucleotides contacted with the R_p -containing primer comprising a shorter homopolymeric region R_p . The correct number/the expected number of nucleotide N_1 incorporated may be obtained by computing/calculating the difference between (i) the length of the homopolymeric R_p region in the further R_p -containing primer and (ii) the length of the shorter homopolymeric region R_p . In various embodiments therefore, the correct number/an expected number of nucleotide N_1 incorporated corresponds to the difference between the length of the homopolymeric R_p region in the further R_p -containing primer and the length of the shorter homopolymeric region R_p .

In one example, the method comprises: (a) contacting the sample/template polynucleotides with a first primer comprising a repeat region; (b) adding a first nucleotide to the sample/template polynucleotides, the first nucleotide corresponding to a nucleotide in the repeat region of the first primer or the reference sequence; (c) determining whether the first nucleotide is incorporated; (d) optionally removing remnants of the first nucleotide; (e) adding a second nucleotide to the sample/template polynucleotides, the second nucleotide corresponding to a nucleotide immediately ascending the repeat region of the reference sequence; (f) determining whether the second nucleotide is incorporated; wherein if the first nucleotide is not incorporated

and the second nucleotide is incorporated, concluding that the repeat region of the sample/template polynucleotides corresponds to the repeat region of the first primer (i.e. the sample/template polynucleotides contains the same number of repeat units/nucleotides as the first primer) and wherein if the result is otherwise, contacting the sample/template polynucleotides with a second or further primer comprising a repeat region of a different length and repeating steps (b) to (f) until the desired result is obtained (i.e. the first nucleotide is not incorporated and the second nucleotide is incorporated) and concluding that the repeat region of the sample/template polynucleotides corresponds to the repeat region of the primer that does not result in the incorporation of the first nucleotide but results in the incorporation of the second nucleotide.

In one example, if the first nucleotide is incorporated, the sample/template polynucleotides is contacted with a second or further primer comprising a repeat region of a longer length. In one example, when a 9A primer is observed to result in the incorporation of two A nucleotides, the sample/template polynucleotides may be contacted with a 11A primer for further analysis. If the first nucleotide is incorporated, one may also conclude the length of the repeat region based on a readout indicative of the number of nucleotides incorporated. For example, when a 9A primer is observed to result in the incorporation of two A nucleotides, one may conclude that the sample contains an 11A repeat region.

In one example, if the first and second nucleotides are not incorporated, the sample/template polynucleotide is contacted with a second or further primer comprising a repeat region of a shorter length. In one example, when a 12A primer is observed not to result in the incorporation the first and second nucleotides, the sample/template polynucleotides may be contacted with a 11A or 10A primer for further analysis.

In one example, steps (a) to (f) are performed on a single sample/template polynucleotides in a single partition/well/containment body. In some embodiments, steps (a) to (c) and optionally (d) are performed on a sample/template polynucleotide in a partition/well/containment body, while steps (e) and (f) are performed on a sample/template polynucleotide in a different partition/well/containment body.

In some examples, the method may further comprise adding a third nucleotide to the sample/template polynucleotides, the third nucleotide corresponding to a nucleotide further ascending the repeat region of the reference sequence (e.g. a

nucleotide 2 position away and ascending from the repeat region of the reference sequence), determining whether the third nucleotide is incorporated, wherein if the third nucleotide is incorporated, concluding that the repeat region of the sample/template polynucleotides corresponds to the repeat region of the first primer (i.e. the sample/template polynucleotides contains the same number of repeat units/nucleotides as the first primer) and wherein if the result is otherwise, contacting the sample/template polynucleotides with a second or further primer comprising a repeat region of a different length and repeating steps (b) to (f) until the desired result is obtained (i.e. the first nucleotide is not incorporated and the second and third nucleotides are incorporated) and concluding that the repeat region of the sample/template polynucleotides corresponds to the repeat region of the primer that does not result in the incorporation of the first nucleotide but results in the incorporation of the second and third nucleotides.

In one example, the method comprises: (a) contacting the sample/template polynucleotides with a first primer comprising a repeat region; (b) adding a first nucleotide to the sample/template polynucleotides, the first nucleotide corresponding to a nucleotide in the repeat region of the first primer or the reference sequence; (c) determining whether the first nucleotide is incorporated; (d) optionally removing remnants of the first nucleotide; (e) adding a second nucleotide that is different from the first nucleotide; (f) determining whether the second nucleotide is incorporated; wherein if the first nucleotide is not incorporated and the second nucleotide is incorporated, concluding that the repeat region of the sample/template polynucleotides corresponds to the repeat region of the first primer (i.e., the sample/template polynucleotides contains the same number of repeat units/nucleotides as the first primer) and wherein if the result is otherwise, contacting the sample/template polynucleotides with a second or further primer comprising a repeat region of a different length and repeating steps (b) to (f) until the desired result is obtained (i.e. the first nucleotide is not incorporated and the second nucleotide is incorporated) and concluding that the repeat region of the sample/template polynucleotides corresponds to the repeat region of the primer that does not result in the incorporation of the first nucleotide but results in the incorporation of the second nucleotide.

In one example, the method comprises: (a) contacting the sample/template polynucleotides with a first primer comprising a repeat region; (b) adding a first

nucleotide to the sample/template polynucleotides, the first nucleotide corresponding to a nucleotide immediately ascending the repeat region of the reference sequence; (c) determining whether the first nucleotide is incorporated; (d) optionally removing remnants of the first nucleotide; (e) adding a second nucleotide to the sample/template polynucleotides, the second nucleotide corresponding to a nucleotide further ascending the repeat region of the reference sequence; (f) determining whether the second nucleotide is incorporated; wherein if the first nucleotide and the second nucleotide are incorporated, concluding that the repeat region of the sample/template polynucleotides corresponds to the repeat region of the first primer (i.e., the sample/template polynucleotides contains the same number of repeat units/nucleotides as the primer) and wherein if the results are otherwise, contacting the sample/template polynucleotides with a second or further primer comprising a repeat region of a different length and repeating steps (b) to (f) until the results are obtained and concluding that the repeat region of the sample/template polynucleotides corresponds to the repeat region of the primer that results in the incorporation of the first nucleotide and the second nucleotide. In some embodiments, steps (a) to (f) are performed on a single sample/template polynucleotides in a single partition//well/containment body. In some embodiments, steps (a) to (c) and optionally (d) are performed on a sample/template polynucleotides in a partition//well/containment body, while steps (e) and (f) are performed on an identical clone/copy in a different partition//well/containment body.

In various embodiments, the method further comprises determining a length of a repeat region/homopolymeric region in a reference sequence of the polynucleotide sample prior to providing the plurality of R_p -containing primers. The R_p -containing primers may then be designed to have repeat regions/homopolymeric regions that are identical and/or similar in length to the repeat region/homopolymeric region in the reference sequence. Advantageously, obtaining prior knowledge on an approximate length of a repeat region/homopolymeric region in a sample (based on target genome sequences) and designing the primers accordingly to have an identical/similar length of a repeat region/homopolymeric region is likely to speed up the sequencing process of the sample, e.g., as compared to contacting with primers of random lengths. In various embodiments, the plurality of R_p -containing primers comprises: a R_p -containing primer which homopolymeric region R_p has a length corresponding to the length of the homopolymeric region in the reference sequence; optionally, a R_p -

containing primer which homopolymeric region R_p has a length shorter than the length of the homopolymeric region in the reference sequence; and further optionally, a R_p -containing primer which homopolymeric region R_p has a length longer than the length of the homopolymeric region in the reference sequence.

In various embodiments, the method further comprises amplifying the sample or a target region/region of interest in the sample to obtain the template polynucleotides. The amplification reactions may include but are not limited to polymerase chain reaction (PCR), ligase chain reaction (LCR), loop mediated isothermal amplification (LAMP), nucleic acid sequence based amplification (NASBA), self-sustained sequence replication (3SR), rolling circle amplification (RCA) or any other process whereby one or more copies of a particular polynucleotide sequence or nucleic acid sequence may be generated from a polynucleotide template sequence or nucleic acid template sequence. Molecular cloning, e.g., using bacteria, may also be used to amplify/expand the sample. In one example, PCR is performed on a target region of the sample, followed by cloning into vector and subsequent PCR expansion to obtain the template polynucleotides. The PCR products may be subjected to alkali denaturation to obtain single-stranded template polynucleotides.

As may be appreciated, a sample may be heterogenous with mixed populations of cells/polynucleotides having different number of repeating units/nucleotides in their respective repeat regions. By use of a cloning step, different polynucleotides having different number of repeating units may be segregated and separately analysed with the primers of the present disclosure. Thus, embodiments of the method may be useful and applicable for sequencing a heterogenous sample having polynucleotides comprising repeat regions of different lengths. In some embodiments, the sample comprises a heterogenous mix of polynucleotides having different number of repeating units in their respective repeat regions. In some embodiments, the sample is substantially homogenous and comprises polynucleotides having repeat regions of a single length/having the same number of repeat units. In various embodiments, the template polynucleotides comprise polynucleotides having repeat regions of substantially the same length.

In various embodiments, the method further comprises adding or contacting the template polynucleotides with a polynucleotide synthesizing enzyme for synthesis of complementary polynucleotides. The polynucleotide synthesizing enzyme may be a DNA synthesizing enzyme. The polynucleotide synthesizing enzyme may be a

polymerase or DNA polymerase. In various embodiments, the method comprises adding or contacting the template polynucleotides a polymerase for incorporating the nucleotide(s) to the complementary polypeptide. In one example, the DNA polymerase comprises DNA Polymerase I of *Escherichia coli*. In one example, the DNA polymerase comprises the Klenow fragment of *Escherichia coli* DNA Polymerase I. It will be appreciated that other suitable polymerases may also be used.

In various embodiments, the method comprises a sequencing by synthesis method, which takes place by taking a single-stranded polynucleotide to be sequenced (i.e., a sample) and then synthesizing its complementary strand enzymatically. In some embodiments, the sequencing method comprises a pyrosequencing method. In some embodiments, the pyrosequencing method comprises a solid-phase pyrosequencing method. In some embodiments, the pyrosequencing method comprises a liquid-phase pyrosequencing method. Embodiments of the method may be semi-quantitative or a quantitative.

In various embodiments therefore, the method may further comprise adding or contacting the template polynucleotides/primers/mixture thereof with one or more components/reagents in a pyrosequencing method, such as, but not limited to, an enzyme apyrase for decomposing nucleotides/dNTPs which have been added as a substrate and remained unreacted, sulfurylase and APS for converting pyrophosphate into ATP; luciferin and luciferase for reacting with ATP to emit chemiluminescence.

Embodiments of the sequencing method may also be implemented as one or more of the following methods: a method of determining the sequence of a sample or of a target region in the sample, a method of determining the presence of a mutation in a sample, a method of determining a variant (e.g. for somatic or germline mutation), a method of verifying a sequence, a method of treatment, a method of diagnosis, a method of prognosis, a method of stratifying a patient for treatment, a method of selecting a treatment regimen, a method of de novo assembly, a method of sequencing a microorganism, a method of sequencing a virus, and the like. Advantageously, embodiments of the method not only allow for accurate determination of sequences at regions with repeats, but it also has a simple workflow that is user-friendly and that does not require complicated equipment or analysis pipeline. Embodiments of the methods also provide a fast turnaround time. Read-outs may be obtained within a day for quick clinical decision making.

In various embodiments, there is provided a plurality/combination of primers comprising repeat regions, e.g., for use in embodiments of the methods as described herein. In various embodiments, the repeat regions comprise homopolymeric regions. In various embodiments, there is provided a plurality/combination of R_p -containing primers. In various embodiments, there is provided a plurality/combination of graduated census primers. The R_p -containing primers may comprises one or more features as described herein. In various embodiments, the plurality of R_p -containing primers comprises, at each of their 3' end, a homopolymeric regions R_p . The repeat regions/homopolymeric regions R_p in the combination/plurality of primers/ R_p -containing primers may be of different lengths.

In various embodiments, the repeat region/homopolymeric regions R_p in the combination/plurality of primers/ R_p -containing primers comprise a length selected from: X nucleotides (nt), X+Y nt, X+2Y nt, X+3Y nt, X+4Y nt, X+5Y nt, X+6Y nt, X+7Y nt, X+8Y nt, X+9Y nt, X+10Y nt and combinations thereof, wherein X and Y are both positive integers. In various embodiments, X is at least about 2. In various embodiments, X is a positive integer of from about 2 to about 300, from about 2 to about 200, from about 2 to about 100, from about 2 to about 50 or from about 2 to about 30. In various embodiments, Y is at least about 1. In various embodiments, Y is from about 1 to about 50, from about 1 to about 40, from about 1 to about 30, from about 1 to about 20, from about 1 to about 10, from about 1 to about 9, from about 1 to about 8, from about 1 to about 7, from about 1 to about 6, from about 1 to about 5, from about 1 to about 4, from about 1 to about 3 or from about 1 to about 2. In various embodiments, the repeat region/homopolymeric regions R_p in the plurality of R_p -containing primers comprise a length selected from: X nucleotides (nt), X+1 nt, X+2 nt, X+3 nt, X+4 nt, X+5 nt, X+6 nt, X+7 nt, X+8 nt, X+9 nt, X+10 nt and combinations thereof, wherein X is an integer no less than 3.

In various embodiments, the combination/plurality of primers/ R_p -containing primers comprise at least at least about two, at least about three, at least about four, at least about five, at least about six, at least about seven, at least about eight, at least about nine, at least about ten, at least about 11, at least about 12, at least about 13, at least about 14, at least about 15, at least about 16, at least about 17, at least about 18, at least about 19 or at least about 20 primers/ R_p -containing primers. In one embodiment, the combination/plurality of primers/ R_p -containing primers comprise at least three primers/ R_p -containing primers.

In various embodiments, the combination/plurality of primers/ R_p -containing primers comprises a first primer comprising a repeat region/homopolymeric region R_p of a first length, and optionally, a second or further primer(s) comprising a repeat region/homopolymeric region R_p of a second or further length(s) that is different from the first length.

In various embodiments, there is provided a kit e.g., for use in embodiments of the methods as described herein. In various embodiments, the kit comprises embodiments of the combination/plurality of primers/ R_p -containing primers as described herein and at least one more components/reagents selected from the group consisting of: nucleotides, polymerase (e.g., DNA polymerase), adenosine phosphosulfate (APS), ATP sulfurylase, luciferin, luciferase and apyrase. In one embodiment, the nucleotides are selected from the group consisting of: dTTP, dCTP, dGTP and dATP α S and combinations thereof.

In various embodiments, there is provided a system, optionally an automated system, comprising: a primer design module configured to add one or more nucleotides to the 3' end of a first primer comprising a repeat region/homopolymeric region of a first length to obtain a second or further primer(s) comprising a repeat region/homopolymeric region of a second or further length(s) that is different from the first length; and/or a primer dispensation module configured to dispense a first primer comprising a repeat region/homopolymeric region of a first length and optionally a second or further primer(s) comprising a repeat region/homopolymeric region of a second or further length(s) that is different from the first length; and optionally a sample processing module for processing the sample (e.g. amplification of sample, purification of sample etc.); a sequencing module for sequencing the sample, optionally a sequencing module that operates according to a pyrosequencing technique (e.g., PyroMark platform such as PyroMark Q48); and an analysis module for transforming data received from the sequencing module to output a sequence result.

In various embodiments, there is provided an apparatus adapted to implement the embodiments of the system as described herein.

In various embodiments, there is provided a method, a product, a system or an apparatus as described herein.

BRIEF DESCRIPTION OF FIGURES

FIG. 1. shows how the semi-quantitative feature of Sanger sequencing, which is a method that does not comply with one or more of the requirements of embodiments of the methods disclosed herein, could not determine the sequence at regions of variable reads.

FIG. 2 shows the primers with repeat regions of different lengths being used in a sequencing method in accordance with an embodiment disclosed herein. After the primers are added to the template polynucleotides, nucleotides are subsequently dispensed and tested for incorporation by detecting for the presence of a peak in the sequencing readout.

FIG. 3 shows the determination of the immediate ascending A and G for each clone with a 9A primer in a pyrosequencing assay in accordance with an embodiment disclosed herein.

FIG. 4 shows the determination of the immediate ascending A and G for each clone with 10A, 11A, 12A, 13A and 14A primers in a pyrosequencing assay in accordance with an embodiment disclosed herein.

EXAMPLES

Example embodiments of the disclosure will be better understood and readily apparent to one of ordinary skill in the art from the following discussions and if applicable, in conjunction with the figures. It will be appreciated that the example embodiments are illustrative, and that various modifications may be made without deviating from the scope of the invention. Example embodiments are not necessarily mutually exclusive as some may be combined with one or more embodiments to form new exemplary embodiments.

The accurate detection of the sequences is important for verification and de novo assembly purposes. Inaccurate detection of sequences could bring great impact to the downstream works e.g., wrong selection of therapy due to wrong detection of mutation or incorrect identification of human, microbial or viral sequences due to wrongly annotated sequences. Optimal PCR assays with specific primers that could amplify the specific target region, as well as efficient sequencing assay for accurate

read-out of the sequences, are important for accurate sequencing. However, accurate sequencing is often complicated at regions of high polymorphisms or repeated sequences.

Sanger sequencing could only provide semi-quantitative readout, it could not give accurate readout for regions with high polymorphism of various frequency or repeated sequences. Though next-generation sequencing technology could provide qualitative and quantitative information for sequences in regions of high polymorphism, it could not provide accurate readout in regions of repeated sequences with routine mapping and analysis pipeline. In contrast, pyrosequencing could address these limitations of Sanger and next-generation sequencing, by providing accurate and quantitative readout without expensive equipment and complicated workflow. However, pyrosequencing has low accuracy in determining the sequences for regions with repeated sequences (Table 1).

Features	Sanger sequencing	Next-generation sequencing	Pyrosequencing
Cost	Low	High	Low
Wet-lab workflow	Simple	Complicated	Simple
Analytical workflow	Simple	Complicated	Simple
Qualitative readout	Yes	Yes	Yes
Quantitative readout	No	Yes	Yes
Accuracy in regions of high polymorphism	Low	High	High
Accuracy in regions of repeated sequences	High	Possible but with complicated sequence annotation pipeline	Low

Table 1. The comparison of different features on Sanger, next-generation and pyrosequencing platforms.

To resolve this, a new protocol that allow for accurate determination and/or increase the accuracy of the determination of target sequences in genomic regions with repeated sequences is developed.

Pyrosequencing serves as an excellent sequencing solution for assay that requires sequencing reads from short or fragmented sample without complicated

equipment and a time-consuming analytical pipeline. However, as mentioned above, the existing pyrosequencing assay is limited in its ability to accurately determine sequences at regions with repeated sequences.

Pyrosequencing is a sequencing by synthesis method which relies on the detection of a signal on nucleotide incorporation. In conventional pyrosequencing, a single-stranded polynucleotide to be sequenced is contacted with a primer for hybridizing to the polynucleotide to determine the starting point of complementary strand synthesis. The four kinds of nucleotides are then added, one at a time, to the reaction mixture in a designated order. If the nucleotide added is the next complementary nucleotide to the sequence of the sample, the nucleotide is incorporated into the complementary strand and the complementary strand is extended by one base length. This incorporation of nucleotide is translated into a detectable signal. If the nucleotide added is not the next complementary nucleotide and not incorporated into the complementary strand, no signal is emitted. The nucleotide is then degraded/decomposed before a next nucleotide is added and monitored for signal emission. The four kinds of nucleotides are added repeatedly in a designated order and the base sequence of the sample is determined one by one according to the presence or absence of signal emitted each time.

It will be appreciated that, in pyrosequencing, theoretically, the peak intensities in a pyrosequencing flow-gram should be directly proportional to the incorporated bases during one nucleotide dispensation. For example, when two dCTPs are incorporated into the complementary strand during the dispensation of the nucleotide dCTP, the intensity of the signal emitted should be twice of that of a signal emitted when one dCTP is incorporated. However, conventional pyrosequencing method is known to be experimentally imprudent to sequence long repeat regions. For example, in conventional pyrosequencing, the signal is known to saturate with the incorporation of three or four identical nucleotides, resulting in inaccurate determination of the sequence in a repeat region with multiple adjacent identical nucleotides/sequences. Without being bound by theory, it is believed that this may be due to non-linear light response following incorporation of several identical nucleotides, or the homopolymeric regions may reduce synchronized extension and synthesis of a polynucleotide strand, causing non-uniform peak heights.

A graduated consensus primer approach was used to tackle this limitation of pyrosequencing.

First, Sanger sequencing was performed on a test sample to obtain its sequence at region with repeated and variable sequences. However, the reads at region of variable sequences could not be determined (FIG. 1).

Since the test sample is heterogenous with mix population of cells with different number of repeats, it has complicated the Sanger sequencing readout. To resolve this, a new strategy was adopted. PCR was performed on the target region, followed by cloning into vector and PCR expansion. Next, graduated census primers were used, which allow pyrosequencing readout on the number of immediate ascending As and Gs (FIG. 2). Then, pyrosequencing was performed using the graduated census primers on these DNA clones with verified number of A repeats. Based on the pyrosequencing result, the counts of A and G (the immediate ascending sequence after the A repeats) were tabulated for each clones (FIG. 3 and FIG.4). The number of As could be determined accurately with the pyrosequencing.

Results

In FIG.3, a sequencing primer with 9A (CATTGCTCTACAAAAAAAAA) was used. For colony 14, the homopolymer region had 10T so 9 out of 10T were complementarily bound to A of the sequencing primer, leaving behind only 1T. Thus, during pyrosequencing, when nucleotide A was dispensed, a signal equivalent to 1A could be observed. Similarly, for colony 12 (containing 11A), colony 11 (containing 12A), colony 16 (containing 13A), when nucleotide A was dispensed, signals equivalent to 2A, 3A and 4A could be observed respectively. From above results, the number of A present in the homopolymer region in the sample/colonies can be easily determined. The total number of A in the homopolymer region = number of A in homopolymer region of sequencing primer (e.g., 9A in this example) + number of A incorporated as determined by pyrosequencing.

Based on above approach, similar predicted results were obtained when primers with 10A, 11A, 12A, 13A and 14A were used (FIG. 4).

Materials and methods

Materials and methods employed at the various stages in the example are summarised in this section. It will be appreciated that embodiments of the sequencing

methods are not limited to the materials and methods described in this section, and may also work with other suitable materials and methods.

1. Amplification of target region

Reagents	Vendor	Catalogue Number
Q5® Hot Start High-Fidelity DNA Polymerase	New England Biolabs (NEB)	M0491
SSA2_FP (Forward primer) Sequence: CATCGAGATCATCCTTTGCAT	Integrated DNA Technologies (IDT)	N/A
SSA4_RP Sequence: TCCTTATGGATCATTGTGGAAT	Integrated DNA Technologies (IDT)	N/A
RNase-Free Water	QIAGEN	129112

Equipment	Vendor	Part Number
Rotor-Gene Q 5plex Platform	QIAGEN	9001570

PCR Setup:

Component	Volume/Reaction
5X Q5 Reaction Buffer	5.00 µL
10 mM dNTPs	0.50 µL
10 µM Forward Primer	1.25 µL
10 µM Reverse Primer	1.25 µL
Template DNA	Variable
Q5 High-Fidelity DNA Polymerase	0.25 µL
Nuclease-Free Water	to 25 µL

PCR Cycling Steps:

Step	Temp.	Time	No. of Cycle
------	-------	------	--------------

Initial Denaturation	95°C	15min	1
Denaturation	94°C	15sec	35
Annealing	56°C	30sec	
Extension	72°C	30sec	
Final Extension	72°C	5min	1

2. Cloning of target DNA into vector

Reagents/consumables	Vendor	Catalog Number
QIAGEN PCR Purification kit	QIAGEN	28104
Nuclease-free water	Promega	P119C
TOPO TA Cloning Kit	Life Technologies	45-0641
S.O.C. medium (provided by FM group)	Invitrogen	46-0700/15544-034
STBL3 cells	In-house preparation	NA
LB+Amp plates	In-house preparation	NA
0.1M IPTG	In-house preparation	NA
X-GAL	Fermentas	R0941

The amplified target DNA was then purified and cloned into vector using TA cloning kit.

Method used for PCR product purification

- Add ethanol (96–100%) to Buffer PE before use (see bottle label for volume).
 - Add 1:250 volume pH indicator I to Buffer PB. The yellow colour of Buffer PB with pH indicator I indicates a pH of ≤ 7.5
1. Add 5 volumes Buffer PB to 1 volume of the PCR reaction and mix. *If the colour of the mixture is orange or violet, add 10 μ l 3 M sodium acetate, pH 5.0, and mix. The colour of the mixture will turn yellow.*

2. Place a QIAquick column in a provided 2 ml collection tube.
3. To bind DNA, apply the sample to the QIAquick column and centrifuge for 30–60 seconds.
4. Discard flow-through and place the QIAquick column back in the same tube.
5. To wash, add 0.75 ml Buffer PE to the QIAquick column centrifuge for 30–60 seconds.
6. Discard flow-through and place the QIAquick column back in the same tube.
7. Centrifuge the QIAquick column once more in the provided 2 ml collection tube for 1 minute to remove residual wash buffer.
8. Place each QIAquick column in a clean 1.5 ml microcentrifuge tube.
9. To elute DNA, add 30 μ l water (pH 7.0– 8.5, Promega, #P119C) to the center of the QIAquick membrane and centrifuge the column for 1 minute.

Adding A overhang to purified PCR product

Prepare the reaction mix according to the below table:

	Final Concentration	Volume(μ L)
Purified PCR product (100ng)	0.15 to 1.5 pmol	variable
dATP (10 mM)	0.2 mM	1
PCR Buffer with Mg (10x)	1x (1.5 mM MgCl ₂)	5
Taq DNA Polymerase (5 U/ μ L)	1U	0.2
ddH ₂ O	-	up to 50 μ L

1. The A-addition reaction works better when a specific amount of the PCR product is used. A good amount is 10–100 ng PCR product for each 100 bp length of the PCR product. This corresponds to 0.15–1.5 pmol PCR product (see table below).

PCR product size	Amount of PCR product to use
100 bp	10–100 ng
250 bp	25–250 ng
1000 bp	100–1000 ng

Use 100ng PCR product.

2. Incubate for 20 minutes at 72 °C.
3. Proceed to TA cloning. For optimal ligation efficiency, it's better to use fresh PCR products, since 3'A-overhangs will gradually be lost during storage.

Ligation

1. Prepare the following mix:

	Ligation A	Ligation B
Reagent	Volume (μL)	Volume (μL)
PCR product (0.5-4μL)	4 ng/ μL)	4ng/ μL)
Salt solution	1	1
dH2O	X	X
TOPO vector	1	1
<i>Total</i>	<i>6</i>	<i>6</i>

2. Mix gently.
3. Incubate at room temperature overnight
4. Place on ice and continue with transformation.

Transformation

Pre-warm LB-Amp plates to 37°C, then spread 40μL X-Gal and 100μL IPTG on top, put back to 37°C.

1. Thaw 50μL STBL3 cells on ice.
2. Add 2μL of ligated vector.
3. Mix gently. Do not pipette up and down.
4. Incubate on ice for 15 minutes (5-30 minutes).
5. Heat shock for 30 seconds at 42°C in water-bath.
6. Immediately put on ice.
7. Add 250μl S.O.C. medium.
8. Shake at 37°C, 300rpm for 1.5 hours.

9. Plate onto LB+Amp+IPTG+X-Gal plates.
 - a. A: 100µL
 - b. B: 20µL ligation + 100µL SOC
10. Check colonies the next day.
 - a. Several hundred colonies should be present.

3. Colony PCR

A quick colony PCR was done using below mentioned reagents.

Reagents	Vendor	Catalogue Number
Q5® Hot Start High-Fidelity DNA Polymerase	New England Biolabs (NEB)	M0491
SSA2_FP (Forward primer) Sequence: CATCGAGATCATCCTTTGCAT	Integrated DNA Technologies (IDT)	N/A
SSA4_RP Sequence: TCCTTATGGATCATTGTGGAAT	Integrated DNA Technologies (IDT)	N/A
RNase-Free Water	QIAGEN	129112

Equipment	Vendor	Part Number
Rotor-Gene Q 5plex Platform	QIAGEN	9001570

PCR Setup:

Component	Volume/Reaction
5X Q5 Reaction Buffer	5.00 µL
10 mM dNTPs	0.50 µL
10 µM Forward Primer	1.25 µL
10 µM Reverse Primer	1.25 µL
Template DNA	variable
Q5 High-Fidelity DNA Polymerase	0.25 µL

Nuclease-Free Water	to 25.00 μ L
---------------------	------------------

PCR Cycling Steps:

Step	Temp.	Time	No. of Cycle
Initial Denaturation	95°C	15min	1
Denaturation	94°C	15sec	35
Annealing	56°C	30sec	
Extension	72°C	30sec	
Final Extension	72°C	5min	1

Colonies that showed amplification for target DNA were further propagated in the LB broth.

4. Preparation of amplicons for pyrosequencing:

Cloned target DNA from various colonies were amplified using biotinylated primers for pyrosequencing.

Reagents	Vendor	Catalogue Number
Q5® Hot Start High-Fidelity DNA Polymerase	New England Biolabs (NEB)	M0491
SSA2_FP (Forward primer) Sequence: CATCGAGATCATCCTTTGCAT	Integrated DNA Technologies (IDT)	N/A
SSA4_BRP (Reverse primer) Sequence: /5Biosg/TCCTTATGGATCATTGTGGAAT	Integrated DNA Technologies (IDT)	N/A
RNase-Free Water	QIAGEN	129112

Equipment	Vendor	Part Number
-----------	--------	-------------

Rotor-Gene Q 5plex Platform	QIAGEN	9001570
-----------------------------	--------	---------

PCR Setup for amplification of cloned target DNA (using biotinylated Reverse primer):

Component	Volume/Reaction
5X Q5 Reaction Buffer	5.00µL
10 mM dNTPs	0.50µL
10 µM Forward Primer (SSA2_FP)	1.25µL
10 µM Reverse Primer (SSA4_BRP)	1.25µL
Template DNA	1.00µL
Q5 High-Fidelity DNA Polymerase	0.25µL
Nuclease-Free Water	15.75µL
Total	25.00µL

PCR Cycling Steps:

Step	Temp.	Time	No. of Cycle
Initial Denaturation	95°C	15min	1
Denaturation	94°C	15sec	35
Annealing	56°C	30sec	
Extension	72°C	30sec	
Final Extension	72°C	5min	1

5. Pyrosequencing Run:

Reagents	Vendor	Catalogue Number
RNase-Free Water	QIAGEN	129112
SQR1-B (8A): CATTGCTCTACAAAAAAA	Integrated DNA Technologies (IDT)	N/A

SQR1-C (9A): CATTGCTCTACAAAAAAAAA	Integrated DNA Technologies (IDT)	N/A
SQR1-D (10A): CATTGCTCTACAAAAAAAAA	Integrated DNA Technologies (IDT)	N/A
SQR1-E (11A): CATTGCTCTACAAAAAAAAA	Integrated DNA Technologies (IDT)	N/A
SQR1-12A: CATTGCTCTACAAAAAAAAA	Integrated DNA Technologies (IDT)	N/A
SQR1-13A: CATTGCTCTACAAAAAAAAA	Integrated DNA Technologies (IDT)	N/A
SQR1-C14A: CATTGCTCTACAAAAAAAAA	Integrated DNA Technologies (IDT)	N/A
PyroMark Q24 Gold Reagents	QIAGEN	970802
Binding Buffer	QIAGEN	979006
Denaturation Solution	QIAGEN	979007
Annealing Buffer	QIAGEN	979009
Streptavidin Mag Sepharose Beads	GE Healthcare	28-9857-38
Waste Strips	BMS	N/A

Equipment	Vendor	Part Number
Qseq	Bio Molecular Systems (BMS)	N/A
Vortex Mixer	Corning	N/A
Microfuge	Myfuge	N/A

Assay setup:

Prepare pyrosequencing assay file using Qseq software (refer to Qseq user manual); set nucleotide dispensation order based on sequence to be analysed.

Aim	Dispensation order
To check with which	AGAGGCTGATGAT

sequencing primer G peak appears	
To check Multiple-A	GCTCTAC <u>ACACACACACACACACACACACAG</u> ACGCTATGAT

Run Setup:

- a. Standard Gold V6 protocol.
- b. Primer Loading: Manual (if more than 3 primers are used).

Disc Setup:

- a. Drag saved assay file onto wells.
- b. Enter sample information.
- c. Save run file to USB drive.

Machine Setup

- a. Select the Run file on the instrument via the USB flash drive.
- b. Check that a waste strip is inserted into the chamber.
- c. Load the required volumes of reagent displayed on instrument touch panel.
- d. Selected last run of the day (appeared on touch panel).
- e. Prime and test the injectors.

Note: always clean injectors before running sequencing run.

Loading of magnetic beads and PCR amplicons:

- a. Insert 48 well disc into Qseq machine
- b. Load 2µL of Streptavidin Mag sepharose beads to each well.
- c. Load 10 µL of PCR amplicon to the respective wells.

Manual primer loading

- a. Machine will prompt for manual primer loading.
- b. Load 2µL of sequencing primer into respective wells.

- c. Follow instruction on display to continue sequencing run.

Upon run completion

- a. Clean injectors, follow instructions on display
- b. Save result file onto USB.
- c. Analyze run.

Result analysis:

After completion of sequencing run copy the result file to USB drive and analysis it using Qseq software.

Conclusion

Presence of repeated sequences can be found in many samples. Existing pyrosequencing assay and protocol could not overcome its limitation for low accuracy in calling sequences at regions with repeated sequences. With the ability to determine the sequences of repeated regions accurately, embodiments of the method as described herein could overcome the current limitations of the pyrosequencing assay. Increased accuracy for pyrosequencing will bring great impacts in its application such as accurate calling of mutation and microbial sequencing.

Embodiments of the method can be implemented in the development of companion diagnostic assays or academic research works that requires accurate determination of the sequences e.g., detection of mutation for selection of therapy, detection of variants for somatic or germline mutation, detection for the presence of microorganisms that requires accurate sequences for annotation.

It will be appreciated by a person skilled in the art that other variations and/or modifications may be made to the embodiments disclosed herein without departing from the spirit or scope of the disclosure as broadly described. For example, in the description herein, features of different exemplary embodiments may be mixed, combined, interchanged, incorporated, adopted, modified, included etc. or the like across different exemplary embodiments. The present embodiments are, therefore, to be considered in all respects to be illustrative and not restrictive.

CLAIMS

1. A pyrosequencing method for sequencing a polynucleotide sample comprising a homopolymeric region R_s , the method comprising:
 - providing template polynucleotides comprising the polynucleotide sample and copies thereof;
 - providing a plurality of R_p -containing primers comprising, at each of their 3' end, a homopolymeric region R_p comprising nucleotides that are complementary to the nucleotides in the homopolymeric region R_s , and
 - contacting the plurality of R_p -containing primers with the template polynucleotides under conditions suitable for hybridization and synthesis of complementary polynucleotides,wherein the homopolymeric regions R_p in the plurality of R_p -containing primers are of different lengths.
2. The pyrosequencing method according to claim 1, wherein the homopolymeric regions R_p in the plurality of R_p -containing primers comprise a length selected from: X nucleotides (nt), X+1 nt, X+2 nt, X+3 nt, X+4 nt, X+5 nt, X+6 nt, X+7 nt, X+8 nt, X+9 nt, X+10 nt and combinations thereof, wherein X is an integer no less than 3.
3. The pyrosequencing method according to claim 1 or claim 2, wherein the plurality of R_p -containing primers comprise at least three R_p -containing primers.
4. The pyrosequencing method according to any one of claims 1-3, wherein contacting the plurality of R_p -containing primers with the template polynucleotides comprises partitioning the template polynucleotides into a plurality of partitions of template polynucleotides and contacting each partition with a R_p -containing primer with a homopolymeric region R_p of a different length.
5. The pyrosequencing method according to any one of claims 1-3, wherein contacting the plurality of R_p -containing primers with the template polynucleotides comprises applying the plurality of R_p -containing primers sequentially to the template polynucleotides.

6. The pyrosequencing method according to any one of claims 1-5, the method further comprising:
 - adding a nucleotide N_1 to the template polynucleotides after the template polynucleotides are contacted with a R_p -containing primer, the nucleotide N_1 corresponding to the nucleotides in the homopolymeric region R_p ; and
 - determining whether the nucleotide N_1 is incorporated in the complementary polynucleotides.
7. The pyrosequencing method according to any one of claims 1-6, the method further comprising:
 - adding a nucleotide N_2 to the template polynucleotides after the template polynucleotides are contacted with a R_p -containing primer, wherein nucleotide N_2 is different from the nucleotides in the homopolymeric region R_p , optionally wherein nucleotide N_2 is complementary to an adjacent nucleotide that is directly next to the homopolymeric region R_s on a 5' side in the polynucleotide sample; and
 - determining whether the nucleotide N_2 is incorporated in the complementary polynucleotides.
8. The pyrosequencing method according to claim 6 or claim 7, the method further comprising:
 - identifying the R_p -containing primer contacted with the template polynucleotides when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides; and
 - concluding that the homopolymeric region R_s has a length corresponding to the length of the homopolymeric region R_p of the identified R_p -containing primer.
9. The pyrosequencing method according to claim 6, the method further comprising:
 - identifying the R_p -containing primer contacted with the template polynucleotides when nucleotide N_1 is incorporated in the complementary polynucleotides;

determining the number of nucleotide N_1 incorporated; and
concluding that the homopolymeric region R_s has a length corresponding to the sum of the number of nucleotide N_1 incorporated and the length of the homopolymeric region R_p of the identified R_p -containing primer.

10. The pyrosequencing method according to claim 8 or claim 9, the method further comprising:

verifying that nucleotide N_1 is incorporated in the complementary polynucleotides of template polynucleotides contacted with a R_p -containing primer comprising a shorter homopolymeric region R_p than that of the identified R_p -containing primer before the concluding step.

11. The pyrosequencing method according to claim 6 or claim 7, wherein where the nucleotide N_1 is not incorporated in the complementary polynucleotides of the template polynucleotides contacted with the plurality of R_p -containing primers, the method further comprises:

- a) providing a further R_p -containing primer comprising a shorter homopolymeric region R_p than those of the plurality of R_p -containing primers;
- b) contacting the template polynucleotides with the further R_p -containing primer;
- c) determining whether the nucleotide N_1 and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of the template polynucleotides of step b); and
- d) concluding that the homopolymeric region R_s has a length corresponding to the length of the homopolymeric region R_p of the further R_p -containing primer when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of step c).

12. The pyrosequencing method according to claim 6 or claim 7, wherein the nucleotide N_1 is incorporated in the complementary polynucleotides of the template polynucleotides contacted with the plurality of R_p -containing primers, the method further comprises:

- a) providing a further R_p -containing primer comprising a longer homopolymeric region R_p than those of the plurality of R_p -containing primers;
- b) contacting the template polynucleotides with the further R_p -containing primer;
- c) determining whether the nucleotide N_1 and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of the template polynucleotides of step b); and
- d) concluding that the homopolymeric region R_s has a length corresponding to the length of the homopolymeric region R_p of the further R_p -containing primer when nucleotide N_1 is not incorporated and/or the nucleotide N_2 is incorporated in the complementary polynucleotides of step c).

13. The pyrosequencing method according to claim 11 or claim 12, the method further comprising:

verifying that nucleotide N_1 is incorporated in the complementary polynucleotides of the template polynucleotides contacted with a R_p -containing primer comprising a shorter homopolymeric region R_p than that of the further R_p -containing primer before the concluding step.

14. The pyrosequencing method according to any one of claims 1-13, the method further comprising amplifying the polynucleotide sample and/or cloning the polynucleotide sample into a vector to obtain the template polynucleotides.

15. The pyrosequencing method according to any one of claims 1-14, the method further comprising:

determining a length of a homopolymeric region in a reference sequence of the polynucleotide sample prior to providing the plurality of R_p -containing primers.

16. The pyrosequencing method according to claim 15, wherein the plurality of R_p -containing primers comprises:

a R_p -containing primer which homopolymeric region R_p has a length corresponding to the length of the homopolymeric region in the reference sequence;

optionally, a R_p -containing primer which homopolymeric region R_p has a length shorter than the length of the homopolymeric region in the reference sequence; and

further optionally, a R_p -containing primer which homopolymeric region R_p has a length longer than the length of the homopolymeric region in the reference sequence.

17. A plurality of R_p -containing primers for use in the method according to any of claims 1-16, wherein the plurality of R_p -containing primers comprises, at each of their 3' end, a homopolymeric region R_p , further wherein the homopolymeric regions R_p in the plurality of R_p -containing primers are of different lengths
18. The plurality of R_p -containing primers according to claim 17, wherein the homopolymeric regions R_p in the plurality of R_p -containing primers comprise a length selected from: X nucleotides (nt), X+1 nt, X+2 nt, X+3 nt, X+4 nt, X+5 nt, X+6 nt, X+7 nt, X+8 nt, X+9 nt, X+10 nt and combinations thereof, wherein X is an integer no less than 3.
19. The plurality of R_p -containing primers according to claim 17 or claim 18, wherein the plurality of R_p -containing primers comprise at least three R_p -containing primers.
20. A kit for use in the method according to any of claims 1-16, the kit comprising:
 - the plurality of R_p -containing primers according to any one of claims 17-19; and
 - at least one more component selected from the group consisting of: nucleotides, polymerase, adenosine phosphosulfate (APS), ATP sulfurylase, luciferin, luciferase and apyrase.

Reference sequence (reverse):

AGCGTTCCTTTTGTAGCAATGAA
TCGCAAGAAATAACAATCTACCTT

Sanger sequencing of PCR product:

C	T	T	T	T	T	T	T	T	G	T	A	G	A	A	A	T	T	G	T	C	T	A	C	A	C	A
98	100	105	110	115	120	125	130	135	140	145	150	155	160	165	170	175	180	185	190	195	200	205	210	215	220	

FIG. 1

Dispensation 1: Measure A peak height
Dispensation 2: Test for G peak

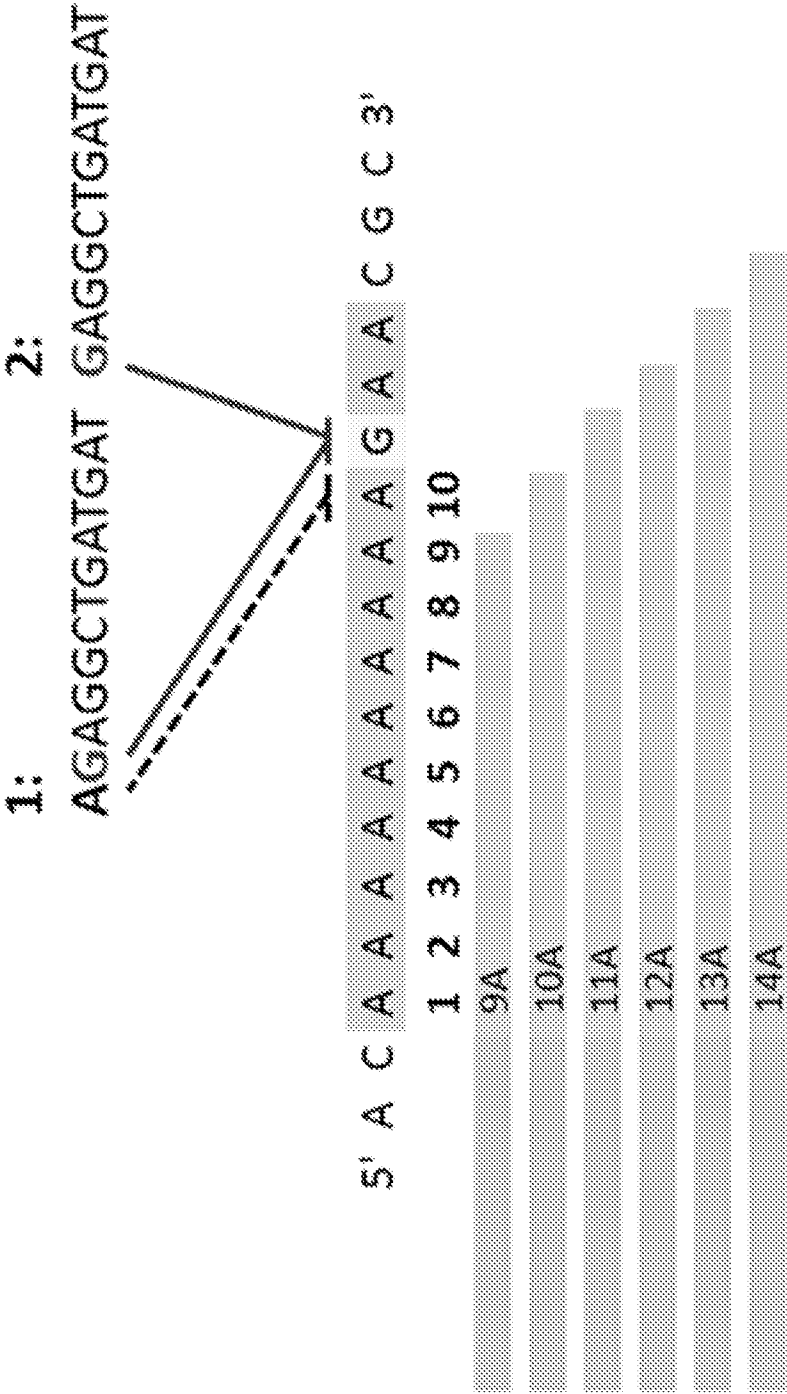


FIG. 2

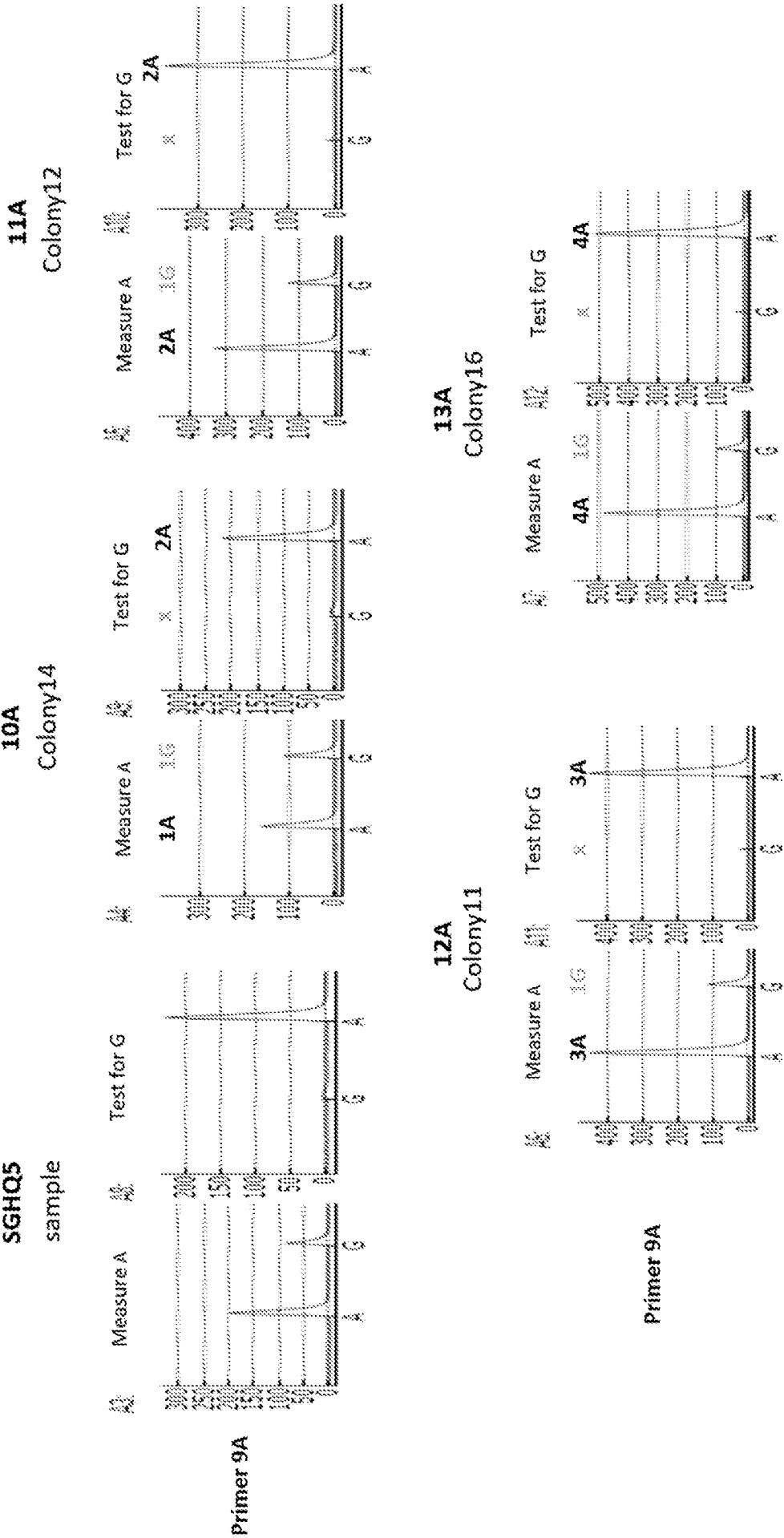


FIG. 3

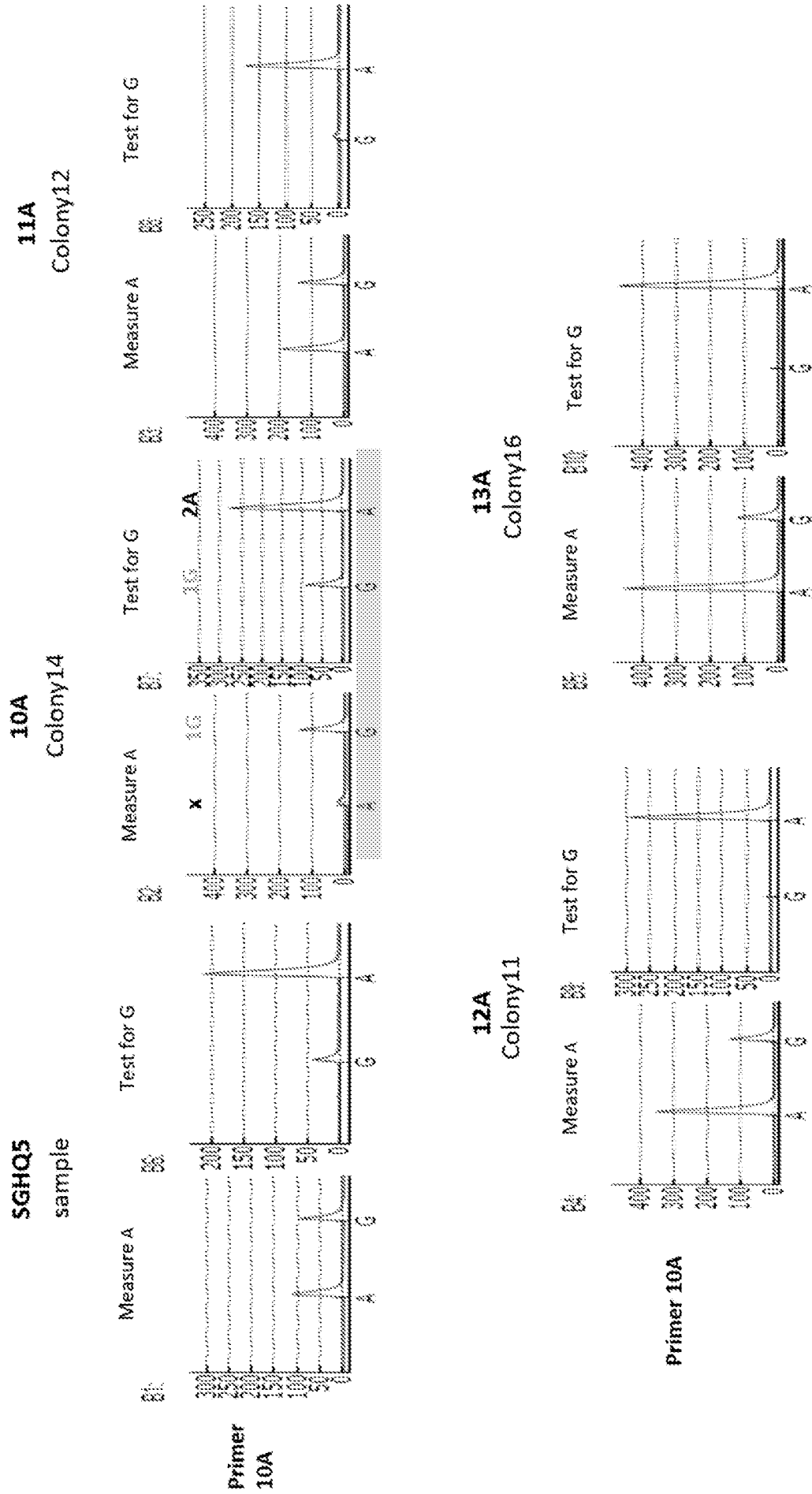


FIG. 4

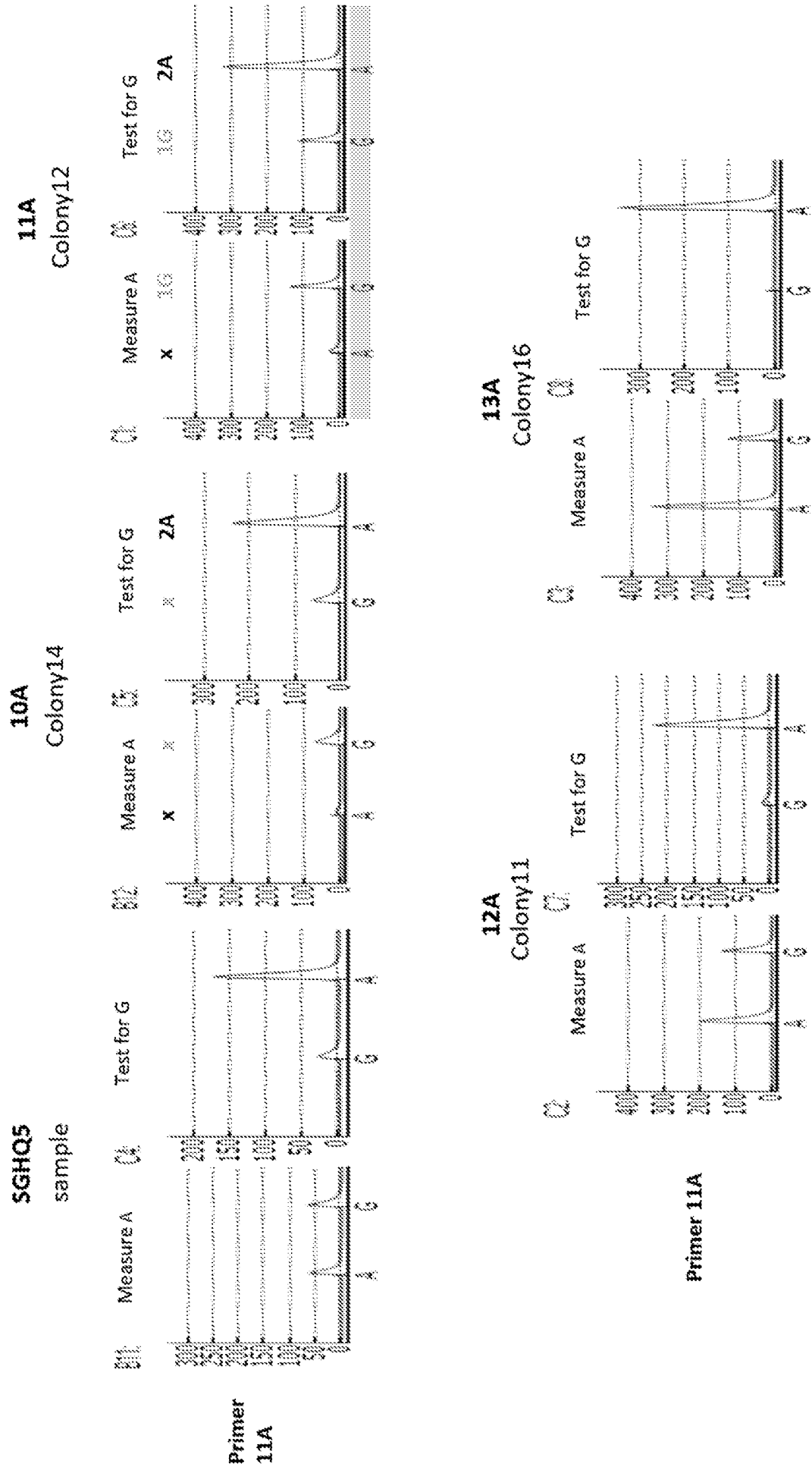


FIG. 4 (continued)

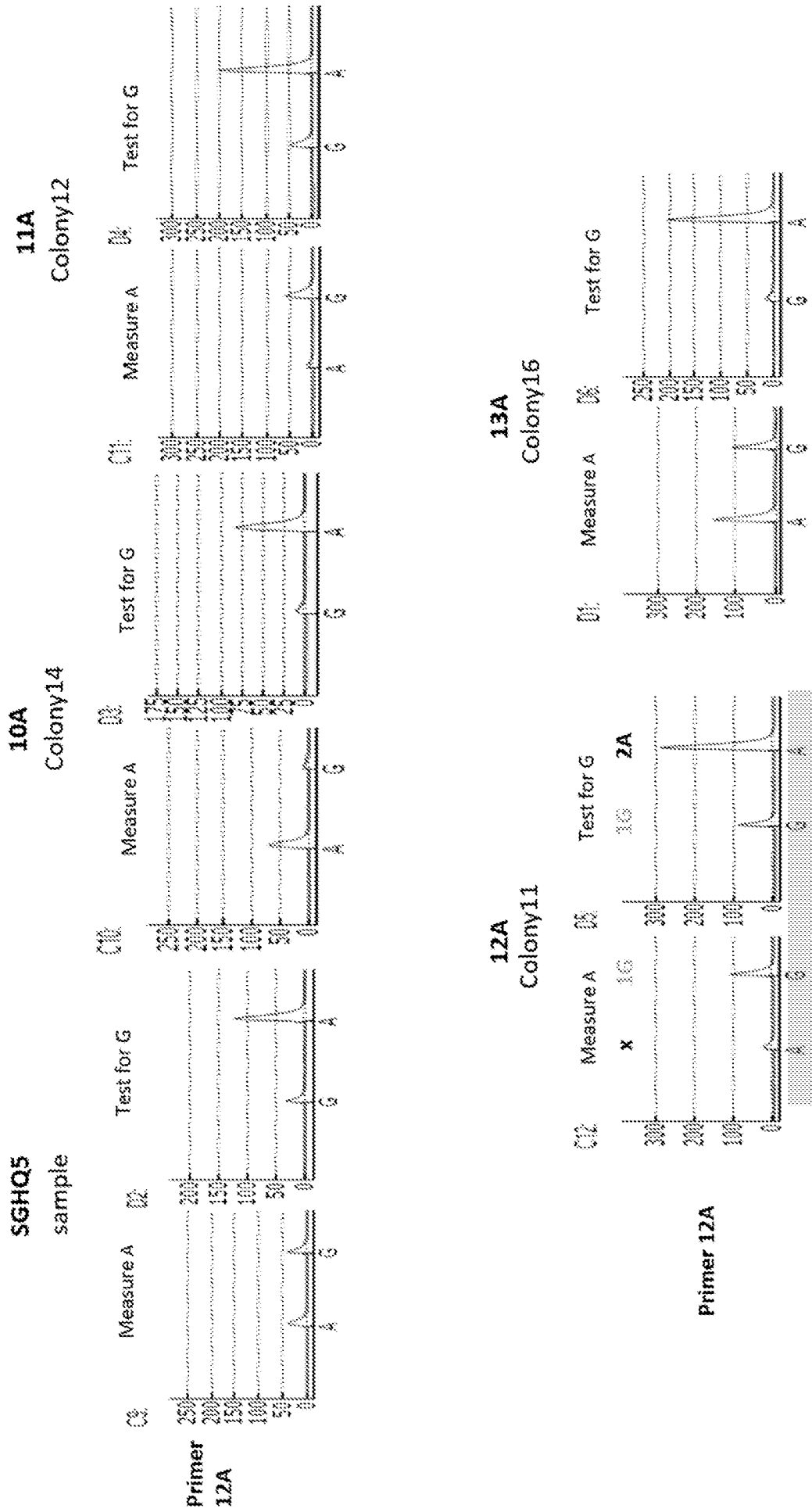


FIG. 4 (continued)

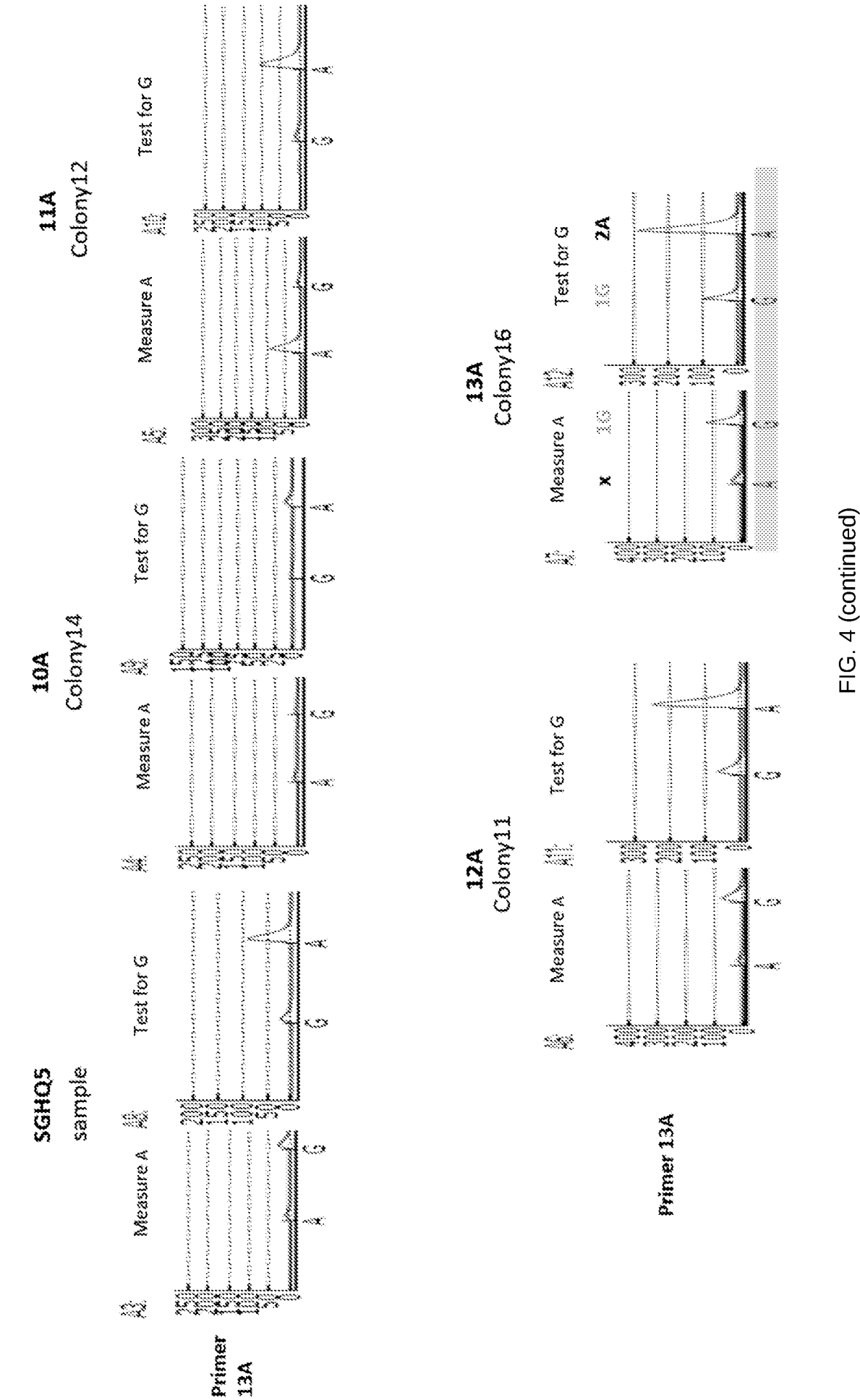


FIG. 4 (continued)

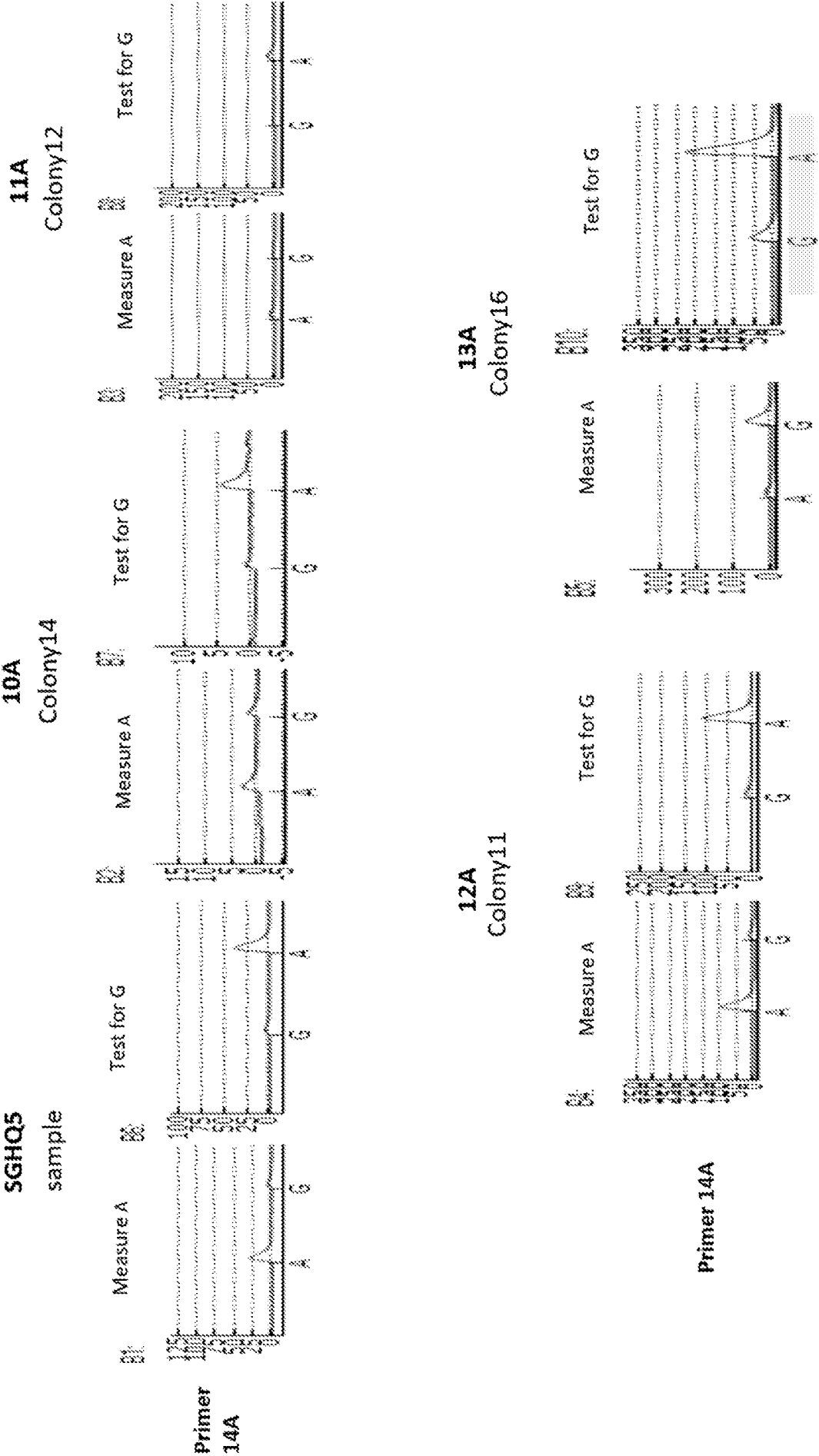


FIG. 4 (continued)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2021/050426

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/6869 (2018.01) C12Q 1/6876 (2018.01)

According to International Patent Classification (IPC)

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

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)
FAMPAT/BIOSIS/EMBASE/MEDLINE/SCISEARCH: pyrosequencing, homopolymeric and similar terms

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102776287 B (ANHUI PROVINCIAL HOSPITAL) 12 March 2014 Para. [0012] & [0023] of the original non-English language document (a machine translation is enclosed only for your reference)	17-20
A	JP 2009-011230 A (HITACHI HIGH TECH CORP) 22 January 2009 Para. [0024]; Fig. 1 of the machine translation	1-16
A	WO 2011/102808 A1 (AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH) 25 August 2011 Whole document	-
A	WO 2007/002204 A2 (THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK) 4 January 2007 Whole document	-

☒ 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

"D" document cited by the applicant in the international application

"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

14/08/2021

(day/month/year)

Date of mailing of the international search report

16/08/2021

(day/month/year)

Name and mailing address of the ISA/SG



Intellectual Property Office of Singapore

1 Paya Lebar Link, #11-03

PLQ 1, Paya Lebar Quarter

Singapore 408533

Email: pct@ipos.gov.sg

Authorized officer

Sung Ying Ying (Dr)

IPOS Customer Service Tel. No.: (+65) 6339 8616

INTERNATIONAL SEARCH REPORT

International application No.

PCT/SG2021/050426

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GONG M. ET AL., Pyrosequencing enhancement for better detection limit and sequencing homopolymers. <i>Biochem Biophys Res Commun</i> , 8 October 2010, Vol. 401, No. 1, pages 117-123 [Retrieved on 2021-08-14] <DOI: 10.1016/J.BBRC.2010.09.022> Whole document	-
A	GHARIZADEH B. ET AL., Improvements in Pyrosequencing technology by employing Sequenase polymerase. <i>Anal Biochem</i> , 15 July 2004, Vol. 330, No. 2, pages 272-280 [Retrieved on 2021-08-14] <DOI: 10.1016/J.AB.2004.03.018> Whole document	-

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/SG2021/050426

Note: This Annex lists known patent family members relating to the patent documents cited in this International Search Report. This Authority is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
CN 102776287 B	12/03/2014	CN 102776287 A	14/11/2012
JP 2009-011230 A	22/01/2009	NONE	
WO 2011/102808 A1	25/08/2011	US 2013/0045876 A1 EP 2536851 A1 US 2015/0132748 A1 SG 183403 A1	21/02/2013 26/12/2012 14/05/2015 27/09/2012
WO 2007/002204 A2	04/01/2007	US 2009/0325154 A1 US 2016/0312279 A1	31/12/2009 27/10/2016