(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 October 2018 (11.10.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/186947 A1

(51) International Patent Classification:
C12N 15/10 (2006.01) *C12Q 1/68* (2018.01)

(21) International Application Number:
PCT/US2018/019788

(22) International Filing Date:
26 February 2018 (26.02.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/482,189 06 April 2017 (06.04.2017) US
PCT/US2018/016778
04 February 2018 (04.02.2018) US

(71) Applicant: **MYOMICSDX, INC** [US/US]; 600 Washington Ave, Suite 100-W, Towson, Maryland 21204 (US).

(72) Inventors: **WANG, Qing**; 9 Rimfire Ct, Owings Mills, Maryland 21117 (US). **ZHANG, Ming**; 9 Rimfire Ct, Owings Mills, Maryland 21117 (US).

(74) Agent: **FENG, Junjie** et al.; 1024 Jewel Creek Drive, Cary, North Carolina 27519 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH,

WO 2018/186947 A1

(54) Title: METHOD AND KIT FOR TARGETED ENRICHMENT OF NUCLEIC ACIDS

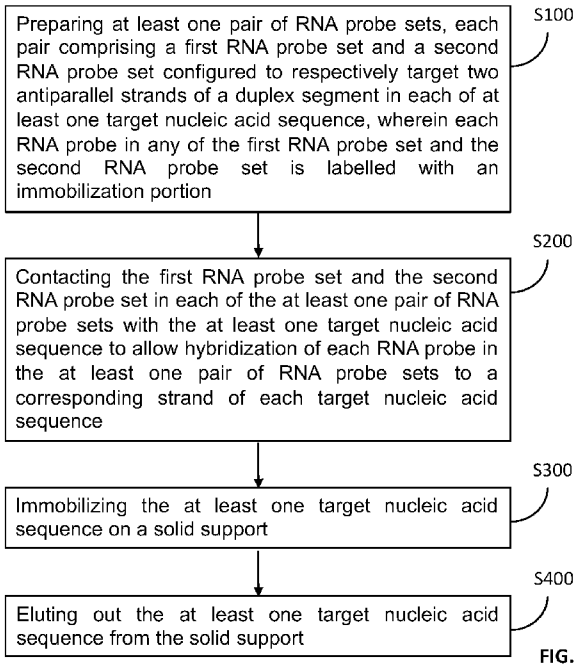

FIG. 1A

(57) Abstract: A kit and a method for enriching target nucleic acid sequences from a biological sample are disclosed. The method includes preparing, and contacting with the biological sample, a first RNA probe set and a second RNA probe set respectively targeting both of the two antiparallel strands of a duplex segment in each target nucleic acid sequence. Each RNA probe in the first RNA probe set and the second RNA probe set can be generated by chemical synthesis or by in vitro or in vivo transcription, and can be biotin-labelled to thereby allow capturing of the target nucleic acid sequences by magnetic beads labelled with streptavidin, or can be engineered to a microfluidic channel to facilitate the capturing. The method can be applied to capture double-stranded nucleic acid sequences or single-stranded nucleic acid sequences having duplex segments, and the nucleic acid sequences can include DNAs, RNAs, or DNA-RNA hybrid molecules.

[Continued on next page]

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

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*
— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

METHOD AND KIT FOR TARGETED ENRICHMENT OF NUCLEIC ACIDS

CROSS-REFERENCE TO RELATED APPLICATION

[001] The present application claims priority to U.S. Provisional Application No. 62/482,189 filed on April 6, 2017, and PCT Application No. PCT/US2018/016778 filed February 4, 2018, the disclosures of which are hereby incorporated by reference in their entirety.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[002] The content of the electronically submitted sequence listing, file name Library_SEQ_Capture_ST25.txt, size 190,883 bytes, and date of creation February 26, 2018, filed herewith, is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[003] This present disclosure relates generally to the area of diagnostics and prognostics, specifically to the field of biomarker technologies, more specifically to nucleic acid assays for genetic and genomic analysis, and in more particular to a method and a kit for targeted enrichment of nucleic acid sequences in nucleic acid assays.

BACKGROUND

[004] Genome sequencing, especially the recent next generation sequencing (NGS) technology, has emerged as a tool for global analysis of all genetic information behind a disease or an individual. For example, NGS has been widely used clinically for disease diagnostics, companion diagnostics for personalized therapeutics and disease monitoring.

SUMMARY OF THE INVENTION

[005] In a first aspect, this disclosure provides a method for enriching at least one target nucleic acid sequence from a biological sample. The method comprises the following two steps:

[006] (1) preparing at least one pair of RNA probe sets, each pair comprising a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a duplex segment in each of the at least one target nucleic acid sequence, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with an immobilization portion configured to allow immobilization onto a solid support; and

[007] (2) capturing each strand of the at least one target nucleic acid sequence from the biological sample through hybridization of both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets respectively with the two antiparallel strands of the

duplex segment in the each of the at least one target nucleic acid sequence and through immobilization onto the solid support via the immobilization portion labelled onto each RNA probe in any of the first RNA probe set and the second RNA probe set.

[008] Herein each of the at least one target nucleic acid sequence can be a double-stranded nucleic acid molecule, such as a double-stranded DNA molecule, a double-stranded RNA molecule, or a DNA-RNA hybrid molecule, or can be a single-stranded nucleic acid molecule (i.e. DNA or RNA) having a hairpin structure. As such, by means of the at least one pair of RNA probe sets where a first RNA probe set and a second RNA probe set in each pair respectively target two antiparallel strands of a duplex segment in each target nucleic acid sequence, each strand of the at least one target nucleic acid sequence can be captured or enriched from the biological sample.

[009] For a typical example, each of the at least one target nucleic acid sequence in the biological sample can be a double-stranded DNA molecule, which has a plus strand and a minus strand that runs antiparallelly to form a duplex. By means of the at least one pair of RNA probe sets where a first RNA probe set and a second RNA probe set in each pair respectively target both of the two antiparallel strands (i.e. the plus strand and the minus strand) of each target nucleic acid sequence, each strand, including both the plus strand and the minus strand, of each target DNA molecule, can be captured or enriched from the biological sample.

[010] Herein, according to some embodiments of the method, one or more of the at least one pair of RNA probe sets may be prepared through chemical synthesis. As such, the step (1) preparing at least one pair of RNA probe sets can comprise: performing chemical synthesis reactions to thereby obtain the one or more of the at least one pair of RNA probe sets. Each RNA probe in each of the one or more of the at least one pair of RNA probe sets can be labelled with the immobilization portion during the chemical synthesis reactions. Herein, the immobilization portion is covalently attached onto the solid support (i.e. each RNA probe is synthesized directly on the solid support conjugated with the immobilization portion thereof). Alternatively, each RNA probe in each of the one or more of the at least one pair of RNA probe sets can be labelled with the immobilization portion after the chemical synthesis reactions, and as such the step (1) preparing at least one pair of RNA probe sets further comprises: performing labelling reactions such that each RNA probe in each of the one or more of the at least one pair of RNA probe sets is labelled with the immobilization portion.

[011] According to some other embodiments of the method, one or more of the at least one pair of RNA probe sets is prepared through transcription. As such, the step (1) preparing at least one pair of RNA probe sets can comprise: performing transcription reactions to thereby obtain the one or more of the at least one pair of RNA probe sets.

[012] Herein, the sub-step of performing transcription reactions to thereby obtain the one or more of the at least one pair of RNA probe sets can include: performing transcription reactions such that each RNA probe in any of the one or more of the at least one pair of RNA probe sets is labelled with the immobilization portion during each of the transcription reactions. According to some embodiments, each of the transcription reactions is performed in presence of NTPs labelled with the immobilization portion, where the NTPs comprises at least one of ATPs, UTPs, GTPs, and CTPs. Herein the NTPs labelled with the immobilization portion can preferably comprise biotin-labelled UTPs, and the biotin-labelled UTPs can have a relative molar percentage of 2%-100% in all UTPs present in each of the transcription reactions.

[013] Alternatively, each RNA probe in any of the one or more of the at least one pair of RNA probe sets is labelled with the immobilization portion after each of the transcription reactions. As such, the sub-step of performing transcription reactions to thereby obtain the one or more of the at least one pair of RNA probe sets can include: performing the transcription reactions; and performing a labelling.

[014] According to some embodiments, the above sub-step of performing the transcription reactions can comprise:

[015] providing a plurality of DNA vectors, comprising at least one pair of DNA vectors, each pair comprising a first DNA vector and a second DNA vector configured to respectively allow transcription of a first RNA molecule and a second RNA molecule respectively targeting two antiparallel strands of a duplex segment in each of the one or more of the at least one target nucleic acid sequence; and

[016] performing the transcription reactions over the plurality of DNA vectors.

[017] Herein, each of the plurality of DNA vectors can include a promoter, selected from one of a T3 promoter, a T7 promoter, or a SP6 promoter. At least one of the transcription reactions can be performed *in vitro*, or *in vivo*.

[018] According to some embodiments of the method, the sub-step of performing transcription reactions over the plurality of DNA vectors comprises:

[019] pooling the plurality of DNA vectors to obtain at least two DNA vector pools, such that the first DNA vector and the second DNA vector in the each pair of DNA vectors are not in a same DNA vector pool; and

[020] performing a transcription reaction over each of the at least two DNA vector pools respectively to obtain RNA molecules corresponding to the each of the at least two DNA vector pools.

[021] After the sub-step of performing a transcription reaction over each of the at least two DNA vector pools respectively and prior to the sub-step of performing a labelling, the method can further

comprise: performing a fragmentation reaction to the RNA molecules corresponding to the each of the at least two DNA vector pools.

[022] According to some other embodiments of the method, rather than pooling the DNA vectors before transcription, the sub-step of performing transcription reactions over the plurality of DNA vectors comprises: performing a transcription reaction over each of the plurality of DNA vectors to thereby obtain an RNA molecule corresponding thereto.

[023] As such, after the sub-step of performing a transcription reaction over each of the plurality of DNA vectors and prior to the sub-step of performing a labelling, the method can further include:

[024] pooling the RNA molecule corresponding to each of the plurality of DNA vectors to obtain at least two RNA pools, such that a pair of RNA molecules respectively targeting two antiparallel strands of a duplex segment in any one of the one or more of the at least one target nucleic acid sequence are not in a same RNA pool; and

[025] performing a fragmentation reaction to each of the at least two RNA pools respectively.

[026] Alternatively, after the sub-step of performing a transcription reaction over each of the plurality of DNA vectors and prior to the sub-step of performing a labelling, the method can further include:

[027] performing a fragmentation reaction to the RNA molecule corresponding to each of the plurality of DNA vectors respectively to obtain fragmented RNA molecules corresponding to each of the plurality of DNA vectors; and

[028] pooling the fragmented RNA molecules corresponding to each of the plurality of DNA vectors such that a pair of fragmented RNA molecules respectively targeting two antiparallel strands of a duplex segment in any one of the one or more of the at least one target nucleic acid sequence are not in a same RNA probe set.

[029] In any of the embodiments of the method as described above, the sub-step of performing a labelling can comprise: performing ligation reactions such that an immobilization portion-labelled nucleotide is ligated to one terminus, or to a middle, of each RNA probe in each of the at least one pair of RNA probe sets. In one example, in each of the ligation reactions, a 5' phosphate terminus of a biotin-labeled nucleotide can be ligated to a 3' hydroxyl terminus of each RNA probe in each of the at least one pair of RNA probe sets, and each of the ligation reactions can be performed by means of an RNA ligase, which can comprise at least one of T4 RNA ligase, or CircLigase RNA ligase.

[030] According to some embodiments, the step of preparing at least one pair of RNA probe sets comprises: performing direct transcription on the solid support to thereby obtain the one or more of the at least one pair of RNA probe sets. This can be done, for example, by means of a RNA polymerase attached on the solid support.

[031] In the method disclosed herein, the step (2) of capturing each strand of the at least one target nucleic acid sequence from the biological sample can comprise the following sub-steps:

[032] contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample; and

[033] immobilizing the at least one target nucleic acid sequence on the solid support.

[034] Herein, each RNA probe in any of the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets can have a length of about 100 - 150 nt. As such, the sub-step of contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample can be performed at a temperature of about 62-70°C, and preferably at a temperature of about 67.5°C, and for about 6-24 hours, and preferably for about 12 hours.

[035] According to some embodiments of the method, the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets are configured to respectively target a different portion of the duplex segment in the each of the at least one target nucleic acid sequence. As such, the sub-step of contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample can comprise: contacting both of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a single hybridization reaction.

[036] According to some embodiments of the method, the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets are configured to respectively target a substantially same portion of the duplex segment in the each of the at least one target nucleic acid sequence. As such, the sub-step of contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample can comprise: contacting one of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a first hybridization reaction; and contacting another of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a second hybridization reaction. Alternatively, the sub-step of contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample can comprise: separately contacting the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a third hybridization reaction and a fourth

hybridization reaction, respectively; and combining the third hybridization reaction and the fourth hybridization reaction to thereby allow a fifth hybridization reaction to proceed.

[037] In the biological sample, one or more of the at least one target nucleic acid sequence in the biological sample are each in a polynucleotide comprising at least one un-targeted sequence. To minimize the interference of the at least one un-targeted sequence in the polynucleotide on capturing the at least one target sequence, prior to the contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample, the method can further include: contacting at least one blocking oligo with the at least one target nucleic acid sequence such that the at least one blocking oligo respectively hybridizes with, and thereby blocks, at least one strand of each of the at least one un-targeted sequence in the polynucleotide.

[038] Herein, if each of the one or more of the at least one target nucleic acid sequence in the biological sample is flanked by a first adaptor sequence and a second adaptor sequence in the polynucleotide, the at least one blocking oligo can be configured to respectively block one strand of the first adaptor sequence and one strand of the second adaptor sequence in the polynucleotide, and according to some embodiments, the at least one blocking oligo can be configured to respectively block both two antiparallel strands of the first adaptor sequence and both two antiparallel strands of the second adaptor sequence in the polynucleotide.

[039] According to some embodiments, the at least one blocking oligo comprises a first blocking oligo set and a second blocking oligo set, each comprising one or more blocking oligo, configured to respectively block two antiparallel strands of one of the first adaptor sequence and the second adaptor sequence in the polynucleotide, wherein the sub-step of contacting at least one blocking oligo with the at least one target nucleic acid sequence comprises: contacting one of the first blocking oligo set and the second blocking oligo set with the at least one target nucleic acid sequence; and contacting another of the first blocking oligo set and the second blocking oligo set with the at least one target nucleic acid sequence.

[040] According to some embodiments of the method, the step (2) of capturing each strand of the at least one target nucleic acid sequence from the biological sample can comprise:

[041] conjugating the at least one pair of RNA probe sets on the solid support via the immobilization portion labelled onto each RNA probe in any of the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets to thereby obtain at least one pair of solid support-conjugated RNA probe sets, each pair comprising a solid support-conjugated first RNA probe set and a solid support-conjugated second RNA probe set; and

[042] contacting both the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence.

[043] In embodiments where the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set may interfere with each other, such as when the first RNA probe sets and the second RNA probe set are labelled on beads that may be mixed in a common reaction, the sub-step of contacting both the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence can include: contacting one of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in a sixth hybridization reaction; and contacting another of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in a seventh hybridization reaction. Alternatively, the sub-step of contacting both the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence can include: separately contacting the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in an eighth hybridization reaction and a ninth hybridization reaction, respectively; and combining the eighth hybridization reaction and the ninth hybridization reaction to thereby allow a tenth hybridization reaction to proceed.

[044] In embodiments where the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set do not interfere with each other, such as when the first RNA probe sets and the second RNA probe set are conjugated on different regions on a solid surface (e.g. a glass/plastic chip, a column or a microfluidic channel), the sub-step of contacting both the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence can include: contacting both of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in a single hybridization reaction.

[045] According to some embodiments of the method, the immobilization portion can be configured to be able to form a stable non-covalent binding with a coupling partner conjugated onto surface of the solid support, and as such, the immobilization portion can comprise a biotin moiety, and correspondingly, the coupling partner conjugated onto surface of the solid support can comprise at least one of streptavidin, avidin, or an anti-biotin antibody. According to some other embodiments of the method, the immobilization portion can be configured to be able to form a covalent connection with a coupling partner conjugated onto surface of the solid support. The solid support can comprise at least one of a magnetic bead, a filter, a resin bead, a nanosphere, a plastic surface, a microtiter plate, a glass surface, a slide, a membrane, a microfluidic channel, a chip, or a matrix.

[046] The method according to any one of the embodiments as described above can further include: eluting out the at least one target nucleic acid sequence from the solid support.

[047] In a second aspect, the disclosure further provides a kit for enriching at least one target nucleic acid sequence from a biological sample utilizing the method according to any one of the embodiments as described above.

[048] According to some embodiments, the kit comprises at least one pair of RNA probe sets and a solid support labelled with a coupling partner on a surface thereof. Each pair of RNA probe sets comprises a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a duplex segment in each of the at least one target nucleic acid sequence, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with an immobilization portion. The coupling partner is configured to be able to form a secure coupling to the immobilization portion to thereby allow immobilization of each RNA probe in any of the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets onto the solid support.

[049] Herein, each RNA probe in any of the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets can have a length of about 100 - 150 nt.

[050] In the kit, the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets can be configured to respectively target a substantially same portion, or different portions, of the duplex segment in the each of the at least one target nucleic acid sequence.

[051] According to some embodiments of the kit, the immobilization portion can comprise a biotin moiety, the solid support comprises at least one of a magnetic bead, a filter, a resin bead, a nanosphere, a plastic surface, a microtiter plate, a glass surface, a slide, a membrane, a microfluidic channel, a chip, or a matrix, and the solid support can be labelled with at least one of streptavidin, avidin, or an anti-biotin antibody.

[052] According to some embodiments, the kit further comprises an apparatus having a working surface as the solid support. The first RNA probe set and the second RNA probe set in each pair are respectively conjugated onto the working surface, arranged such that each RNA probe in the solid support-conjugated first RNA probe set does not substantially interact with each RNA probe in the solid support-conjugated second RNA probe set. Herein the apparatus can be one of a column, a microfluidic channel, or a chip.

[053] Optionally, the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in each pair can be respectively arranged at at least one, and preferably more than one, pair of two different regions of the working surface of the apparatus.

[054] Optionally, the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in each pair can be mixedly arranged on the working surface of the apparatus, configured such that each RNA probe from the first RNA probe set has a relatively large distance to each RNA probe from the second RNA probe set to thereby substantially prevent an interaction therebetween.

[055] In any of the above embodiments, the apparatus can be configured to allow the biological sample to flow sequentially through the working surface for more than one round.

[056] In the biological sample, one or more of the at least one target nucleic acid sequence are each in a polynucleotide comprising at least one un-targeted sequence. As such, the kit can further include at least one blocking oligo, configured to respectively hybridize with, and to thereby block, at least one strand of each of the at least one un-targeted sequence in the polynucleotide. If the at least one un-targeted sequence in the polynucleotide comprises a first adaptor sequence and a second adaptor sequence flanking each of the one or more of the at least one target nucleic acid sequence, the at least one blocking oligo can be configured to respectively block one strand of the first adaptor sequence and one strand of the second adaptor sequence in the polynucleotide, or to respectively block both two antiparallel strands of the first adaptor sequence and both two antiparallel strands of the second adaptor sequence in the polynucleotide. As such, in the kit, the at least one blocking oligo can comprise a first blocking oligo set and a second blocking oligo set, configured to respectively block two antiparallel strands of one of the first adaptor sequence and the second adaptor sequence in the polynucleotide. The first blocking oligo set and the second blocking oligo set can be configured to respectively target two different portions within the one of the first adaptor sequence and the second adaptor sequence in the polynucleotide.

[057] According to some other embodiments, the kit can include a plurality of DNA vectors, NTPs comprising each of ATPs, UTPs, GTPs, and CTPs; immobilization portion-labelled NTPs, wherein NTPs comprises at least one of ATPs, UTPs, GTPs, and CTPs; and a solid support labelled with a coupling partner on a surface thereof. The plurality of DNA vectors comprises at least one

pair of DNA vectors, each pair comprising a first DNA vector and a second DNA vector configured, via transcription thereover, to respectively obtain a first RNA probe set and a second RNA probe set targeting respectively two antiparallel strands of a duplex segment in each of the at least one target nucleic acid sequence. The coupling partner is configured to be able to form a secure coupling to the immobilization portion.

[058] Herein, the immobilization portion can include a biotin moiety, and the NTPs labelled with the immobilization portion can comprise biotin-labelled UTPs, having a relative molar percentage of 2%-100% among all UTPs in the kit. The solid support can comprise comprises at least one of a magnetic bead, a filter, a resin bead, a nanosphere, a plastic surface, a microtiter plate, a glass surface, a slide, a membrane, a microfluidic channel, a chip, or a matrix, and the solid support can be labelled with at least one of streptavidin, avidin, or an anti-biotin antibody.

[059] According to some embodiments, the kit can further include an RNA ligase, comprising at least one of T4 RNA ligase or CircLigase RNA ligase, which is configured to ligate a 3' hydroxyl terminus of each RNA probe in any of the first RNA probe set and the second RNA probe set generated from each of the at least one pair of DNA vectors with one of the immobilization portion-labeled NTPs.

[060] Each of the plurality of DNA vectors may comprise a DNA template and a promoter. The DNA template may comprise a sequence corresponding to one of two antiparallel strands of a duplex segment in each of the at least one target nucleic acid sequence. The promoter can be configured to initiate a transcription reaction of the DNA template in a presence of an RNA polymerase compatible with the promoter. The promoter is selected from one of a T3 promoter, a T7 promoter, a SP6 promoter, and the kit can correspondingly further comprise a T3 RNA polymerase, a T7 RNA polymerase, or a SP6 RNA polymerase, corresponding to the promoter in each of the plurality of DNA vectors.

[061] The kit can further comprise cells or viruses containing the RNA polymerase compatible with the promoter in each of the plurality of DNA vectors. The cells can include at least one of a bacterial cell line, a yeast cell line, or a mammalian cell line. Depending on the different host cells or host viruses, the promoter can be any of the T3 promoter, T7 promoter, and SP6 promoter, and can optionally be a tissue or cell line-specific promoter. There are no limitations herein.

[062] In the kit, each of the plurality of DNA vectors can be a double-stranded DNA vector or a single-stranded DNA vector.

[063] In the biological sample, one or more of the at least one target nucleic acid sequence are each in a polynucleotide comprising at least one un-targeted sequence. As such, the kit can further include at least one blocking oligo, configured to respectively hybridize with, and to thereby block, at least one strand of each of the at least one un-targeted sequence in the polynucleotide. If the at

least one un-targeted sequence in the polynucleotide comprises a first adaptor sequence and a second adaptor sequence flanking each of the one or more of the at least one target nucleic acid sequence, the at least one blocking oligo can be configured to respectively block one strand of the first adaptor sequence and one strand of the second adaptor sequence in the polynucleotide, or to respectively block both two antiparallel strands of the first adaptor sequence and both two antiparallel strands of the second adaptor sequence in the polynucleotide. As such, in the kit, the at least one blocking oligo can comprise a first blocking oligo set and a second blocking oligo set, configured to respectively block two antiparallel strands of one of the first adaptor sequence and the second adaptor sequence in the polynucleotide. The first blocking oligo set and the second blocking oligo set can be configured to respectively target two different portions within the one of the first adaptor sequence and the second adaptor sequence in the polynucleotide.

[064] In the above-mentioned method and kit, the nucleic acid sequences may comprise DNA and/or RNA sequences from a test sample. The nucleic acid sequences can be natural sequence obtained from an organism, or can be artificial sequences manufactured manually. There are no limitations herein.

[065] The test sample can be a biological sample from an organism, or can be a sample obtained artificially or obtained after appropriate handling, as long as the test sample contain one or more nucleic acid sequences. The biological sample can be DNA, RNA or any samples composed of nucleic acid sequence(s), and the samples can be chemical synthesis products or extracted from an organism, such as multicellular animals, plants, fungi, protists, bacteria, archaea, or any types of the tissue culture of the above organisms. The organism can be from any domains from prokaryota and eukaryote. The test samples may be treated prior to enrichment. For example, the nucleic acid sequences in the biological sample can be amplified prior to enrichment and testing.

[066] The nucleic acid sequences may contain genetic markers associated with human diseases including cancer, diabetes, heart diseases, and so on, but can also include markers associated with a phenotype such as height, weight, skin color, etc. The markers may include qualitative or quantitative genetic information.

[067] Test samples can be from any appropriate sources in the patient's body that will have nucleic acids from a cancer or lesion that can be collected and tested. Test samples can be also from any appropriate sources derived from patient tissue, such as FFPE slides, FFPE tissue blocks, and test samples can be also from any appropriate sources derived from other biological specimens, such as fossils, body remains of ancient human species or animal species. Suitable test samples may be obtained from body tissue, stool, and body fluids, such as blood, tear, saliva, sputum, bronchoalveolar lavage, urine and different organ secreted juices. In some cases, the nucleic acids will be amplified prior to testing. The samples may be collected using any means conventional in

the art, including from surgical samples, from biopsy samples, from endoscopic ultrasound (EUS), phlebotomy, etc. Obtaining the samples may be performed by the same person or a different person that conducts the subsequent analysis. Samples may be stored and/or transferred after collection and before analysis. Samples may be fractionated, treated, purified, enriched, prior to assay.

[068] According to one aspect of the invention, a portion or all nucleic acids in a sample are enriched for nucleic acid analysis. A set of probes for one or more analytes of interest is synthesized. The probes are RNA probes. The probes are transcribed from DNA template sharing the same sequences as the target nucleic acids that are going to be enriched. The RNA probes are complementary to both plus and minus strands of a target nucleic acid analyte. The RNA probes complementary to plus and minus strands of a target nucleic acid analyte are synthesized in parallel reaction systems. The set of RNA probes can be massively generated by in vitro transcription using the target nucleic acid sequences as the templates. The RNA probes are bound to a solid support. The solid support is contacted with the sample comprising nucleic acids under hybridization conditions so that complementary nucleic acids in the sample are captured on the solid support, and the solid support is washed to remove non-complementary nucleic acids. Captured nucleic acids are eluted from the solid support for further analysis.

[069] According to another aspect of the invention, a portion or all nucleic acids in a sample are enriched for nucleic acid analysis. A set of probes for one or more analytes of interested is synthesized. The probes are RNA probes. The probes are transcribed from DNA template sharing the same sequences as the target nucleic acids that are going to be enriched. The RNA probes are complementary to both plus and minus strands of a target nucleic acid analyte. The RNA probes complementary to plus and minus strands of a target nucleic acid analyte are synthesized in parallel reaction systems. The set of RNA probes can be massively generated by in vitro transcription using the target nucleic acid sequences as the templates. The transcribed or synthesized RNA probes are sheared into fragments by sonication. Modified nucleic acids (such as biotinylated Uracil) can be incorporated into the RNA probes or additional nucleic acid modifications can happen on the ends or within nucleic acid sequences of RNA probes, either of which allows the RNA probes to be immobilized or captured and extracted after the capture procedure. RNA probes are contacted with the sample compromising nucleic acids under hybridization conditions so that complementary nucleic acids in the sample are captured. The hybridization reaction mixture is contacted with a solid support which compromises chemical structures (such as avidin or streptavidin) that specifically react with the modification groups (such as biotin) on the RNA probes. RNA probes with their captured complementary nucleic acids are immobilized, or captured, on the solid support. The solid support is washed to remove non-complementary nucleic acids. Captured nucleic acids are eluted from the solid support for further analysis.

[070] These and other embodiments which will be apparent to those of skill in the art upon reading the specification provide the art with methods for assessing, characterizing, and detecting genetic markers, such as cancer markers, and genetic analysis, such as SNV identification. In particular, it provides methods for enriching nucleic acids for desired analytes.

[071] Throughout the disclosure, "upstream" and "downstream" are respectively defined as nucleic acid sequences at 5' end and 3' end of a strand of nucleic acid sequence (DNA strand or RNA strand), unless indicated otherwise.

[072] Unless indicated otherwise, all DNA sequences disclosed herein have a direction from a 5' end to a 3' end thereof.

[073] Unless indicated otherwise, an oligo can be a single-stranded DNA oligo, or a single-stranded RNA oligo, having a sequence of at least 2 nt.

[074] The term "about" in the disclosure generally refers to plus or minus 10% of the indicated number. For example, "about 20" may indicate a range of 18 to 22, and "about 1" may mean from 0.9-1.1. Other meanings of "about" may be apparent from the context, such as rounding off, so, for example "about 1" may also mean from 0.5 to 1.4.

[075] The term "probe" in this disclosure is referred to as a bait molecule that can be used for capturing or enriching a target molecule, unless indicated otherwise. Specifically, in this disclosure, an RNA probe is substantially a bait RNA molecule that are used to capture and enrich a target nucleic acid sequence, including a DNA sequence or an RNA sequence or any other nucleic acid sequences that could form hybridization molecules with the RNA probes.

[076] As used herein, the term "hybridization" or "binding" or "annealing" refers to the pairing of complementary (including partially complementary) polynucleotide strands. Hybridization and the strength of hybridization (e.g., the strength of the association between polynucleotide strands) is impacted by many factors well known in the art including the degree of complementarity between the polynucleotides, stringency of the conditions involved affected by such conditions as the concentration of salts, the melting temperature (Tm) of the formed hybrid, the temperature of the hybridization reaction, the presence of other components, the molarity of the hybridizing strands and the G:C content of the polynucleotide strands. When one polynucleotide is said to "hybridize" to another polynucleotide, it means that there is some complementarity between the two polynucleotides or that the two polynucleotides form a hybrid under high stringency conditions. When one polynucleotide is said to not hybridize to another polynucleotide, it means that there is no sequence complementarity between the two polynucleotides or that no hybrid forms between the two polynucleotides at a high stringency condition. Related, the terms "target", "targeting", or alike in the disclosure, such as in a phrase "an RNA probe targeting one strand of

a DNA molecule" refers to the situation that the RNA probe can specifically hybridize, or bind, or anneal with the one strand of the DNA molecule.

[077] As used herein, the term "complementary" refers to the concept of sequence complementarity between regions of two polynucleotide strands (e.g. a double-stranded structure) or between two regions of the same polynucleotide strand (e.g. a "loop" or "hairpin" structure). It is known that an adenine base of a first polynucleotide region is capable of forming specific hydrogen bonds ("base pairing") with a base of a second polynucleotide region which is antiparallel to the first region if the base is thymine or uracil. Similarly, it is known that a cytosine base of a first polynucleotide strand is capable of base pairing with a base of a second polynucleotide strand which is antiparallel to the first strand if the base is guanine. A first region of a polynucleotide is complementary to a second region of the same or a different polynucleotide if, for example, when the two regions are arranged in an antiparallel fashion, at least one nucleotide of the first region is capable of base pairing with a base of the second region. Therefore, it is not required for two complementary polynucleotides to base pair at every nucleotide position. "Complementary" refers to a first polynucleotide that is 100% or "fully" complementary to a second polynucleotide and thus forms a base pair at every nucleotide position. "Complementary" also refers to a first polynucleotide that is not 100% complementary (e.g., 90%, or 80% or 70% complementary) contains mismatched nucleotides at one or more nucleotide positions. In one embodiment, two complementary polynucleotides are capable of hybridizing to each other under high stringency hybridization conditions.

[078] As used herein, the term "target nucleic acid" or "target" refers to a nucleic acid containing a target nucleic acid sequence to be identified. A target nucleic acid may be single-stranded or double-stranded, and often is DNA, RNA, a derivative of DNA or RNA, or a combination thereof. A "target nucleic acid sequence," "target sequence" or "target region" means a specific sequence comprising all or part of the sequence of a single-stranded nucleic acid. A target sequence may be within a nucleic acid template, which may be any form of single-stranded or double-stranded nucleic acid. A template may be a purified or isolated nucleic acid, or may be non-purified or non-isolated.

[079] A target or target nucleic acid usually exists within a portion or all of a polynucleotide, and is usually a polynucleotide analyte. The identity of the target nucleotide sequence generally is known to an extent sufficient to allow preparation of various probe/bait sequences hybridizable with the target material. The target material is generally a fraction of a larger pool of molecules or it may be substantially the entire molecule such as a polynucleotide as described above.

[080] The term "antiparallel strands" of a duplex segment of a nucleic acid sequence refers to the situation where two strands of nucleic acid sequences align in opposite directions (i.e., one strand

in a 5' end-3' end direction, and the other in a 3' end-5' end direction) and form a double-stranded (i.e. "duplex") structure due to the hybridization of the two strands. It is possible that such two antiparallel strands of the duplex segment are two complimentary DNA strands ("DNA-DNA") of a double-stranded DNA segment, two complimentary RNA strands ("RNA-RNA") of a double-stranded RNA segment, or two complimentary DNA and RNA strands ("DNA-RNA") of a double-stranded DNA-RNA segment. The nucleic acid sequence can be double-stranded, and the two antiparallel strands of the duplex segment thereof are respectively the two complimentary strands of the nucleic acid sequence. Alternatively, the nucleic acid sequence can be single-stranded, where the two antiparallel strands of the duplex segment substantially forms a "hairpin" or a "loop" segment of the single-stranded nucleic acid sequence.

[081] The term "RNA polymerase compatible with a promoter" is defined as an RNA polymerase that can recognize the promoter to thereby initiate a transcription of RNA molecules using a DNA template at a downstream of the promoter.

[082] The terms "first", "second", "third", "fourth", "fifth", etc., are intended to refer to a different object (i.e. component, composition, process, etc.) and indicates or suggests no actual order in the disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[083] FIGS. 1A-1C respectively illustrate a flow chart of a method for enriching at least one target nucleic acid sequence from a biological sample according to two embodiments of the disclosure;

[084] FIGS. 1D-1J respectively illustrate a schematic structure of an apparatus having each pair of RNA probe sets arranged on an inner surface thereof to allow the capture of target nucleic acid sequences from the biological sample according to several different embodiments of the disclosure;

[085] FIG. 2A and FIG. 2B respectively illustrate a flow chart of step S100 as shown in FIG. 1A according to two embodiments of the disclosure;

[086] FIG. 3A and FIG. 3B respectively illustrate a structural diagram of the first DNA vector and the second DNA vector for transcriptionally obtaining the first RNA molecule targeting a plus strand and the second RNA molecule targeting a minus strand of a target nucleic acid sequence according to some embodiments of the disclosure;

[087] FIG. 3C and FIG. 3D respectively illustrate a structural diagram of the first DNA vector and the second DNA vector according to some other embodiments of the disclosure;

[088] FIG. 4A and FIG. 4B respectively illustrate the process of preparing one pair of RNA probe sets respectively targeting a plus strand and a minus strand of one target nucleic acid sequence according to two embodiments of the disclosure;

[089] FIG. 5A and FIG. 5B respectively illustrate a flow chart of step S200 of the method according to two embodiments of the disclosure;

[090] FIG. 6A is a schematic diagram of the hybridization process of RNA probes with target nucleic acid sequences according to some embodiments with blocking oligos applied;

[091] FIG. 6B illustrates a double-stranded target nucleic acid sequence consisting of a plus strand and a minus strand, and a single-stranded target nucleic acid sequence whose duplex segment having a hairpin structure, can be targeted for capturing and enrichment using the method disclosed herein;

[092] FIGS. 7A and 7B illustrate the process of enriching target sequences from a DNA library generated for a NGS-based sequencing assay;

[093] FIGS. 8A-8D show a performance evaluation of DNA double strand capture based on the method. (FIG. 8A) For each of the six NGS DNA libraries derived from different amounts (500ng, 20ng, 1ng, 100pg, 20pg and 10pg) of input genomic DNA, enrichment efficiency of 298 cancer related genes were calculated. Recovery ratios of double strand DNA capture and single strand DNA capture for all 298 genes in six libraries were quantified by real-time PCR assays detecting each gene's abundance in the libraries before and after captures with different approaches (targeting double DNA strands or targeting single DNA strand). (FIG. 8B) A plasmid was constructed with an insert sequence composed of amplicon regions of five genes whose GC contents cover a broad range (27.3% to 74.1%). (FIG. 8C) Real-time PCR analysis of sequential dilutions of the plasmid illustrated in 3E. 1, 10, 100, 1,000 and 10,000 femtomoles of the plasmids were added as template for the detection. C_t value for each gene obtained from each plasmid amount was plotted, and trend lines were shown. (FIG. 8D) A original genomic DNA NGS library, and the library molecules captured by RNA probes targeting a single DNA strand or both DNA strands of a target sequence region were analyzed on an agarose gel;

[094] FIGS. 9A and 9B show SNV-calling trends and statistics of RNA probe-based DNA double strand capture WES (Whole Exome Sequencing) study;

[095] FIGS. 10A-10C show read statistics. (FIG. 10A) Bar plot of percentage of initial reads, mapped reads and reads remained after filtering. Results were obtained from three technical replicates. Numbers of reads were shown under each bar with the unit of 1 million reads. (FIG. 10B) Stacked bar plot of subgroups of filtered reads in triple replicates. (FIG. 10C) Coverage efficiency correlation with read numbers. The percentage of target bases covered at $\geqslant$ 10X, $\geqslant$ 20X, $\geqslant$50X and $\geqslant$100X depths with 5 million to 50 million reads were shown;

[096] FIGS. 11A and 11B show density plots of read depths to demonstrate the relationship between GC content and normalized mean read depth for (FIG. 11A) an NGS WES study using RNA probe-based DNA double strand capture approach with DNA extracted from normal human

tissue; (FIG. 11B) an NGS whole genome sequencing study with DNA extracted from normal human tissue (without whole exome enrichment through any methods);

[097] FIG. 12 shows detection of ultra-rare SNVs in libraries created from normal DNA spiked with sequentially diluted tumor DNA samples;

[098] FIGS. 13A-13N show the 298-gene panel real-time PCR parameters;

[099] FIG. 14 shows data yield from RNA probe-based DNA double strand capture WES sequencing; and

[0100]FIGS. 15A-15E show results of mutation and ultra-rare mutation detection by RNA probe-based DNA double strand capture NGS.

DETAILED DESCRIPTION OF THE INVENTION

[0101]In a sequencing technology such as NGS, the enrichment of nucleic acids from diluted samples or the capture of specific nucleic acid sequences from a sample comprising a complex pool of nucleic acids is often a crucial step, but can be challenging tasks in many cases. Enrichment for desired sequences can make assays feasible that would otherwise fall below detection limits, and can improve the performance of a genetic or genomic assay.

[0102]The present disclosure provides a method for enriching nucleic acid sequences from a biological sample, which substantially utilizes RNA probes that target both of two antiparallel strands of a duplex segment of a target sequence to be enriched in a sample.

[0103]FIG. 1A illustrates a flow chart of the method for enriching at least one target nucleic acid sequence from a biological sample according to some embodiments of the disclosure. As shown in FIG. 1A, the method comprises steps as set forth in S100-S400:

[0104]S100: Preparing at least one pair of RNA probe sets, each pair comprising a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a duplex segment in each of at least one target nucleic acid sequence, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with an immobilization portion.

[0105]Specifically, each pair of RNA probe sets comprises a first RNA probe set and a second RNA probe set, corresponding to one another, and each comprising one or more RNA probes. The one or more RNA probes in the first RNA probe set and the one or more RNA probes in the second RNA probe set are configured to respectively target two antiparallel strands of a duplex segment (i.e. double-stranded segment) of one of the at least one target nucleic acid sequence in the biological sample. Each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with, or comprises, an immobilization portion, configured to allow an immobilization by a solid support which will be described below in detail.

[0106]Herein, there is no limitation to the source of the biological sample, to the type of the at least one target nucleic acid sequence, or to the type of the duplex segment in each of the at least one target nucleic acid sequence.

[0107]The biological sample may be a tissue sample, from which at least one target nucleic acid sequence is obtained through a DNA or RNA purification protocol. The biological sample may be a cell-free DNA sample obtained from plasma, which contains the at least one target nucleic acid sequence in the sample. The biological sample may be derived from a treated sample, and contain, for example, a barcoded DNA library as disclosed in patent application PCT/US2018/016778, where each of the at least one target nucleic acid sequence contains a barcoded adaptor at one or both ends thereof. Other possibilities are possible as well.

[0108]The at least one target nucleic acid sequence in the biological sample can comprise one or more DNA molecules, one or more RNA molecules, one or more DNA-RNA hybrid molecules, or any of their combinations.

[0109]The duplex segment in each of the at least one target nucleic acid sequence may be formed by two separate DNA strands of a double-stranded DNA molecule, two separate RNA strands of a double-stranded RNA molecule, or one DNA strand and one RNA strand of a DNA-RNA hybrid molecule. The duplex segment may also be an intra-strand hairpin or alike formed within one single DNA strand or within one single RNA strand. In any of the aforementioned cases, a duplex segment substantially comprises two strand segments from one single strand (i.e. intra-strand duplex) or from two separate strands (i.e. inter-strand duplex), each having a sequence allowing a hybridization therebetween (e.g. having a sequence substantially complimentary to each other) and each running antiparallelly to each other to thereby form a double-stranded (i.e. duplex) structure.

[0110]Regardless of the type of a target nucleic acid sequence (DNA or RNA), or the type of the duplex segment therein (inter-strand or intra-strand), a first RNA probe set and a second RNA probe set can be configured to respectively target the two antiparallel strands of a duplex segment of each of the at least one target nucleic acid sequence in the biological sample, which together form a pair of RNA probe sets corresponding to the each of the at least one target nucleic acid sequence. Herein, throughout the disclosure, and also unless indicated otherwise, the two antiparallel strands of a duplex segment of each of the at least one target nucleic acid sequence in the biological sample are termed a plus strand and a minus strand of the each of the at least one target nucleic acid sequence.

[0111]In other words, the at least one RNA probe in the first RNA probe set and the at least one RNA probe in the second RNA probe set are respectively configured to target the plus strand and the minus strand, or alternatively the minus strand and the plus strand, of one of the at least one

target nucleic acid sequence. By means of the at least one pair of RNA probe sets where a first RNA probe set and a second RNA probe set in each pair respectively target two antiparallel strands of a duplex segment in each target nucleic acid sequence, each strand of the at least one target nucleic acid sequence can be captured or enriched from the biological sample. For a typical example, each target nucleic acid sequence can be a double-stranded DNA molecule, which has a plus strand and a minus strand that runs antiparallelly to form a duplex. By means of the at least one pair of RNA probe sets where a first RNA probe set and a second RNA probe set in each pair respectively target both of the two antiparallel strands (i.e. the plus strand and the minus strand) of each target nucleic acid sequence, each strand, including both the plus strand and the minus strand, of each target DNA molecule, can be captured or enriched from the biological sample.

[0112]This notably contrasts with a conventional approach where only one strand of each target nucleic acid sequence is captured since only one RNA probe or one RNA probe set targeting only one strand of each target nucleic acid sequence is utilized. More specifically, in conventional approach a valid capture is typically a 1st degree reaction between two complementary sequences that are respectively from a target nucleic acid sequence and a probe. However, in this present disclosure, a valid capture is a 2nd degree reaction between a target nucleic acid molecule having a duplex segment (e.g. a duplex double-stranded DNA molecule) and each of a pair of RNA probes (i.e. a first RNA probe and a second RNA probe) that respectively target the two antiparallel strands of the duplex segment, where the hybridization of one strand in the duplex segment of the target nucleic acid molecule with one of the pair of RNA probes can help expose the other strand of the duplex segment to thereby facilitate the hybridization of the other of the pair of RNA probes therewith. Therefore a higher capture efficiency can be realized, and such an effect has been observed in the experiment as detailed below.

[0113]Herein, the solid support can comprise at least one of a magnetic bead, a microfluidic channel, a filter, a resin bead, a nanosphere, a plastic surface, a microtiter plate, a glass surface, a slide, a membrane, a microfluidic channel, a chip, or a matrix, which is labelled, conjugated, or attached, with the coupling partner corresponding to the immobilization portion. The solid support can be part of an apparatus, such as a chip, a column, a tube, or a channel (such as a microfluidic channel in a microfluidic chip).

[0114]The immobilization portion labelled on each RNA probe can include a biotin moiety, and correspondingly the solid support can comprise comprises at least one of a magnetic bead, a filter, a resin bead, a nanosphere, a plastic surface, a microtiter plate, a glass surface, a slide, a membrane, a microfluidic channel, a chip, or a matrix, and the solid support can be labelled with at least one of streptavidin, avidin, or an anti-biotin antibody. As such, the RNA probes labelled with, or carrying, the biotin moiety can form a secure non-covalent binding with the solid support

conjugated with a biotin-coupling partner, such as streptavidin, avidin, or an anti-biotin antibody, which facilitates the capture of target nucleic acid sequences hybridized by the RNA probes. Other examples of the immobilization portion-coupling partner pair can include, but is not limited to, a carbohydrate-lectin pair, an antigen-antibody pair and a negative charged group-positive charged group static interacting pair.

[0115]In addition to the above embodiments where each RNA probe binds to the solid support through a non-covalent binding between the immobilization portion and the coupling partner pair, the secure coupling between each RNA probe and the solid support can be via a covalent connection (or cross-linking). As such, the immobilization portion and the coupling partner can respectively be one and another of a cross-linking pair. Examples of the cross-linking pair include an NHS ester-primary amine pair, a sulfhydryl-reactive chemical group pair (e.g. cysteines, or other sulfhydryls such as maleimides, haloacetyls, and pyridyl disulfides), an oxidized sugar-hydrazide pair, photoactivatable nitrophenyl azide's UV triggered addition reaction with double bonds leading to insertion into C-H and N-H sites or subsequent ring expansion to react with a nucleophile (e.g., primary amines), or carbodiimide activated carboxyl groups to amino groups (primary amines), etc. It is noted that the RNA probes can be conjugated on the solid support after synthesis, or can be synthesized directly on the solid support. Any method that could result in RNA probes to be linked on a solid support can be adopted. There is no limitation herein.

[0116]Herein it is noted that the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets can be configured to target a same portion, or a distinct portion, of the duplex segment in one of at least one target nucleic acid sequence.

[0117]According to some embodiments, a first RNA probe set and a second RNA probe set in a corresponding pair of RNA probe set are configured to target a same portion of the duplex segment of one target nucleic acid sequence, and as such, the RNA probe(s) in the first RNA probe set and the RNA probe(s) in the second RNA probe set have substantially complimentary sequences. As such, in certain embodiments where the RNA probes between the first RNA probe set and the second RNA probe set could potentially interact, or interfere, with one another in a single hybridization due to the unwanted hybridization therebetween, such as in cases where the RNA probes are labelled onto beads or matrix, contacting of the biological sample containing target nucleic acid sequences by the first RNA probe set and the second RNA probe set is preferably performed in a sequential manner. This can be realized by sequentially contacting the biological sample containing target nucleic acid sequences with magnetic beads respectively carrying one and another of the first RNA probe set and the second RNA probe set, which will be described in detail below. This can also be realized by allowing the biological sample to flow sequentially

through two different layers (10A and 20A) of a column which comprise a matrix conjugated respectively with the first RNA probe set and the second RNA probe set, as illustrated in FIG. 1D.

[0118]It is noted, however, in some other embodiments where the RNA probes are separately conjugated onto a solid support which has no or little interactions due to, for example, positional compartmentation on the surface of the solid support, such as in cases where RNA probes from the first RNA probe set and RNA probes from the second RNA probe set are separately conjugated onto different regions of a microfluidic channel (FIG. 1E) or different regions of a chip (FIG. 1F), or are mixedly arranged on a common region of a microfluidic channel (FIG. 1G) or a chip (FIG. 1H) with a relatively large distance between each RNA probe from the first RNA probe set and each RNA probe from the second RNA probe set (thereby substantially preventing an interaction therebetween), there is no need for a sequential contact between the biological sample with the first RNA probe set and the second RNA probe set, and the biological sample can be applied to contact both the first RNA probe set and the second RNA probe set simultaneously. Yet depending on certain practical needs, the biological sample can still be allowed to sequentially contact RNA probes in the first RNA probe set and the second RNA probe set by flowing sequentially through two different regions of a microfluidic channel or a chip corresponding respectively to the first RNA probe set and the second RNA probe set as illustrated in FIG. 1E and FIG. 1F.

[0119]It is further noted that the employment of a column or a microfluidic channel further provides a convenience for repeated contacts of the biological sample with RNA probes, either by arranging the first RNA probe set and the second RNA probe set in series (as illustrated in FIG. 1I), by arranging the biological sample to recirculate (as illustrated in FIG. 1J), or by combination of these two approaches. The details for each of these above different embodiments of the method for capturing target nucleic acid sequences from the biological sample by means of the at least one pair of RNA probe sets will be provided below.

[0120]As such, the above FIGS. 1D-1J respectively illustrate several different embodiments where each pair of RNA probe sets arranged on an inner surface of an apparatus (microfluidic channel or a chip) to allow the capture of target nucleic acid sequences from the biological sample. It is noted that these embodiments are illustrating only, and there can be other apparatuses as well.

[0121]According to some other embodiments, a first RNA probe set and a second RNA probe set in a corresponding pair of RNA probe set are configured to target a different portion of the duplex segment of one target nucleic acid sequence. In other words, a first RNA probe set and a second RNA probe set in a corresponding pair of RNA probe set are respectively configured to target a first portion in a first strand and a second portion in a second strand of the one target nucleic acid sequence, wherein the first strand and the second strand are the two antiparallel strands of the duplex segment, and the first portion and the second portion are two different portions of the

duplex segment of the one target nucleic acid sequence. As such, the RNA probe(s) in the first RNA probe set and the RNA probe(s) in the second RNA probe set have substantially no complimentary sequences. As such, the biological sample containing target nucleic acid sequences can be applied to contact both the first RNA probe set and the second RNA probe set simultaneously, regardless of what type of solid support is utilized for coupling with the RNA probes.

[0122]In order to prepare the at least one pair of RNA probe sets in step S100, any one of the first RNA probe set and the second RNA probe set in each pair can be prepared by a manner of direct chemical synthesis or by a manner of transcription, or by a various combination of these approaches, and can be labelled with the immobilization portion during or after the synthesis/transcription process. These different embodiments will be described below in detail.

[0123]According to some embodiments, each RNA probe in any of the first RNA probe set and the second RNA probe set corresponding to each of the at least one pair of RNA probes can be synthesized directly by chemical reactions (i.e. the manner of direct chemical synthesis). Depending on different situations, the immobilization portion can be labelled to each RNA probe during the synthesis process or after the synthesis process, or each RNA probe can be directly synthesized from a solid support that is covalently connected with one end of each RNA probe through the immobilization portion.

[0124]According to some other embodiments, each RNA probe in any of the first RNA probe set and the second RNA probe set for each pair of the RNA probe sets can be respectively and separately obtained through a transcription reaction (i.e. the manner of transcription) over a pair of DNA vectors corresponding thereto, and the immobilization portion can be labelled to each RNA probe during the transcription process or after the transcription process.

[0125]Regarding the transcription process, each of the pair of DNA vectors can comprise a DNA template and a promoter (i.e. transcription promoter). The DNA template can comprise a sequence whose transcription gives rise to an RNA molecule corresponding to either of the two antiparallel strands of the duplex segment of the one target nucleic acid sequence. The promoter is configured to be recognized by an RNA polymerase to thereby allow the transcription reaction to occur, and can be at an upstream of, a downstream of, or within a target DNA sequence in the DNA vector.

[0126]Compared with the manner of direct chemical synthesis, this transcription-based approach to obtain RNA probes is relatively more cost-effective, and allows for the production of a relatively larger amount of the RNA probes.

[0127]In the aforementioned embodiments where the at least one RNA probe in each of the at least one pair of RNA probe sets is prepared by transcription, step S100 can comprise:

[0128]S110: preparing a plurality of DNA vectors, comprising at least one pair of DNA vectors, each pair comprising a first DNA vector and a second DNA vector configured to respectively allow a separate transcription of a first RNA molecule and a second RNA molecule targeting respectively two antiparallel strands of a duplex segment of each of the at least one target nucleic acid sequence.

[0129]According to some embodiments as illustrated in FIG. 3A and FIG. 3B, the first DNA vector 001A and the second DNA vector 001B in each of the at least one pair of DNA vectors can respectively comprise a first DNA template 100A at a downstream of a first transcription promoter 200A and a second DNA template 100B at a downstream of a second transcription promoter 200B. Each of the first DNA template 100A and the second DNA template 100B is substantially a double-stranded DNA segment (indicated by the box with dotted lines in the two figures), configured to allow transcription of RNA molecules using one strand thereof as a template under action of the transcription promoter in a transcription reaction.

[0130]Specifically, the first DNA vector 001A comprises a first transcription promoter 200A at a 5' end (i.e. upstream) of a strand of the double-stranded first DNA template 100A that corresponds to a plus strand of the one target nucleic acid sequence (as indicated by the "+" sign), and thus is configured to allow the transcription of a first RNA molecule complementary to a minus strand of the one target nucleic acid sequence (as indicated by the "-" sign) as a transcription template. As such, the RNA molecules produced by the first DNA vector 001A can specifically target (i.e. hybridize or bind or anneal with) the minus strand of the one target nucleic acid sequence.

[0131]The second DNA vector 001B comprises a second transcription promoter 200B at a 5' end (i.e. upstream) of a strand of the double-stranded first DNA template 100B that corresponds to a minus strand of the one target nucleic acid sequence (as indicated by the "-" sign), and thus is configured to allow the transcription of a second RNA molecule complementary to a plus strand of the one target nucleic acid sequence (as indicated by the "+" sign) as a transcription template. As such, the RNA molecules produced by the second DNA vector 001B can specifically target (i.e. hybridize or bind or anneal with) the plus strand of the one target nucleic acid sequence.

[0132]As such, in each corresponding pair of DNA vectors, the first DNA vector 001A and the second DNA vector 001b respectively allow transcription of a first RNA molecule that specifically target the minus strand and a second RNA molecule that specifically target the plus strand of the one target nucleic acid sequence.

[0133]In each of the at least one pair of DNA vectors (i.e. the first DNA vector and the second DNA vector) as described above, the first promoter 200A and the second promoter 200B can be substantially same or different, and different pairs of DNA vectors can have same or different promoters. These promoters can include a T3 promoter, a T7 promoter, a SP6 promoter, or a

species-specific or tissue-specific promoter. Herein, the double-stranded DNA segment in each of the first DNA vector and the second DNA vector that correspond to each target nucleic acid sequence in the sample can comprise a genomic DNA fragment, a gene coding sequence (CDS) or such sequences in an existing construct (such as commercially available gene expression constructs), or can be derived from reverse-transcription of an RNA sequence, such as an mRNA sequence, or can comprise segments that are artificially synthesized or assembled.

[0134]It is noted that the above embodiments as shown in FIG. 3A and FIG. 3B, where each of the first DNA vector 001A and the second DNA vector 001B in each pair of DNA vectors is substantially a double-stranded DNA vector, and each of the first DNA template 100A and the second DNA template 100B exists as a double-stranded DNA segment and is at a downstream of a transcription promoter, serves as an illustrating example only and does not impose a limitation to the scope of the present disclosure. Other embodiments are also possible.

[0135]For example, according to some other embodiments as illustrated in FIG. 3C and FIG. 3D, each of the first DNA vector 001A' and the second DNA vector 001B' can be a single-stranded DNA vector (such as a phagemid or phasmid, or a vector containing a cDNA molecule produced from a reverse-transcription reaction from an RNA sequence). The first DNA vector 001A' comprises a first promoter 200A' at a 3' end of a first DNA template 100A' which corresponds to a plus strand of a duplex segment of one target nucleic acid sequence (as indicated by the "+" sign). The second DNA vector 001B' comprises a second promoter 200B' at a 3' end of a second DNA template 100A' which corresponds to a minus strand of the duplex segment of the one target nucleic acid sequence (as indicated by the "-" sign). Thus transcription of the first DNA vector 001A' and the second DNA vector 001B' can respectively produce RNA molecules that target the plus strand and the minus strand of the duplex segment of the one target nucleic acid sequence.

[0136]According to yet some other embodiments of the disclosure, the first DNA vector and the second DNA vector in each pair of DNA vectors can be of a different type. For example, the first DNA vector can be a double-stranded DNA vector whereas the second DNA vector can be a single-stranded DNA vector, and it is further configured such that transcription of the first DNA vector and the second DNA vector can respectively produce RNA molecules that target the two antiparallel strands (e.g. plus strand/minus strand or minus strand/plus strand) of the duplex segment of the one target nucleic acid sequence.

[0137]In any of the above embodiments of the DNA vectors, a transcription promoter is disposed at an upstream of a target DNA sequence. The relative position of the transcription promoter is not limited to the upstream of a target DNA sequence, and can be within, or at a downstream of a target DNA sequence as well, depending on specific cases.

[0138]As such, there is no limitation to the nature and type of the first DNA vector and the second DNA vector in each pair of DNA vectors, as long as the respective transcription reaction over the first DNA vector and the second DNA vector can give rise to the first RNA molecule and the second RNA molecule targeting respectively two antiparallel strands of a duplex segment of each target nucleic acid sequence.

[0139]After sub-step S110, there are several embodiments of step 100 of the method where the first RNA probe set and the second RNA probe set in each of at least one pair of RNA probe sets are obtained by transcription reactions over the first DNA vector and the second DNA vector in each of the plurality of DNA vectors, and each probe in any of the first RNA probe set and the second RNA probe set is labelled with, or carries, the immobilization portion.

[0140]According to some preferred embodiments as illustrated in FIG. 2A, the immobilization portion is labelled during transcription process. As such, after sub-step S110, step S100 of the method further comprises:

[0141]S120: performing a transcription reaction separately over each DNA vector to generate a plurality of RNA molecules corresponding to the each DNA vector and labelled with the immobilization portion.

[0142]Herein the transcription reaction can be a regular *in vitro* transcription reaction, and involves an RNA polymerase and four nucleoside triphosphates (ATP, UTP, GTP, CTP, collectively as NTPs). The RNA polymerase can recognize a transcription promoter (i.e. the first DNA transcription promoter or the second DNA transcription promoter) of a DNA template (i.e. the first DNA template or the second DNA template corresponding respectively to the two antiparallel strands of each target nucleic acid sequence), to thereby allow the transcription of RNA molecules having sequences targeting/complementary to the plus strand/minus strand of each nucleic acid sequences (i.e. the first RNA molecule and the second RNA molecule). The RNA polymerase can be any enzyme that triggers DNA-dependent RNA polymerization, and can be, for example a T7 RNA polymerase.

[0143]The RNA molecules can be purified from the reaction using an RNA purification protocol, and the DNA molecules in the reaction can be eliminated by applying enzymes that can degrade DNA molecules, such as DNase. Such enzymes need to be completely removed from the RNA molecules if the targeted nucleic acid sequences being captured include DNA molecules, but this removal step can be skipped if the targeted nucleic acid sequences are nucleic acids that are not susceptible to DNase-induced damage, such as RNA molecules. Other approaches to remove DNA molecules or to separate RNA molecules from the DNA molecules are also possible. For example, the DNA vectors are pre-immobilized to solid support that could be readily removed from the reaction after the transcription reaction is finished without removing the transcribed RNA.

[0144]The RNA molecules can be directly extracted from the system through applying the interaction between the coupling pairs, such as using streptavidin beads to extract RNA molecules that are biotinylated. The transcription reaction can also be an *in vivo* transcription reaction, and can be synthesized in an organism, such as a bacterium (e.g. *E. coli*), a fungus (e.g. yeast), a mammalian cell line, etc. The RNA molecules can be extracted based on a regular RNA extraction protocol, and can be left in the system to perform real-time labeling or capturing of its target nucleic acid sequences.

[0145]Herein the sub-step S120 can directly generate a plurality of RNA molecules, each labelled with the immobilization portion. Specifically, the immobilization portion labelled on each of the RNA molecules, which is further labelled on the RNA probes derived from the RNA molecules, can facilitate the immobilization (or capturing) of the at least one target nucleic acid sequence on a solid support in step S300 (see below), due to a stable coupling between the immobilization portion and the solid support.

[0146]The stable coupling can be mediated by a secure and stable non-covalent binding, or by a covalent connection (i.e. cross-linking) between the immobilization portion and a corresponding coupling partner conjugated onto the solid support. Depending on different types of a solid support, and different types of coupling between each RNA probe labelled with the immobilization portion and the solid support conjugated with the coupling partner, there can be different embodiments. For example, in some preferred embodiments, the immobilization portion is a biotin moiety, and the coupling partner can be a streptavidin, avidin, or an anti-biotin antibody, which is attached onto, or conjugated with a solid support such as magnetic beads, as illustrated in FIG. 7A and FIG. 7B. In another example, each RNA probe is conjugated onto an inner surface of a microfluidic channel or a chip via a covalent connection, as illustrated in FIGS. 1E-1H.

[0147]As such, through the stable coupling between each immobilization portion on each RNA probe that derives from the RNA molecules obtained thereby and the coupling partner corresponding thereto that is conjugated onto the solid support, the at least one RNA probe can be immobilized by the solid support, in turn facilitating subsequent enrichment, isolation, and purification of target nucleic acid sequences.

[0148]It is noted that in order to introduce convenience for other applications (e.g. visualization), each RNA probe can be further labelled with other functional portion(s) in addition to the immobilization portion. Examples include a dye, a fluorophore group, or a chemical group, etc.

[0149]In some specific embodiments that are preferred, the transcription reaction in sub-step S120 can include addition of a mix of UTPs comprising biotin-labelled UTPs (i.e. herein the biotin moiety serves as an immobilization portion in the at least one functional portion) and non-biotin-labelled UTPs, wherein the biotin-labelled UTPs have a relative molar percentage of ~2%-100%

of the total UTPs. In other words, among all the UTPs added in the reaction, the biotin-labelled UTPs can take a molar percentage between about 2% and about 100% (i.e. all the UTPs added are biotin-labelled UTPs). As such, after transcription, each of the plurality of RNA molecules with a length of about 200 nt can be labelled with biotin in some or all of its U residue.

[0150]S130: pooling the plurality of RNA molecules to obtain at least two RNA pools, each comprising at least one RNA molecule, configured such that a corresponding pair of RNA molecules respectively targeting two antiparallel strands of a duplex segment of one target nucleic acid sequence are not in a same RNA pool.

[0151]After transcription in sub-step S120, the plurality of labelled RNA molecules can be pooled into at least two RNA pools (or called RNA libraries), each comprising at least one RNA molecule. It is configured that a corresponding pair of RNA molecules (i.e. the first RNA molecule and the second RNA molecule that respectively targets two antiparallel strands of a duplex segment of one target nucleic acid sequence) are not in a same RNA pool to thereby avoid an interference in subsequent steps of hybridization and enrichment of target nucleic acid sequences.

[0152]In one specific embodiment, a plurality of RNA molecules are pooled into two RNA pools (a first RNA pool and a second RNA pool), and the first RNA pool and the second RNA pool respectively includes RNA molecules that each specifically target one, but not both, of the plus strand and the minus strand of each target nucleic acid sequence.

[0153]It is noted that the sub-step S130 can be modified depending on different needs. For example, when doing pooling, each of the plurality of RNA molecules can have a same ratio, or can have a different ratio in order to ensure a highly efficient capture/enrichment of the different sequence fragments of the at least one target nucleic acid sequences in the biological sample. For example, in cases where some specific target nucleic acid sequences are difficult to capture (which can be based on previous knowledge, or can be known by a preliminary experiment applying all of the steps S100-S400, and identifying some target nucleic acid sequences that are unsatisfactorily captured), the abundance for the pair of RNA probe sets, or the abundance for the RNA probes targeting one specific strand corresponding to these specific target nucleic acid sequences can be increased (e.g. by ~10 fold or higher, or by ~1.5 fold; there is no limitation herein) in the sub-step S130 to thereby increase the efficiency for capture.

[0154]It is further noted that in addition to the above optimization (i.e. adjustment of ratios) in sub-step S130, a different segment in one target sequence can be selected for generation RNA probes in order for an optimized capture if RNA probes generated from one particular target segment are not able to offer expected capturing efficiency.

[0155]S140: performing a fragmentation reaction to each of the at least two RNA pools respectively to thereby obtain the at least one pair of RNA probe sets.

[0156]In order to facilitate the efficiency in the subsequent steps S200 and S300, the relatively long RNA molecules in each RNA pool can be preferably fragmented into relatively shorter fragments of ~100-150 nt, which can be done, for example, by enzymatic reactions or sonication. Conditions for the enzymatic reactions or sonication reactions for nucleic acids are well-known in the field and can be used as is appropriate and convenient. After fragmentation, the at least one RNA probe in each of the at least two RNA probe pools can have a length of at least 2 nt, preferably 100-150 nt.

[0157]Herein it is noted that it is possible to reverse the order of sub-steps S130 and S140. For example, after sub-step S120 to obtain the plurality of RNA molecules labelled with at least one functional portion, the plurality of RNA molecules can be fragmented (i.e. S140) before pooling (i.e. S130) to thereby obtain the at least one pair of RNA probe sets.

[0158]Besides the embodiments as mentioned above, according to some other embodiments of the disclosure, a plurality of RNA molecules can alternatively be transcribed without labeling, and the at least one functional portion can be labelled thereafter. FIG. 2B illustrates one embodiment of step S100. As shown in the figure, after sub-step S110, step 100 further comprises:

[0159]S120': performing a transcription reaction separately over each DNA vector to obtain a plurality of RNA molecules corresponding thereto.

[0160]Herein the sub-step S120' is substantially identical to the aforementioned sub-step S120, except that no functional portion-labelled NTPs (e.g. biotin-labelled UTP), is added to the transcription reaction.

[0161]S130': pooling the plurality of RNA molecules to obtain at least two RNA pools, each comprising at least one RNA molecule, configured such that a corresponding pair of RNA molecules respectively targeting two antiparallel strands of a duplex segment of one target nucleic acid sequence are not in a same RNA molecule pool.

[0162]Herein the sub-step S130' is substantially identical to the aforementioned sub-step S130, and the technical details are thus skipped herein.

[0163]S140': performing a fragmentation reaction to each of the at least two RNA pools.

[0164]Herein the sub-step S140' is substantially identical to the aforementioned sub-step S140, and the technical details are thus skipped herein.

[0165]S150': performing a labelling reaction to each of the at least two RNA pools to thereby obtain the at least one pair of RNA probe sets.

[0166]Herein through S150', the immobilization portion can be labelled onto each RNA probe in any of the at least two RNA pools. According to some preferred embodiment, the immobilization portion is a biotin moiety, and can be labelled at a 5' end, and/or a 3' end, and/or an intra-strand nucleic acid residue of each RNA probe in this sub-step. It is noted that other functional portion(s)

may also be labelled onto each RNA probe in any of the at least two RNA pools. The technical details have been provided above and are skipped herein.

[0167]Herein it is noted that it is possible to alter the order of sub-steps S130'-S150'. For example, after sub-step S120' to obtain the plurality of RNA molecules, the plurality of RNA molecules can be fragmented (i.e. S140') and labelled (i.e. S150') before pooling (i.e. S130) to thereby obtain the at least one pair of RNA probe sets. Alternatively, after sub-step S120', the plurality of RNA molecules can be labelled (i.e. S150') before fragmentation (i.e. S140') and pooling (i.e. S130) to thereby obtain the at least one pair of RNA probe sets. There are no limitations herein.

[0168]In the following, two specific embodiments are provided to respectively illustrate two different processes of preparing one pair of RNA probe sets (i.e. a first RNA probe set and a second RNA probe set), which respectively target the two antiparallel strands (i.e. the plus strand and the minus strand) of a double-stranded segment of one target nucleic acid sequence.

[0169]In one embodiment as illustrated in FIG. 4A, the pair of RNA probe sets are prepared and biotin-labelled during a same transcription and labelling process. Specifically, the two transcription reactions are respectively performed over the first DNA vector 001A and the second DNA vector 001B as shown in FIGS. 3A and 3B to thereby separately generate a plurality of first RNA molecules 300A and a plurality of second RNA molecules 300B, configured to respectively target the minus strand and the plus strand of the one target nucleic acid sequence to be enriched or captured. At the same time, the plurality of first RNA molecules 300A and the plurality of second RNA molecules 300A' are each also labelled with one or more biotin moieties (shown as a filled dot linked with each RNA probe in FIG. 4A). This can be done by providing to the transcription reaction a mixture of UTPs comprising biotinylated UTPs (i.e. biotin-labelled UTPs) and regular UTPs (i.e. non-biotin-labelled UTPs), wherein the biotin-labelled UTPs can have a relative molar percentage of ~2%-100% of the total amount of UTPs.

[0170]In one specific example, a mixture of biotin-labelled UTPs and regular UTPs having a ratio of 1/3 (i.e. the biotin-labelled UTPs have a relative molar percentage of 25% of the total amount of UTPs) can be added to the transcription reaction. As such, about ¼ of U residues in the whole RNA sequence can be found to be labelled with a biotin moiety. Then after the process of transcription and labelling, the plurality of first RNA molecules 300A and the plurality of second RNA molecules 300A' are separately fragmented into the first RNA probe set 500A and the second RNA probe set 500B, which can preferably be ~100-150 nt in length.

[0171]In another embodiment as illustrated in FIG. 4B, the pair of RNA probe sets are biotin-labelled after the transcription process. Specifically, the two transcription reactions can be respectively performed over the first DNA vector 001A the second DNA vector 001B as shown in FIGS. 3A and 3B to thereby separately generate a plurality of first RNA molecules 300A' and a

plurality of second RNA molecules 300B', configured to respectively target the minus strand and the plus strand of the one target nucleic acid sequence to be enriched or to be captured. After transcription, the plurality of first RNA molecules 300A' and second RNA molecules 300B' are separately fragmented into the first RNA probe set 400A' and the second RNA probe set 400B', which are then respectively labelled with the biotin moiety (shown as a filled dot linked with one end of each RNA probe) to thereby obtain biotin-labelled first RNA probe set 500A' and second RNA probe set 500B'.

[0172]Herein labelling of the biotin moiety at a 5' end of each RNA probe is for illustrating purpose only, which can be at a 3'-end or an intra-strand segment thereof as well. According to some specific embodiment, in each of the ligation reactions, a 5' phosphate terminus of a biotin-labeled nucleotide can be ligated to a 3' hydroxyl terminus of each RNA probe in each of the at least one pair of RNA probe sets, which can be catalyzed by means of an RNA ligase. The RNA ligase can comprise at least one of T4 RNA ligase or CircLigase RNA ligase.

[0173]According to some other embodiment, the immobilization portion can be labelled in an intra-strand segment of an RNA probe, in a form of an RNA adduct (or a chemical addon) by means of a technology known to the field, whose description is skipped herein.

[0174]As such, after the above mentioned sub-steps S110-S140 (illustrated in FIG. 2A) or sub-steps S110-S150' (illustrated in FIG. 2B), at least two RNA probe pools are generated, each comprising at least one immobilization portion-labelled RNA probe. Together, the at least two RNA probe pools substantially includes at least one pair of RNA probe sets, each pair comprising a first labelled RNA probe set and a second labelled RNA probe set, respectively targeting two antiparallel strands of a duplex segment of one of the at least one target nucleic acid sequence.

[0175]After preparation of the at least one pair of RNA probe sets in step S100, the at least one pair of RNA probe sets can be used to contact with target nucleic acid sequences for hybridization before capture and enrichment of the at least one target nucleic acid sequence using the at least one pair of RNA probe sets as baits.

[0176]S200: Contacting the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence to allow hybridization of each RNA probe in the at least one pair of RNA probe sets to a corresponding strand of each target nucleic acid sequence.

[0177]Depending on whether the first RNA probe set and the second RNA probe set corresponding to a pair of RNA probe sets interfere with each other, such as having complimentary sequences to thereby be able to form an unwanted duplex structure when mixed together, S200 can be carried out in different manners.

[0178]According to some embodiments, the first RNA probe set and the second RNA probe set corresponding to a pair of RNA probe sets do not have substantially complimentary sequence (such as in one above-mentioned embodiment where the first RNA probe set and the second RNA probe set corresponding to a pair of RNA probe sets target two different portions of the duplex segment of the target nucleic acid sequence), or are conjugated to a solid support that do not readily interfere with each other (such as in one above-mentioned embodiment where the first RNA probe set and the second RNA probe set corresponding to a pair of RNA probe sets are conjugated to different regions of a microfluidic channel on a microfluidic chip), step S200 can comprise:

[0179] S200a: Contacting both the first RNA probe set and the second RNA probe set with the at least one target nucleic acid sequence in a single hybridization reaction.

[0180]According to some other embodiments, the first RNA probe set and the second RNA probe set in each pair have substantially complimentary sequences and thus can interfere with each other in a single hybridization reaction with the target nucleic acid sequences, as such, different embodiments of the method can be employed.

[0181]According to some embodiments, the first RNA probe set and the second RNA probe set in each pair can be configured to hybridize with the at least one target nucleic acid sequence in a sequential manner. Herein the "sequential manner" is referred to as a manner where the first RNA probe set and the second RNA probe set in each pair are allowed to contact with, and thereby to hybridize with, the at least one target nucleic acid sequence in the biological sample one after another. There is no limitation to the specific order: for example, the first RNA probe set is added to the hybridization reaction first, followed by the second RNA probe set, or alternatively, the second RNA probe set is added to the hybridization reaction first, followed by the first RNA probe set.

[0182]Specifically, according to some embodiments where the first RNA probe set and the second RNA probe set in each pair of RNA probe sets is labelled onto magnetic beads, the step S200 includes the following sub-steps, as illustrated in FIG. 5A:

[0183]S210: Contacting one of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a first hybridization reaction; and

[0184]S220: Contacting another of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a second hybridization reaction.

[0185]Alternatively, the first RNA probe set and the second RNA probe set in each pair can be configured to hybridize with the at least one target nucleic acid sequence in two separate hybridization reactions, which are then combined to react in another hybridization reaction. This

is substantially a special situation for the sequential approach, and specifically, as shown in FIG. 5B, step S200 includes the following sub-steps:

[0186]S210': Separately contacting the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a third hybridization reaction and a fourth hybridization reaction, respectively; and

[0187]S220': Combining the third hybridization reaction and the fourth hybridization reaction to allow the hybridization to proceed in a fifth hybridization reaction.

[0188]Herein it is noted that "first", "second", "third", "fourth", and "fifth" are intended to refer to a different hybridization reaction and indicates or suggests no actual order in the reactions. It is further noted that any of the aforementioned hybridization reactions (i.e. the first, second, third, fourth, or fifth hybridization reaction) can occur at a temperature that substantially allows each RNA probe to hybridize efficiently with a corresponding strand of each target nucleic sequence.

[0189]For example, in some embodiments, each RNA probe has a length of ~100 - ~150 nt, and the temperature a hybridization reaction can be at a range of ~40-90°C, preferably at ~62-70°C, and more preferably at ~67.5°C. Such a hybridzation temperature allows a high efficiency and a balanced specificity for the capture and enrichment of target nucleic acid sequences (see below). However, it is noted that these embodiments serve as illsutrating examples only, and do not limit the scope of the disclosure. Other hybridization conditions, for example, RNA probes of a different length, at different abundances, a hybridization temperature, etc., can also be applied depending on specific needs.

[0190]Furthermore, the incubation time for each hybridization reaction can vary, depending on different configurations. According to some embodiments where the hybridization reaction occurs between RNA probes carrying the biotin moiety and the target nucleic acid sequences in the sample, the incubation time can be 6-24 hours, and preferably 12 hours, to ensure an efficient hybridization of each RNA probe to its corresponding strand of the target nucleic acid targets. According to some other embodiments where the hybridization reaction occurs between RNA probes conjugated onto an inside surface microfluidic channel and the target nucleic acid sequences in the sample, the incubation time can be several seconds to several hours, depending on the temperature and the pressure for the reaction. There are no limitations herein.

[0191]By sequentially adding the two RNA probe sets (i.e. the first RNA probe set and the second RNA probe set) in each pair in the hybridization reaction as described in the above embodiments, the unwanted formation of probe-probe hybridizations between themselves if simultaneously added can be effectively reduced. In addition, the sequential approach as described above has been proved to be more efficient in capturing target nucleic acid sequences. As has been demonstrated in an experiment that is described below, the capturing of a target DNA sequence

utilizing a pair of complementary RNA probes that target both strands of the target DNA sequence can achieve an over 3-fold increase in the capture efficiency compared to the approach utilizing only one RNA probe that targets one single strand of the same target DNA sequence, as illustrated in FIG. 8A which will be described in detail below.

[0192]It is noted that in the above mentioned two approaches as illustrated in FIGS. 5A and 5B, the first RNA probe set and the second RNA probe set in each pair, once allowed to contact with the biological sample containing the at least one target nucleic acid sequence, are not removed from the reaction. Other embodiments are also possible.

[0193]For example, after each of the first RNA probe set and the second RNA probe set in each pair has contacted with the biological sample containing the at least one target nucleic acid sequence, each of the first RNA probe set and the second RNA probe set can be separated from the biological sample, allowing a capture of a portion of the at least one target nucleic acid sequence by the RNA probes in each of the first RNA probe set and the second RNA probe set, and the biological sample can be allowed to contact with each of the first RNA probe set and the second RNA probe set in each pair again.

[0194]It is noted that these embodiments actually allow the biological sample containing the at least one target nucleic acid sequence to repeatedly contact with the first RNA probe set and the second RNA probe set in each pair of RNA probe sets corresponding thereto. As such, the capture of the at least one target nucleic acid sequence from the biological sample can have a relatively higher efficiency after multiple rounds of sequential contact.

[0195]As such, the step S200 can include:

[0196]S211: Sequentially contacting the at least one target nucleic acid sequence with one and another of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets; and

[0197]optionally S212: Repeating S211 for at least one more time.

[0198]In one specific embodiment, the first RNA probe set and the second RNA probe set in each corresponding pair are respectively labelled onto magnetic beads, and as such, step S200 specifically comprises:

[0199]S211a: Contacting one of the first RNA probe set and the second RNA probe set with the biological sample containing the at least one target nucleic acid sequence;

[0200]S211b: Removing the one of the first RNA probe set and the second RNA probe set from the biological sample;

[0201]S211c: Contacting another of the first RNA probe set and the second RNA probe set with the biological sample; and

[0202]S211d: Removing the another of the first RNA probe set and the second RNA probe set from the biological sample.

[0203]These sub-steps S211a-S211d can be performed for only one time, or optionally can be repeated for at least one more time (i.e. S212).

[0204]In another specific embodiment, the first RNA probe set and the second RNA probe set in each corresponding pair are respectively conjugated onto the solid support (e.g. a matrix or a surface), which are respectively arranged at two different regions of an apparatus, such as a column or a microfluidic channel of a microfluidic chip. In one example as illustrated in FIG. 1D, a matrix conjugated with the first RNA probe set and a matrix conjugated with the second RNA probe set can be arranged at two different layers (10A and 10B) of a column. In another example as illustrated in FIG. 1E or FIG. 1F, the first RNA probe set and the second RNA probe set in each corresponding pair can be respectively immobilized onto two different regions of a microfluidic channel or chip along the direction of flow.

[0205]The apparatus (i.e. the column or the microfluidic channel) can be further configured to allow the biological sample to flow sequentially through the two different regions of the apparatus to thereby allow the biological sample containing the at least one target nucleic acid sequence to sequentially contact with the first RNA probe set and the second RNA probe set in each pair, which further supports repeated contact/separation to thereby increase the capture efficiency.

[0206]As such, step S200 can comprise:

[0207]S211a': Allowing the biological sample containing the at least one target nucleic acid sequence to sequentially flow through the two different regions of the apparatus to thereby allow a sequential contact thereof with, one and another of the first RNA probe set and the second RNA probe set; and

[0208]Optionally S212': Repeating S211a' for at least one more time.

[0209]It is noted that in the embodiment having S211a' and S212' where a column or a microfluidic channel is employed, it is possible that the repeating step (i.e. S212') is realized by arranging more than one pair of the first RNA probe set and the second RNA probe set in series on the column or the microfluidic channel along a direction of flow (illustrated in FIG. 1F), by arranging the sample to recirculate (i.e. by arranging the sample to flow into the column and the microfluidic channel again after flow out for more than one rounds, as illustrated in FIG. 1G), or by a combination of these two above approaches as well.

[0210]According to some embodiments of the disclosure, one or more target nucleic acids sequence in the biological sample may each be present in a polynucleotide which also contains at least one not-desired-to-be-captured sequences (termed "un-targeted sequence" hereafter). For example, in one specific embodiment as illustrated in FIG. 6A, the at least one target nucleic acid sequence is

in a DNA library, and each target nucleic acid sequence is flanked by a pair of adaptor sequences, such as PE sequencing adapters having a length of ~70 bp, which are substantially the un-targeted sequences. The presence of the at least one un-targeted sequence (e.g. one or two adaptor sequences) may potentially interfere with the enrichment or capture of the at least one target nucleic acid sequence by the RNA probes.

[0211]As such, in order to effectively eliminate the interference of the at least one un-targeted sequence, according to some embodiments of the disclosure, prior to step S200, the method further comprises:

[0212]S199: contacting at least one blocking oligo with the at least one target nucleic acid sequence, such that the at least one blocking oligo respectively hybridizes with, and thereby blocks, at least one strand of each of the at least one un-targeted sequence.

[0213]In one example, each target nucleic acid sequence is a double-stranded DNA sequence flanked by a pair of adaptor sequences (i.e. a first adaptor sequence 600A and a second adaptor sequence 600B as illustrated in FIG. 6A). As such, a set of blocking oligos comprising a first blocking oligo specifically targeting one strand of the first adaptor sequence 600A and a second blocking oligo specifically targeting one strand of the second adaptor sequence 600B, can be utilized in sub-step S201 to facilitate the hybridization of corresponding RNA probes with the each target nucleic acid sequence without the interference from the un-targeted sequences (i.e. the flanking adaptor sequences).

[0214]Herein, it is noted that there can be multiple arrangements for the set of blocking oligos. For example, the set of blocking oligos can consist of two blocking oligos, and can have a blocking oligo pair of (611 and 621), (611 and 622), (612 and 621), or (612 and 622), as long as the two blocking oligos target the two adaptor sequences 600A and 600B respectively, as shown in FIG. 6A. Alternatively, the set of blocking oligos can consist of three blocking oligos, having a combination of (611, 612 and 621), (611, 612 and 622), (611, 621 and 621), or (612, 621 and 622). Alternatively, the set of blocking oligos can consist of four blocking oligos (611, 612, 621 and 622), which substantially form two pairs of blocking oligos, 611/612 and 621/622, each pair corresponding to two strands of the first adaptor sequence 600A and two strands of the second adaptor sequence 600B, respectively, as illustrated in FIG. 6A.

[0215]It is noted that in order to avoid the unwanted annealing, hybridization, or binding, between two blocking oligos that target the two complimentary strands in any of the two adaptor sequences (600A or 600B), the two oligos can be added for hybridization and blocking in a sequential manner, or in a separate-and-combining manner, just as the addition of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets as illustrated in FIG. 5A or FIG. 5B.

[0216]Multiple embodiments for the at least one blocking olio are possible. For example, each of the at least one blocking oligo employed in S199 can be a single-stranded DNA oligo, or a single-stranded RNA oligo, which has a length of at least 2 nt and can be obtained based on a conventional technology known by people of ordinary skills in the field.

[0217]According to some embodiments, the at least one blocking oligo can comprise one or more blocking oligo sets, configured such that each blocking oligo set comprises one or more oligos which specifically target one of the two antiparallel strands of each of the at least one un-targeted sequence in a target nucleic acid sequence.

[0218]In one illustrating example of the at least one blocking oligo, it can be configured such that two blocking oligos (or two blocking oligo sets) respectively target two different portions of a duplex segment of each of the one or two adaptor sequences in a target nucleic acid sequence (i.e. a first portion in a first strand and a second portion in a second strand, where the first strand and the second strand are the two antiparallel strands in a duplex segment of each of the one or two adaptor sequences in the target nucleic acid sequence, and the first portion and the second portion are two different portions of the duplex segment of the target nucleic acid sequence.) This arrangement allows the simultaneous addition of the at least one blocking oligo in a blocking reaction without temporal separation, and brings about convenience.

[0219]In another example, two blocking oligos (or two blocking oligo sets) respectively target a same portion of a duplex segment of each of the one or two adaptor sequences in a target nucleic acid sequence, and as such, the two blocking oligos (or oligo sets) need to be allowed to contact the at least one target sequence in the biological sample in a sequential manner, or in a separate and then combined manner.

[0220]FIG. 6A shows one specific embodiment as an illustrating example.

[0221]In order for the set of blocking oligos to hybridize with, and thereby to block, the corresponding strand in each of the pair of two adaptor sequences respectively, the blocking hybridization reaction can occur at a substantially same temperature as the aforementioned hybridization reactions respectively required for hybridizing the first RNA probe set 500A and the second RNA probe set 500B to each corresponding strand of each target nucleic acid sequence. As such, according to some embodiments, the blocking hybridization reaction can be at a range of ~40-90°C, preferably at ~62-70°C, and more preferably at ~67.5°C.

[0222]As such, by applying a corresponding set of blocking oligos 611, 612, 621, and/or 622 to block each of the two strands of the first adaptor sequence 600A and each of the two strands of the second adaptor sequence 600B that flank each of the at least one target nucleic acid sequence in the DNA library in the sample (as illustrated in FIG. 6A), the potential interference by each strand of the first adaptor sequence 600A and the second adaptor sequence 600B can be minimized.

[0223]According to some other embodiments, only one adaptor sequence is next to each double-stranded target nucleic acid sequence, and as such, one blocking oligo (which targets one of the two strands of the adaptor sequence) or a set of two blocking oligos (which respectively target two strands of the adaptor sequence) can be used.

[0224]According to yet some other embodiments, each target nucleic acid sequence is single-stranded in the biological sample and is flanked by two adaptor sequences, and as such, a set of two blocking oligos respectively targeting the two adaptor sequences can be used. There are other possibilities as well, and the specific design is dependent on actual situation. There are no limitations herein.

[0225]It is noted that one or more adaptor sequences as described above serve only as illustrating examples for the at least one un-targeted sequence in a polynucleotide sequence containing a target nucleic acid sequence, and do not limit the scope of the disclosure.

[0226]S300: immobilizing the at least one target nucleic acid sequence on a solid support;

[0227]Herein step S300 can be performed by means of the immobilization portion in the at least one functional portion that has been labelled onto each of the at least one RNA probe, as mentioned above.

[0228]Specifically, S300 can be carried out by means of a stable binding (i.e. non-covalent binding) between the immobilization portion in each RNA probe and a coupling partner immobilized on a surface of a solid support. In some specific embodiments where the immobilization portion is a biotin moiety, the coupling partner can be a streptavidin, avidin, or an anti-biotin antibody, attached onto, or conjugated with a solid support such as magnetic beads. Other examples of the immobilization portion-coupling partner pair can include, but is not limited to, a carbohydrate-lectin pair, an antigen-antibody pair and a negative charged group-positive charged group static interacting pair.

[0229]Alternatively, S300 can be carried out by means of a cross-link (i.e. covalent connection) between the immobilization portion and a coupling partner attached onto a solid support. In some specific embodiments, the immobilization portion can be a first coupling partner, which can form a cross-link with a second coupling partner, allowing for the further immobilization of the captured sequences. As such, the first coupling partner and the second coupling partner are respectively one and another of a cross-linking pair, selected from one of an NHS ester-primary amine pair, a sulfhydryl-reactive chemical group pair (e.g. cysteines, or other sulfhydryls such as maleimides, haloacetyls, and pyridyl disulfides), an oxidized sugar-hydrazide pair, photoactivatable nitrophenyl azide's UV triggered addition reaction with double bonds leading to insertion into C-H and N-H sites or subsequent ring expansion to react with a nucleophile (e.g., primary amines),

or carbodiimide activated carboxyl groups to amino groups (primary amines), etc.. There are no limitations herein.

[0230]By means of this step of the method, the target nucleic acid molecules are isolated, enriched, or captured from the biological sample via the labelled RNA probes. Solid supports that can be used may be any that are convenient for the particular purpose and situation.

[0231]S400: eluting out the at least one target nucleic acid sequence from the solid support.

[0232]After the immobilization of the at least one target nucleic acid sequence to the solid support through the at least one immobilization portion-labelled RNA probe in step S300, the plurality of immobilized target nucleic acid sequences are enriched or captured, which can then be eluted from the solid support in step S400 to facilitate the subsequent treatment and analysis, such as PCR amplification, a sequencing assay (such as next generation sequencing and other sequencing assays), PCR-based detection, microarray assays, construction of gene fragments into clones, transfection and transduction, and all other nucleic acid based applications. There are no limitations herein.

[0233]According to some embodiments of the method, the step S400 can include a washing process followed by an elution process. In the washing process, the hybridized molecules extracted by the solid supports can be washed to remove unspecific nucleic acid binders, including nucleic acids that are binding to probes or binding to solid supports or binding to any other moieties due to unspecific interactions. Specifically, the washing process can be carried out by saline-sodium citrate (SSC) buffer with 0.1% SDS.

[0234]In the elution process, nucleic acid sequences that are hybridized to the RNA probes can be eluted through heat-induced strand dissociation or through nucleic acid denaturing reagents, such as 0.1M sodium hydroxide. After separation of the probe from the target nucleic acid molecule through elution, a neutralizing buffer, such as Tris-HCl pH7.5, can be used to treat the target nucleic acid molecule portion of the elution reaction to further neutralize the effects initiated by the denaturing buffer if initially utilized.

[0235]According to some embodiments of the method, rather than immobilizing the at least one target nucleic acid sequence on the solid support (i.e. step S300) after hybridizing the RNA probes generated by step S100 with the at least one target nucleic acid sequences (i.e. step S200), the RNA probes generated by step S100 can be first immobilized on the solid support (and thus become solid support-immobilized RNA probes or solid support-conjugated RNA probes), and then can be allowed to contact with the target nucleic acid sequences for capturing. As such, the method can comprise the following steps, as illustrated in FIG. 1B:

[0236]S100': Preparing at least one pair of RNA probe sets, each pair comprising a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a

duplex segment in each of at least one target nucleic acid sequence, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with an immobilization portion;

[0237]S200': Conjugating each of the at least one pair of RNA probe sets on a solid support;

[0238]S300': Contacting one and another of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set with the at least one target nucleic acid sequence to allow hybridization of each RNA probe in the at least one pair of RNA probe sets to a corresponding strand of each target nucleic acid sequence; and

[0239]S400': Eluting out the at least one target nucleic acid sequence from the solid support.

[0240]Herein, steps S100' and S400' in the embodiments of the method as described above are substantially same as steps S100 and S400 in the aforementioned embodiments of the method as illustrated in FIG. 1A. The solid support can be magnetic beads, non-magnetic beads, resin matrix, filter, membrane, or a different type that has been mentioned above.

[0241]Specifically, in a manner similar to the above embodiments where the RNA probes are immobilized to the solid support only after their hybridizations with target nucleic acid sequences, a pair of solid support-immobilized RNA probe sets (i.e. the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the pair) can be allowed to sequentially contact with the at least one target nucleic acid (i.e. one after another), or to separately contact with the at least one target nucleic acid in two hybridization reactions and then to combine the two hybridization reactions.

[0242]As such, according to a first embodiment which is substantially a sequential manner, S300' includes:

[0243]S310: Contacting one of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set with the at least one target nucleic acid sequence in a sixth hybridization reaction; and

[0244]S320: Contacting another of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set with the at least one target nucleic acid sequence in a seventh hybridization reaction.

[0245]According to a second embodiment, S300' includes:

[0246]S310': Separately contacting the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in each pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in an eighth hybridization reaction and a ninth hybridization reaction, respectively; and

[0247]S320': Combining the eighth hybridization reaction and the ninth hybridization reaction to thereby allow a tenth hybridization reaction to proceed.

[0248]It is noted that depending on the specific properties of the solid support, it is possible that a pair of RNA probe sets conjugated on the solid support have little or acceptable level of interactions among RNA probes having complimentary sequences. As such, according to a third embodiment, a solid support-conjugated first RNA probe set and a solid support-conjugated second RNA probe set in each corresponding pair can be allowed to contact with the at least one target nucleic acid in a single hybridization reaction without being separated temporally. For example, if beads are used, the beads-conjugated first and second RNA probe set in each corresponding pair can be combined in a single reaction for the simultaneous capture of different strands of a duplex segment of a target nucleic acid sequence. Alternatively, a glass surface can be used for conjugation of both the first and the second RNA probe set in each corresponding pair thereon and can allow a simultaneous capture of different strands of a duplex segment of a target nucleic acid sequence in a single reaction.

[0249]It is further noted that according to a fourth embodiment, a first RNA probe set and a second RNA probe set in each corresponding pair can be conjugated at a different region of the solid support, such as at a different segment of a microfluidic channel on a microfluidic chip, which allows a sample containing the at least one target nucleic acid sequence to sequentially flow through the different segments of the microfluidic channel corresponding to the pair of solid support-conjugated RNA probe sets to thereby allow a sequential capture of the two different strands of a duplex segment of a target nucleic acid sequence. Such a configuration allows a repeated use of RNA probes conjugated on the solid support.

[0250]According to some embodiments of the method, rather than conjugating each of the at least one pair of RNA probe sets on a solid support after preparing the at least one pair of RNA probe sets (i.e. steps S100' and S200' in the above mentioned method shown in FIG. 1B), the RNA probes in each of the at least one pair of RNA probe sets can be directly prepared on the solid support. As such, the method can comprise the following steps, as illustrated in FIG. 1C:

[0251]S100'': Preparing at least one pair of RNA probe sets directly on a solid support, each pair comprising a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a duplex segment in each of at least one target nucleic acid sequence;

[0252]S200'': Contacting one and another of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set with the at least one target nucleic acid sequence to allow hybridization of each RNA probe in the at least one pair of RNA probe sets to a corresponding strand of each target nucleic acid sequence; and

[0253]S300'': Eluting out the at least one target nucleic acid sequence from the solid support.

[0254]Herein, steps S200'' and S300'' in the embodiments of the method as described above are substantially same as steps S300' and S400' in the aforementioned embodiments of the method as

illustrated in FIG. 1B. The solid support can be magnetic beads, non-magnetic beads, resin matrix, filter, membrane, or a different type that has been mentioned above.

[0255]Herein, step S100'' can be realized by direct chemical synthesis or by direct transcription on the solid support. In cases where the direct transcription on the solid support, a RNA polymerase can be attached onto the solid support.

[0256]The method, as illustrated by the various embodiments as described above, has the following advantageous features:

[0257]First, a pair of RNA probes which respectively target two antiparallel strands (i.e. a plus strand and a minus strand) of a duplex segment of each target nucleic acid sequence are both employed for the enrichment or capture of both strands of each of the target nucleic acid sequences, resulting in a higher enrichment efficiency compared with a conventional method where there is only one set of RNA probes targeting only one of the two strands of each target nucleic acid sequence.

[0258]Second, the nucleic acid sequence capture approach as described above can substantially capture both of the two antiparallel strands that typically belong to a same molecule of each target nucleic acid sequence and have complementary sequences, and can thus bring additional advantages in certain applications having a high requirement for sequence accuracy, such as single-nucleotide variation (SNV) calling. In an exemplary application in detecting SNVs in a sample, only an SNV observed on both of the two complementary strands in a target DNA sequence that have been respectively captured by a corresponding pair of RNA probe sets respectively targeting both of the two complementary strands is a bona fide SNV, whereas an SNV observed only on one, but not on both, of the two complementary strands in a target DNA sequence is likely to represent a false positive signal introduced by PCR. As such, the nucleic acid sequence capture approach disclosed herein allows for an error-proof application.

[0259]Third, each RNA probe can be prepared by transcription, which allows a large amount of RNA probes to be cost-effectively and conveniently obtained. This allows the target nucleic acid sequences to be enriched/captured at a significantly higher efficiency due to the much higher probe/target ratio. Additionally, this further allows a quantification of each RNA probe, in turn causing each RNA probe to be conveniently tweaked to increase the efficiency of capturing some specific target nucleic acid sequences that are difficult to capture by, for example, using a 10-fold higher amount of RNA probes for capturing.

[0260]Fourth, the above advantageous features together can have an additional advantage such that the hybridization of a first RNA probe with its corresponding strand of each target nucleic acid sequence could help expose the other complementary strand of each target nucleic acid sequence, thereby could kinetically favor the subsequent hybridization of another RNA probe with the

complimentary strand of the each target nucleic acid sequence, leading to a favorable capturing of each of the two strands of each target nucleic acid sequence in the biological sample.

[0261]Due to these above advantages, the method disclosed herein is expected to have a variety of applications, as follows:

[0262]First, it allows for the capture and analysis of sequence variants and/or copy number variants, including without limitation, a point mutation, a deletion, an amplification, a loss of heterozygosity, a rearrangement, and/or a duplication. These genetic variants may be associated with human diseases, specific phenotypes, etc. The analyses include sequencing, hybridization assay, ligation assay, etc. Due to its ultra-sensitivity, the method disclosed herein allows for the capture and analysis of rare variants, especially ultra-rare variants/mutations, and copy number variants, and can be employed in capture and enrichment of nucleic acids from mitochondria, chloroplast, plastid, bacterial or viral pathogens, environmental DNA (eDNA), etc., and can also be employed in capture and enrichment of nucleic acids for population genetics studies, SNP typing and deep phylogenies, RAD (Restriction-site-Associated DNA sequencing) or GBS (Genotyping By Sequencing) locus enrichment. There is no limitation herein.

[0263]Second, it allows for enrichment and molecular cloning of specific target nucleic acid sequences blended in a DNA sample, such as ancient DNA and nucleic acid materials from museum specimens, which typically have a poor DNA quality and have a typically serious contamination of other organisms (esp. microorganisms) over many years' storage in nature environment, or contamination of modern human DNA due to inappropriate sample handlings.

[0264]Third, the method can be applied broadly to capture a variety of nucleic acid sequences in a test sample, which can include a double-stranded nucleic acid sequence, which consists of a plus strand and a minus strand, as illustrated in FIG. 6B, left panel), and can also include a single-stranded nucleic acid sequence having a duplex segment, which substantially forms a loop or a hairpin (as illustrated in FIG. 6B, right panel), and the nucleic acid sequences can include DNAs, RNAs, or DNA-RNA hybrid molecules.

[0265]It is noted that even in situations where any of the target nucleic acid sequences does not form a duplex structure (i.e. the target nucleic acid sequence has only a single-stranded segment), such as a case where cDNA is synthesized via reverse transcription yet the mRNA template is degraded due to an RNase-induced digestion, the method as described above can still be applied to capture a target nucleic acid sequence as long as any of the first RNA probe set or the second RNA probe set can target the single-stranded segment of the target nucleic acid sequence.

[0266]Fourth, the method as disclosed herein, if combined with the use of a method and a kit for constructing a barcoded nucleic acid library as disclosed in the patent application PCT/US2018/016778 (i.e. if used to enrich and capture target nucleic acid sequences in the

barcoded DNA library by the kit and method disclosed therein), can allow an ultra-sensitive error-proof assays for the detection and characterization of target nucleic acid sequences in a biological sample.

[0267]In the following, several illustrating examples are provided to illustrate the several applications of the method disclosed herein.

[0268]FIG. 7A and FIG. 7B show a diagram of the nucleic acid sequence enrichment method according to some embodiments of the present disclosure.

[0269]As illustrated in FIG. 7A and FIG. 7B, double-stranded DNA molecules in a NGS DNA library are dissociated and the shared adapter, index and universal primer sequences among all molecules were hybridized by blocking oligos, and the target DNA sequences are further captured by RNA probes that are complementary to the DNA molecules that are composed of antiparallel double strands. Both the + strand and the – strand from the same DNA molecule are captured respectively by at least one complementary RNA probe. After immobilization of the captured DNA sequences on the solid support (i.e. streptavidin magnetic beads), the target sequences are extracted from the original library and the target DNA molecules being captured can be eluted from the probe after a series of wash and elution steps. After amplification, the library is ready for direct NGS sequencing or other assays.

[0270]Any means of testing for a sequence variant or sequence copy number variant, including without limitation, a point mutation, a deletion, an amplification, a loss of heterozygosity, a rearrangement, a duplication, may be used. Sequence variants may be detected by sequencing, by hybridization assay, by ligation assay, etc. The defined locations of some mutations permit focused assays limited to an exon, domain, or codon. But un-targeted assays may also be used, where the location of a mutation is unknown. If locations of the relevant sequence variants are defined, specific assays which focus on the identified locations may be used. Any assay that is performed on a test sample involves a transformation, for example, a chemical or physical change or act. Assays and determinations are not performed merely by a perceptual or cognitive process in the body of a person.

[0271]Probes and/or primers and/or template for RNA probe synthesis may contain the wild-type or a sequence variant, including without limitation, a point mutation, a deletion flanking sequence, a rearrangement location, may be used. These can be used in a variety of different assays, as will be convenient for the particular situation. Selection of assays may be based on cost, facilities, equipment, electricity availability, speed, reproducibility, compatibility with other assays, invasiveness of sample collection, sample preparation, etc.

[0272]Any of the assay results may be recorded or communicated, as a positive act or step. Communication of an assay result, diagnosis, identification, or prognosis, may be, for example,

orally between two people, in writing, whether on paper or digital media, by audio recording, into a medical chart or record, to a second health professional, or to a patient. The results and/or conclusions and/or recommendations based on the results may be in a natural language or in a machine or other code. Typically, such records are kept in a confidential manner to protect the private information of the patient or the project.

[0273]Collections of RNA probes, primers, control samples, and reagents can be assembled into a kit for use in the methods. The reagents can be packaged with instructions, or directions to an address or phone number from which to obtain instructions. An electronic storage medium may be included in the kit, whether for instructional purposes or for recordation of results, or as means for controlling assays and data collection.

[0274]Control samples can be obtained from the same patient from a tissue that is not apparently diseased. Alternatively, control samples can be obtained from a healthy individual or a population of apparently healthy individuals. Control samples may be from the same type of tissue or a different type of tissue than the test sample. Control samples may be provided together with the RNA probes, primers, and reagents in a kit for use in the method, where the control samples may be a standard reference sample for the purpose of validating the performance of the kit and the operation performed by the user.

[0275]The data described below document the results for the identification of ultra-rare mutations from a whole exome sequencing study using RNA probes targeting both strands of the target DNA molecules. There is no doubt that SNVs can be detected with confidence only when the sequencing system's error rate is significantly lower than the frequency of identified SNVs. Therefore, baseline error rate of an NGS pipeline is critical for its performance of detecting ultra-rare SNVs. Combing the RNA probe-based capture of both DNA strands with the single-stranded library construction, an improved NGS methods with the base line error rate as 2.25×10^{-10} was created. Such high-accuracy pipeline is dependent on the massive amount of DNA library target molecules captured by the RNA probes targeting both DNA strands.

[0276]FIGS. 8A-8D show that reduced amounts of variants were re-detected from sequentially diluted samples. No variant was re-detected from 1:10,000 diluted group. Coverage of re-sequencing is ~5,000 X. The efficiency of capture through a paired sets of RNA probes is significantly stronger than the efficiency of capture through a single set of RNA probe for any target sequence.

[0277]FIGS. 15A, 15B, 15C, 15D and 15E show sequence variants detected by RNA probe-based DNA double strand capture NGS, validation results by Sanger sequencing and ultra-rare mutation redetection results are shown and ranked by Mutant Allele Fraction.

[0278]The ultra-rare mutation detection performance of this method with the target molecules captured by RNA probes targeting both strands of the DNA molecules was then evaluated by the success rate of re-detecting the 38 Sanger sequencing validated sequence variants in the libraries created from normal DNA samples which were spiked with sequential dilutions of tumor DNA. The library is constructed by a barcoded single-strand molecule-based approach and the target enrichment of the whole exome region of the human genome is performed by RNA probe-based DNA double strand capture. As the dilution folds increased, as expected, fewer and fewer variants were detected (FIG. 12), and when the tumor DNA sample was diluted 1,000 folds (the diluted sample containing 0.1ng tumor DNA and 100ng normal DNA), only 21 out of the 38 validated variants can be detected (FIGS. 15 A-E). The allelic fractions of these 21 SNVs in the 1:1000 diluted sample range from 0.03% to 0.005% with an average of 0.013% (FIGS. 15 A-E). No sequence variant was detected in 1:10,000 diluted sample which may presumably be due to the limitation of sequencing depth that has been achieved. For each sample, the targeted sequencing was performed with an average depth of 5,000X, which theoretically only allows us to see SNVs down to the frequency of 1/5000 (0.02%). To observe ultra-rare SNVs with even lower frequency, a greater than 5000X coverage is needed. It is also helpful to design capturing probes targeting only a small number of genes. With a smaller number of different nucleic acid sequences captured by the RNA probes targeting both strands of a smaller cohort of target genes, instead of the entire human genome exome, more copies of the target nucleic acids can be enriched, and this method can achieve a much greater sequencing depth with a significantly improved accuracy of ultra-rare SNV calling. The extremely low baseline error rate of this method allows ultra-rare SNV calling at the whole exome level with high accuracy, and the depth of NGS sequencing becomes the only limiting factor for such applications.

[0279]An RNA probe-based DNA double strand capture approach was reported as an improved method to enrich DNA molecules for NGS purpose, particularly targeted NGS. Such improved performance has been demonstrated in a human genome WES study. Aside from WES, another very important application of RNA probe-based DNA double strand capturing would be the targeted resequencing of a gene panel. Targeted re-sequencing is one of the most popular NGS applications, and it allows people to sequence a small cohort of gene targets to extreme depths, usually thousands of folds of coverage. And such sequencing depth can facilitate the detection of ultra-rare mutations with great sensitivity. In an RNA probe-based DNA double strand capture pipeline, attempts were made to capture the entire exome of all human genes, where an over 98% coverage with the depth of over 200X was achieved on a standard NGS platform. More importantly, the detection limit of this method for rare-mutation detection on whole exome scale is as low as 0.03%, which is made possible by the massively improved capture efficiency of the target

molecules by the RNA probes targeting both DNA strands. For an even smaller cohort of target genes, the depth and coverage of RNA probe-based DNA double strand capturing NGS can be further increased, and the performance of ultra-rare mutation detection can be subsequently improved over several additional orders of magnitude.

[0280]Other than identifying ultra-rare SNVs with high sensitivity and accuracy, RNA based double strand capture method can also be adopted for gene copy number variant (CNV) assays. Barcoded single-stranded library construction links a unique barcode to every single-stranded DNA molecules. Such barcode information can not only be used to label the molecules and create super reads to reduce PCR errors, but also be used as a location marker for DNA fragments. After RNA probe-based DNA double strand capturing, NGS sequencing, and mapping the super reads back to the human genome, the barcode on each super read can be assigned to the position where the super read sequence is mapped. Therefore, a human genome can be reconstructed by unique barcodes. Copy number information can be represented by the diversity of barcodes at subgenomic loci. A highly efficient capture reaction with equal efficiency for all genomic regions offered by RNA probe-based DNA double strand capture is the key to a successful CNV calling.

[0281]Aside from CNV analysis, large structural variants frequently observed in cancer genomes can also be analyzed through RNA probe-based DNA double strand enrichment. RNA probes can be designed to enrich subgenomic regions flanking popular genome breakpoints specifically. A highly sensitive pipeline for translocation and large indel identification could be built based on the high efficiency of RNA probe-based DNA double strand capture.

[0282]In addition to applications in basic research, RNA probe-based DNA double strand capturing has a great potential in clinical NGS fields. It has been demonstrated that this method can highly efficiently construct NGS DNA libraries with very low amount of DNA materials ($\leqslant$ 20pg). Meanwhile it can detect ultra-rare mutations with high confidence. Such features are critical for NGS based clinical diagnostics where the samples are often limited and highly heterogeneous. A typical example would be the NGS sequencing of FFPE samples. FFPE has been a standard sample preparation method for many decades. Historically archived FFPE sample is a very valuable resource for retrospective studies in biomedical research. However, due to chemical modifications during specimen preparation and chronic damages to the tissue blocks or slides over long-term storage, it has been a challenging task to conduct NGS studies with FFPE samples. Poor DNA quality and artificial sequence changes are two major issues coming along with FFPE based NGS studies. RNA probe-based DNA double strand capturing is offering great benefit for FFPE based WES studies. WES data have been reported to be discordant between FFPE and fresh frozen samples at lower coverage levels (~20X), however, this discrepancy can be reduced when higher coverages are achieved. And recently, Allen *et al.* reported a reciprocal overlap of 90% somatic

mutations between FFPE and fresh frozen tissue samples for the positions with sufficient sequencing (Van Allen, Wagle et al. 2014). In Allen's study, an RNA probe-based DNA single strand capture approach was applied, where its capture efficiency is far lower than RNA probe-based DNA double strand capture approach as was shown in data. With the enhanced capture efficiency by this method, WES studies with FFPE samples will offer comparable data quality to WES studies with fresh frozen tissues.

[0283]This method has a great potential to discover novel low-frequency disease-causing variants in biomedical and clinical applications, and can identify more actionable therapeutic targets for patients. This method can fulfill an unprecedented level of personalized precision medicine by revealing the most complete patient genomic profile to date including high-frequency, low-frequency and particularly ultra-low-frequency mutations. This method can also be applied in other clinical applications, like circulating DNA sequencing from body fluid samples, where only limited amount of DNA materials is available. In clinical NGS applications, it is critical to highly efficiently capture target DNA molecules from NGS libraries constructed with very limited amount of highly heterogeneous samples thus being less- or non-invasive; to highly efficiently enrich target sequences thereby reaching a great sequencing depth with limited cost and improved diagnostic sensitivity; and to remove artificial sequencing errors as completely as possible for the best diagnostic specificity. This method, utilizing two sets of RNA probes to capture the two strands of DNA antiparallel molecules simultaneously, has been demonstrated to meet these needs with great potentials in numerous NGS applications.

[0284]EXAMPLE 1

[0285]MATERIALS AND METHODS

[0286]Tumor and normal tissue sample

[0287]The paired tumor and normal tissue samples from a pancreatic cancer patient of Asian race were obtained in accordance with guidelines and regulations from Tianjin Medical University Cancer Institute & Hospital, P.R. China after Institutional Review Board (IRB) approval at Tianjin Medical University, and under full compliance with HIPAA guidelines. An informed consent for conducting this study was obtained from the patient. The tumor tissue sample has an estimated neoplastic content of 43.4%.

[0288]Library preparation

[0289]Genomic DNA from patient normal and tumor fresh frozen tissues was extracted using DNeasy Blood & Tissue Kit (Qiagen) and sheared into 150bp fragments with Diagenode's Bioruptor at a program of 7 cycles of 30 seconds ON/90 seconds OFF using 0.65 ml Bioruptor® Microtubes. Barcoded single-stranded library preparation starts from a complete dissociation of DNA duplex to form single-stranded DNA and tagging the 3' end of each DNA single strand

individually with a unique digital barcode. Barcoded single-stranded adapters have been disclosed in international application PCT/US2018/016778, the disclosure is incorporated herein in its entirety. Pre-dephosphorylated fragmented DNA samples were mixed with barcoded single-stranded adapter (final concentration 0.15uM), 20% PEG-8000, 100U CircLigase II, and incubated at 60°C for 1 hour. After immobilizing the ligation product on Streptavidin-coupled Dynabeads (ThermoFisher Scientific), each barcoded single-stranded DNA molecule is subject to an individual single-cycled PCR reaction to form its complementary strand. A DNA primer complimentary to the single-stranded adapter was annealed and extended using Bst 3.0 polymerase at 50°C for 30 minutes. Blunt-end repair using T4 DNA polymerase was performed at 25°C for 15 minutes. A double-stranded adapter was then ligated to the 5' end of the DNA duplex using T4 DNA ligase with an incubation at 16°C for 1 hour. The library is eluted from the beads by an incubation at 95°C for 1 minute. High fidelity PCR amplification is performed to amplify the DNA sequence as well as the unique barcode. Adapter sequences are designed to be compatible with Illumina sequencing platforms.

[0290]RNA probe synthesis

[0291]To obtain RNA probes complementary to both strands of target subgenomic regions, particularly the exome regions, the entire exome sequences for every human gene were cloned by sequence synthesis and molecular cloning based on Hg19 reference human genome sequences. In brief, exome sequences in 32,524 CCDS IDs containing -50bp and +50bp intronic sequences were cloned into pcDNA 6.2 vector. For extremely large human genes, e.g. DMD, PTPRD, CNTNAP2, etc., their related target sequences were separated and subcloned into multiple vectors. The total DNA sequences used to generate RNA probes cover a 72.6 Mb genome region, where all the exomes with their -50bp and +50bp flanking intronic sequences, as well as 5' and 3' UTRs for each gene were included. Two clones for each target sequence were constructed, where a T7 promoter was inserted at the 5' end of the plus strand in the "+" clone and at the 5' end of the minus strand in the "-" clone (FIG. 4B and FIGS. 7A and 7B). Two pools of clones were established for any give number of genes following the rule that the two clones for the same DNA sequence are separated into two systems, where one system (FIG. 4B and FIGS. 7A and 7B, left panel, "+" Clone) produced the RNA probes targeting the plus strand of the DNA target, and the other system (FIG. 4B and FIGS. 7A and 7B, right panel, "-" Clone) produced the RNA probes targeting the minus strand of the DNA target through *in vitro* transcription. ATP, CTP, GTP, UTP, and Biotin-16/11-UTP were added in each transcription system at the concentration of 1mM, 1mM, 1mM, 0.7mM and 0.3mM. RNA products were further sheared into 100-150nt fragments with a Covaris S220 focused-ultrasonicator (Covaris). The fragmented RNA probes are ready for RNA probe-based DNA double strand capture applications. The two RNA probe libraries for each target

DNA sequence were created separately and were never mixed until the actual capture procedure was carried out. In this study, RNA probes targeting both plus and minus strands of the whole exome sequences were created (including -50bp and +50bp flanking intronic sequences and 5' / 3' UTRs) for all human genes and a cancer-related 298-gene panel.

[0292]RNA probe-based DNA double strand capture

[0293]RNA probe-based DNA double strand capture was performed to capture the whole exome of human genome following a library construction or a standard NGS library construction. In RNA probe-based DNA double strand capture, both DNA strands of the target regions are captured by a pair of complementary RNA probes where each DNA strand is targeted by its complementary RNA probe, individually. A hybridization mixture was prepared containing 500ng DNA library, 2ug of RNA probes (1ug from "+" clone transcripts and 1ug from "-" clone transcripts, targeting Hg19 human exomes including- 50bp and +50bp flanking intronic sequences, as well as 5' and 3' UTRs as described before), 7ul Human Cot-1 DNA (ThermoFisher Scientific), 3ul Herring Sperm DNA Solution (ThermoFisher Scientific), 10µl blocking Oligos (1nmol/ul each) with following sequences:

[0294]Blocking Oligo 1: 5'-AAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTG GTCGCCGTATCATT-3' Inverted dT (SEQ ID NO. 913, where the last T is labelled with an inverted dT)

[0295]Blocking Oligo 2: 5'-CCTCAGCAAGAGCACACGTCTGAACTCCAGTCAC-NN NNNNN-ATCTCGTATGCCGTCTTCTGCTTG-3' Inverted dT (SEQ ID NO. 914, where N can be any of A, T, C, and G, and the last G is labelled with an inverted dT)

[0296]The hybridization mixture is heated for 5 minutes at 95°C, then held at 67.5°C. 25ul pre-warmed (67.5°C) 2.8 X hybridization buffer (14 X SSPE, 14 X Denhardt's, 14 mM EDTA, 0.28% SDS) was added. The mixture was slowly pipetted up and down 8 to 10 times. The hybridization mixture was incubated for 24 hours at 67.5°C with a heated lid.

[0297]After hybridization, 50µl Dynal MyOne Streptavidin C1 magnetic beads (ThermoFisher Scientific) were washed three times by adding 200 µl of binding buffer (1M NaCl, 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA), and re-suspended in 200µl Binding buffer. The hybridization mixture was added to the bead solution gently and was subsequently incubated on a thermomixer at 850 rpm for 30 minutes at room temperature. To wash the beads, the supernatant was removed from beads on a Dynal magnetic separator and the beads were re-suspended in 500µl Wash Buffer A (1X SSC/0.1% SDS), and incubated for 15 minutes at room temperature. Beads were then washed three times, each with 500 µl pre-warmed Wash Buffer B (0.1XSSC and 0.1% SDS) after incubation at 65°C for 10 minutes. To elute captured DNA, beads were re-suspended in 50µl 0.1M NaOH at RT for 10 minutes. The supernatant was transferred into a new 1.5 ml microcentrifuge

tube after magnetic separation, and mixed with 50µl Neutralizing Buffer (1M Tris-HCl, pH 7.5). DNA was purified with a Qiagen MinElute column and eluted in 17µl of 70°C EB buffer to obtain 15 µl of captured DNA library. The captured DNA library was amplified by Phusion Hot Start polymerase (New England Biolabs) using Illumina PE primer 1 and 2. The PCR program used was: 98°C for 30 seconds; 6 ~ 10 (depending on capture yield) cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds; and a final incubation at 72°C for 5 minutes. The PCR product was purified using GeneJET PCR Purification kit (ThermoFisher Scientific).

[0298]Real-time PCR assay

[0299]FIGS. 13A-13N illustrate 298 cancer-related gene targets, primer pairs, amplicon sequences, amplicon GC% and amplification efficiency constant for each real-time PCR detection are listed. Real-time PCR assays with SYBR green detection was carried out using an ABI PRISM 7500 Sequence Detection System (Applied Biosystems). Briefly, the reaction conditions consisted of 500ng of genomic DNA or DNA library products, 0.2 µM primers, and SYBR Green Real-Time PCR Master Mix (ThermoFisher Scientific) in a final volume of 20µl. Each cycle consisted of denaturation at 95°C for 15 seconds, annealing at 58.5°C for 5 seconds and extension at 72°C for 20 seconds, respectively. Gene specific primers were designed using Primer 3 (Untergasser, Cutcutache et al. 2012) and their sequences are provided in FIG. 8. Reactions were run in triplicate in three independent experiments. The primer pair's standard amplification curve for each gene was established through using sequential dilutions of the "+" clone constructs containing the amplicon sequence. Amplification efficiencies for 298 target amplicons were established and listed in FIG. 8. Gene abundance ratios between different samples were calculated by the raising the gene specific amplification efficiency (AE) to the power of ΔC_t value between different samples. For example, the ratio (r) of gene abundance in sample A vs sample B can be calculated through real-time PCR assay by: $r_{(A/B)} = AE^{\Delta Ct}$, where $\Delta C_t = C_{t\ (\text{sample B})} - C_{t\ (\text{sample A})}$

[0300]Whole exome sequencing

[0301]FIGS. 14 show that initial mapped reads represent raw reads that contain the 12nt barcode and mapped to the reference genome. Unique read family represents the number of URF. Each URF has a unique barcode and its sequence is obtained by consolidating read sequences arise from the same DNA molecule by PCR amplification. PCR errors are removed by requesting a sequence uniformity for over 95% of the reads within a URF. Super read duplexes represent the number of DNA duplex whose two strands are coming from two super reads.

[0302]Whole exome sequencing was performed on an Illumina HiSeq 2500 platform according to manufacturer's manual. Total numbers of on-target reads from randomly chosen 5 million to 50 million reads were calculated. After trimming and barcoded super read grouping, SNVs were called with GATK (version 3.6) in a default mode as recommended by the GATK documentation

with reference genome of Hg19 (McKenna, Hanna et al. 2010). In brief, for every sample (tumor or normal DNA), sequencing result was preprocessed by mapping to reference genome with BWA (version 0.7.10), and duplicates were marked with Picard (version 2.0.1). Base Recalibration was performed to generate the reads ready for SNV analysis. For individually processed T/N pair reads, Indel Realignment was performed to generate pairwise-processed T/N pair reads. HaplotypeCaller was used for raw SNV calling. Output from variant calling was directly used for SNV detection by MuTect (version 1). Mutations were filtered through a 4-step approach introduced in the section "Mutation and ultra-rare mutation detection". Low-quality variant with a Phred score <30.0 was abandoned. Paired SNVs from complementary reads bearing different barcodes were identified as true mutations and subject to further validation through Sanger sequencing. The data yields after each step of data analysis for an RNA probe-based DNA double strand capturing NGS study were shown in FIG. 14. SNVs identified and Sanger sequencing validation results were provided in FIGS. 15 A-E.

[0303]Mutation and ultra-rare mutation detection

[0304]The significantly increased number of unique reads obtained through RNA probe-based DNA double strand capturing enabled us to apply stringent filters with the following 4-step procedure.

[0305]Step 1) group reads with the same barcode that are representing PCR duplicates of an original barcoded single-stranded DNA molecule, and call it a unique read family (URF);

[0306]Step 2) combine reads within each URF obtained from Step 1) by requesting >95% sequence identity among the reads;

[0307]Step 3) extract the unique DNA sequence and the barcode sequence for each URF, and call it a "super read"

[0308]Step 4) for all the super reads identified in Step 3), find their paired complementary super reads, and only score sequence variants with matched complementary sequences from paired super reads. To accommodate damaged DNA molecules in the sample, complementary super reads may not be at the same length.

[0309]To evaluate the performance of RNA probe-based DNA double strand capture in detecting low frequency (ultra-rare) mutations, 100ng tumor DNA sample was sequentially diluted by 10, 100, 1,000 and 10,000 folds, and spiked each of them into the same amount (100ng) of genomic DNA extracted from the paired normal tissue of the aforementioned cancer patient. This design can simulate early stages of cancer occurrence, and represent the major obstacles in early cancer diagnostics using NGS, which is the very low allelic fractions of tumor specific mutations in the sample.

[0310]Build a highly accurate reference exome for ultra-rare mutation identification

[0311]To highly accurately assess the baseline mutation frequency of barcoded single-stranded library and RNA probe-based DNA double strand capturing pipeline, six replicates of standard NGS DNA libraries were constructed in parallel, each using 100ng normal DNA input. These six replicates of exome datasets were used to re-build reference exome database for this particular patient by requesting that if the same SNV was observed in $\geqslant$ 5 out of 6 independent datasets, the SNVs can be considered as germline variants and the reference exome sequence database will be updated. For a standard NGS pipeline, the error rate is 1%, and the chance to see the same random error at a fixed position for 5 times is $(⅓ \times 1\%)^5 = 4.12X10^{-13}$. This number means that if this approach is used to to sequence the whole human genome once, there is presumably going to be only one artificial error, because 3×10^{12} human genome bases X $(4.12 \times 10^{-13}) = 1.24$. However, only the human exome sequences are enriched and sequences, which is occupying only 1.5% of human genome, therefore the chance to see a single artificial error within the entire human exome is only 1.86% (=1.5% X 1.24). An updated highly accurate normal exome reference database of the patient was built accordingly.

[0312]EXAMPLE 2

[0313]RNA probe-based DNA double strand capture achieved a high enrichment efficiency

[0314]A method is developed to enrich targeted subgenomic sequences by RNA probe-based capture of both strands from the same DNA molecule, simultaneously. To assess the capture efficacy, RNA probes were created for the exome regions of the 298-gene panel adopted in this study. Real-time PCR assays were performed to detect and quantify the subgenomic regions of this gene panel in NGS library before and after RNA probe-based enrichment. No re-amplification of the library was performed after the capture to ensure that the amounts of DNA molecules obtained from RNA probe-based enrichment represent the captured yields for each gene.

[0315]FIG. 8A shows that for each of the six DNA libraries derived from different amounts (500ng, 20ng, 1ng, 100pg, 20pg and 10pg) of input genomic DNA, enrichment efficiency of 298 cancer-related genes were calculated. Recovery ratios of DNA single and double strand captures by RNA probes for all 298 genes in six libraries were quantified by real-time PCR assays detecting each gene's abundance in the libraries before and after single strand capture or double strand capture; FIG. 8B shows that an insert sequence composed of amplicon regions of five genes whose GC contents cover a broad range (27.3% to 74.1%) was cloned into a pcDNA vector; FIG. 8C shows that Real-time PCR analysis of sequential dilutions of the plasmid. 1, 10, 100, 1,000 and 10,000 femtomoles of the plasmids were added as templates for the assays. C_t value for each gene observed from different plasmid template amount was plotted, and trend lines were shown. No significant GC-dependent amplification bias was observed for real-time PCR assays; FIG. 8D

shows that a whole genome library, and whole exome libraries captured by RNA based DNA single and double strand captures were analyzed on an agarose gel.

[0316]The same set of six libraries created from a sequentially diluted DNA input was adopted again. Such results indicate that each gene's capture ratio is in consistency for all six libraries (FIG. 8A). Such findings demonstrate that RNA probe-based DNA double strand capture efficiency for each gene is not dependent on the initial However, the capture ratios of different genes did vary to a significant extent (10.4% to 49.8%). Further investigation showed that RNA probe-based capture ratios for different genes were loosely correlated with the GC content of their amplicons (FIG. 8A). A very weak correlation (average R^2=0.12) between amplicon GC contents and their RNA probe-based capture efficiencies were shown (FIG. 8A).

[0317]Library construction, RNA probe-based capture or the Real-time PCR assays are all potentially responsible for this GC content associated enrichment bias. Previous results have demonstrated that library construction is not significantly biased by GC content. Next, real-time PCR was investigated to check its potential GC content bias. 5 genes were chosen, PTEN, PALB2, ESR1, CSF1R, and NSD1, with distinct GC% in their amplicon sequences at 27.3%, 39.3%, 50.5%, 62%, and 74.1%, respectively (FIGS. 13A-13N). A plasmid (pcDNA 6.2 vector) containing a DNA insert composed of all five genes' amplicon regions separated by a 100bp flanking sequence is cloned (FIG. 8B) and sequentially diluted to simulate variable amounts of the gene fragments after capture. All 5 gene fragments were cloned into the same plasmid to ensure the equal abundance between different genes in every diluted sample. Real-time PCR was performed to detect the copy number of plasmids by detecting each of the five genes in a series of diluted plasmid samples. As shown in FIG. 8C, there is no obvious GC-content-associated bias observed from the real-time PCR amplifications of all five genes using these primers. The 298 pairs of primers adopted in real-time PCR assays were designed with their T_m values all falling into a narrow range of 57°C to 61°C. This restriction of T_m and short amplicon sequences with similar lengths (~around 150bp) helped ensure the uniformity of real-time PCR assays for all genes (FIGS. 13A-13N). Therefore, the only possible step, where the GC content related capture ratio bias is created, should be the RNA probe-based capture itself. Enrichment bias for hybridization-based subgenomic capture has been reported to be owing to GC content. Whole exome capture NGS studies were conducted to assess further the impact of GC content to the RNA probe-based capture efficiency of target subgenomic regions.

[0318]It is important to note that in this RNA probe-based capture, complementary RNA probes were used to capture both DNA strands of the target regions, and attempts were made to assess if there is any capture efficiency difference between using only one set of RNA probes to capture only one strand of target DNA and using two sets of RNA probes to capture both strands of target

DNA simultaneously. These two capture methods were performed in parallel with two equal aliquots (500ng) of DNA libraries, where each library was created from the same 20ng genomic DNA. A whole genome library, captured yields by RNA probes targeting both strands of DNA molecules or RNA probes targeting only a single strand of the DNA target molecule were analyzed on an agarose gel (FIG. 8D). Real-time PCRs for the 298-gene panel were performed to evaluate the capture ratios of RNA probe single strand capture and double strand capture for all the genes (FIG. 8A). The average capture ratios for the target nucleic acid sequence capture method as disclosed herein (i.e. the double strand-targeting RNA probe-based capture method) is 29.2% across all genes in different libraries, much higher than the ratios observed from the conventional single strand-targeting RNA probe-based capture approach (~8.5% on average). These results have demonstrated that to capture target DNA sequences by complementary RNA probes through hybridizing to both strands of the DNA duplex molecule simultaneously achieved an over 3-fold increase in capture efficiency compared to capturing a DNA single strand alone through RNA probes.

[0319]EXAMPLE 3

[0320]Whole Exome Sequencing

[0321]FIG. 10A shows a bar plot of percentage of initial reads, mapped reads and reads remained after filtering. Results were obtained from three technical replicates. Numbers of reads were shown under each bar with the unit of 1 million reads. FIG. 10B shows a stacked bar plot of subgroups of filtered reads in triple replicates. FIG. 10C shows a coverage efficiency correlation with read numbers. The percentage of target bases covered at $\geqslant$10X, $\geqslant$20X, $\geqslant$50X and $\geqslant$100X depths with 5 million to 50 million reads were shown.

[0322]To evaluate the performance of RNA probe-based DNA double strand capture in NGS, WES assays were performed using this method and compared the data to what obtained through standard NGS library preparation with a standard exome enrichment procedure. All libraries were constructed with 100ng genomic DNA derived from the normal tissue of the cancer patient, and three technical replicates were performed for each sample. All NGS runs were carried out on the same Illumina HiSeq 2500 platform with the same technical specifications of the runs. As shown in FIG. 10A, an average of 188 million reads were obtained from RNA probe-based DNA double strand capturing WES, where 98.3% were aligned to the human genome, and the total read counts were significantly more (1.6 folds) than that from the standard sequencing pipeline. The higher numbers of reads for the libraries presumably came from the ultra-sensitive single-stranded DNA library construction, and the much more efficient RNA probe-based enrichment designed to capture both DNA strands (including DNA molecules that have damages ranging from minor single strand breaks to major damages on both strands).

[0323]All NGS data were analyzed on the same software pipeline with the same settings. Raw reads were filtered to remove duplicates, multiple mappers, improper pairs, and off-target reads. On average 75.4% reads were retained after filtering (FIG. 10A). For the reads that were removed, 71.8% were off-target reads, which were mapped to the human genome but outside of the target regions,21.6% were PCR duplicates, and the remaining reads were mapped to multiple sites of the genome or not mapped at all (FIG. 10B). No statistically significant difference was observed in all the specifications measures for the three technical replicates in this experiment, which indicates that library construction and RNA probe-based DNA double strand capturing pipeline is technically highly reproducible (FIGS. 10A, and 10B).

[0324]Next, the correlation between coverage efficiency and sequencing depth in NGS library with RNA probe-based DNA double strand capturing was evaluated. Filtered reads were randomly selected in 5 million read increments from 5 million to 50 million. The fractions of the retained on-target reads covering the depths of at least 10X, 20X, 50X, and 100X were plotted using randomly selected 5 to 50 million reads (FIG. 10C). 20 million reads could cover close to 90% of the target bases with no less than 10X depth. With 50million reads, over 90% target bases were covered by at least 20X. The efficiency of coverage is not only dependent on the efficiency of library construction but also dependent on the length of the sheared molecules that were initially incorporated into the pipeline. For the current study, the average length of sheared DNA molecule is 150bp. These real-time PCR results for the 298-gene panel indicated that enrichment efficiency of the library construction approach is not significantly biased by GC content (FIG. 8). Density plots were created to show GC content against normalized mean read depth for RNA probe-based DNA double strand capture WES study with normal tissue DNA (FIG. 11A), and DNA library WGS study with normal tissue DNA (without enrichment for whole exome, FIG. 11B).

[0325]To assess the impact of GC content on WES result, normalized mean read depth against GC content was plotted. There is a correlation between GC content and read depth in the WES experiment (FIG. 11A), and this bias is reduced in a WGS study (FIG. 11B). In this method, the mean read depth ratios of GC50%/GC20% =1.55, which is significantly lower than the ratio of 2.0 reported by numerous studies (Benjamini and Speed 2012, Meienberg, Zerjavic et al. 2015), which demonstrates a lower GC bias in this method.

[0326]EXAMPLE 4

[0327]Detection of SNVs

[0328]FIG. 9A shows that total number of SNVs detected at increasing read count thresholds. Sensitivity increases at higher read counts but quickly reaches a plateau with more than 80 million reads. FIG. 9B shows average SNV frequencies of normal tissue DNA measured by three approaches: a standard NGS approach where barcodes were directly trimmed off, a super read

based approach by barcoded single-stranded library based NGS without matching variants from both DNA strands (without the last step of the 4-step procedure), and a super read approach by barcoded single-stranded library based NGS matching the SNV on both strands (all steps in the 4-step procedure were performed). All three approaches were performed with RNA probe-based DNA double strand capture WES.

[0329]One of the most important goals of exome sequencing is to identify sequence variants that are disease-causing or of clinical significance. To evaluate the sensitivity and specificity of sequence variant identification performance of library construction and RNA probe-based DNA double strand capturing, a WES study was conducted with 100ng genomic DNA from a pair of normal and tumor tissue samples obtained from the same cancer patient. The same SNV calling pipeline was used for all data analysis in this study. Briefly, the normal DNA libraries created by library construction and RNA probe-based DNA double strand capturing method was sequenced and the data were analyzed using a standard data analysis pipeline, where the single-stranded barcodes were directly trimmed off, and 78,721 SNVs were detected from the exonic sequences of normal DNA sample at a read count of 30 million (error frequency $2.6X10^{-3}$, FIG. 9A). The total number of SNVs detected from 30 million reads of the normal tissue DNA is significantly higher than what was reported on other platforms (Clark, Chen et al. 2011). Next, further investigation was made to check if there is any bias in SNVs identified using the standard NGS data analysis workflow. Transition-transversion (ts/tv) ratio is routinely used to evaluate the specificity of new SNP calls. The ts/tv ratio on the target regions of WES was calculated to be 2.766, higher than the reported ts/tv ratios of 2.0-2.1 for WGS data. The ts/tv ratio in CCDS exonic regions as was then determined as 3.225, which falls into the range of 3.0~3.3 for reported exonic variations. The reason for RNA probe-based DNA double strand capture for whole exome sequencing to have a higher ts/tv ratio than reported WGS studies is because target regions of sequencing are enriched for exons, and only contain UTRs and short flanking sequences within introns.

[0330]The accuracy of mutations enriched by DNA based mutation calling was then examined. Following the 4-step data analysis procedure introduced in Materials and Methods, super reads were generated after Step 3). Steps 1~3 helped to reduce the mutation frequency by over two orders of magnitude from $2.6X10^{-3}$ down to $2.5X10^{-5}$ by removing most PCR related errors (FIG 9B). This result indicates that PCR related artificial mutations dramatically reduce NGS sequencing accuracy. To detect rare mutations, or even ultra-rare mutations using NGS, a correction for PCR errors is mandatory. As outlined in Step 4), attempts were then made to further reduce artificial errors of mutation calling by using the redundant sequence information offered by complementary DNA strands that were originally from the same DNA duplex molecule. These results indicated

that such procedure resulted in a single base mutation frequency of $1.6X10^{-6}$ (FIG. 9B). For any single base in the DNA sequences, the possibility of having the same artificial error on a paired position is $\frac{1}{3} X (2.5 X 10^{-5})^2 = 2.08 X 10^{-10}$, which is equivalent to one artificial error per $4.8 X 10^9$ nucleotides. This is the theoretical error rate for the pipeline. The total amount of DNA sequence data and the remaining amount of data after each step can be found in FIG. 14, where a stepwise drop of data amount is correlated to the increase of mutation calling stringency.

[0331]To determine the accuracy of variant detection by library construction and RNA probe-based DNA double strand capturing for clinically relevant mutations, the WES data generated from the normal and tumor tissue pair were analyzed side-by-side. For all assessed heterozygous exonic positions, the result was filtered through such 4-step procedure. The filtered result showed that for RNA probe-based DNA double strand capturing, WES study identified 97 sequence variants that were exclusively detected in tumor tissue DNA sample with $\geqslant$ 100X coverage at different fractions. 40 moderate- to high-abundance (> 5%) variants were subject to Sanger sequencing validation, and 38 were confirmed (FIGS. 15A-15E). Two variants failed to be validated where both allelic fractions were low and beyond the detection limit of Sanger sequencing. 57 sequence variants (with mutant allele fractions < 5%) were not subject to Sanger sequencing validation at all, due to the limited sensitivity of Sanger sequencing (Tsiatis, Norris-Kirby et al. 2010).

[0332]EXAMPLE 5

[0333]A protocol for RNA probe-based DNA double strand capture

[0334]Production of the RNA probes

[0335]The entire exome sequences for every human gene by sequence synthesis and molecular cloning based on Hg19 reference human genome sequences. In brief, exome sequences in 32,524 CCDS IDs containing -50bp and +50bp intronic sequences were cloned into pcDNA 6.2 vector. The total DNA sequences used to generate RNA probes cover a 72.6 Mb genome region, where all the exomes with their -50bp and +50bp flanking intronic sequences, as well as 5' and 3' UTRs for each gene were included.

[0336]Two clones for each target sequence were constructed, where a T7 promoter was inserted at the 5' end of the plus strand in the "+" clone and the 5' end of the minus strand in the "-" clone (FIG. 4B and FIGS. 7A and 7B).

[0337]Two pools of clones were established for any give number of genes following the rule that the two clones for the same DNA sequence are separated into two systems, where one system (FIG. 4B and FIGS. 7A and 7B, left panel, "+" Clone) produced the RNA probes targeting the plus strand of the DNA target, and the other system (FIG. 4B and FIGS. 7A and 7B, right panel "-" Clone) produced the RNA probes targeting the minus strand of the DNA target through in vitro transcription.

[0338]AmpliScribeTM T7 FlashTM Biotin-RNA Transcription kit is used for RNA probe production and amplification

[0339]Prepare the mix as following:

component	volume per reaction (µl)
Plasmid library with T7 promoter(100ng)	5.5
T7 flash buffer 10X	2
NTP/biotin-UTP premix	8
100mM DTT	2
RNase Inhibitor	0.5
AmpliScribeT7 Flash enzyme	2
Total	20

[0340]Mix well and incubate at 30°C for 4 hours.

[0341]Add 1µl DNaseI. Incubate at 37°C for 15 minutes.

[0342]Purify the RNA probes using 2×RNA AMPure beads. Elute into 80µl. You should have 150µg probe now.

[0343]sonication of the RNA products to generate 100-150nt RNA fragments as probes

[0344]Turn on BioRuptor and water bath (set to 3°C) at least 45 minutes before starting.

[0345]Place up to 1 µg of RNA adjusted to 57 µl with 1×TE buffer in a BioRuptor microtube.

[0346]Shear with below setting for a target size range of 100-150 nt:

Setting	value
Intensity	H
On:Off	30:30
Cycles	35

[0347]Hybridize the probes with the DNA library

[0348]Mix the following components as DNA library+block mix at room temperature:

4.3 µl of DNA library (150ng/ul)
3 µl of Human Cot-1 DNA (Life Technologies 15279-101)
3 µl of Salmon sperm (Life Technologies 15632-011)
0.7 µl Customized blocking oligos mix (1000µM), sequence shown below:
Blocking Oligo 1(as set forth in SEQ ID NO. 913)
Blocking Oligo 2 (as set forth in SEQ ID NO. 914)

[0349]Mix well by pipetting.

[0350]Transfer the DNA-library+block mix to 384-well PCR plate. Seal the plate will microAmp clear adhesive film (cat# 4306311 from ABI) for tight sealing.

[0351]Centrifuge the plate briefly to collect the liquid at the bottom of the well.

[0352]Run the following thermocycler program (with 105°C heated lid):

95°C for 5 minutes;
67.5°C forever

[0353]Prepare the Hybridization Buffer immediately after putting DNA-library+block mix in the thermocycler. Mix the following components at room temperature to prepare the Hybridization Buffer:

component	volume in 1 reaction (µl)
20×SSPE	12.5
0.5M EDTA	0.5
50×Denhardt's	5
10%SDS	6.5
Total	24.5

[0354]With the 384-well plate in the thermocycler at 67.5°C, transfer 24.5 µl of Hybridization Buffer into a new well of the plate, seal the plate with adhesive film.

[0355]Incubate the Hybridization Buffer at 67.5°C for at least 5 minutes (could be longer) while the RNA-Probe Library get prepared.

[0356]Prepare the RNA probes immediately after putting the Hybridization Buffer in the thermocycler. Mix the following components on ice to prepare the RNA probes:

component	volume in 1 reaction (µl)
RNA probes from "+" clone transcripts (800ng/µl)	2.5
RNA probes from "-" clone transcripts (800ng/µl)	2.5
RNase Inhibitor (20U/ µl)	0.5
Nuclease-free water	1.5
Total	7

[0357]With the 384-well plate in the thermocycler at 67.5°C, transfer 7 µl of RNA probes into a new well of the plate, seal the plate with adhesive film.

[0358]Incubate the RNA probes at 75.5°C for 2 minutes with the heated lid.

[0359]Open the lid and maintain the plate at 75.5°C. Take 13µl of pre-heated Hybridization Buffer and add it to the RNA probes.

[0360]Transfer 10 µl of DNA-library+block mix to the RNA probes.

[0361]Mix well by slowly pipetting up and down several times. The hybridization mixture should be ~30µl.

[0362]Seal the well with double adhesive film.

[0363]Incubate the hybridization mixture for 12~48 hours at 67.5°C with heated lid at 105°C.

[0364]Extracting the hybridized RNA-DNA molecules

[0365]Wash 50 µl Dynal MyOne Streptavidin C1 magnetic beads (ThermoFisher Scientific) with 200 µl Binding buffer (1M NaCl, 10 mM Tris-HCl, pH 7.5, and 1 mM EDTA) three times in 1.5ml microfuge tube and resuspend in 200 µl binding buffer.

[0366]Add the entire hybridization mixture directly from the thermocycler to the bead solution, and invert the tube to mix several times.

[0367]Rotate the hybridization mixture /bead solution 360 deg. for 15 minutes at room temperature.

[0368]After binding, the beads are separated from the solution on a Dynal magnetic separator and the supernatant is removed.

[0369]Washing Procedure

[0370]Wash the beads for 15 minutes at room temperate in 500µl wash buffer I (1×SSC/0.1% SDS).

[0371]Wash the beads for 15 minutes at 67.5°C on a heating block with shaking, three times in 500µl 67.5°C pre-heated wash buffer II (0.1×SSC/0.1% SDS).

[0372]Mix the beads in 50µl 0.1M NaOH at room temperature for 10 minutes.

[0373]Transfer the supernatant to a new tube.

[0374]Add 50µl of Neutralizing buffer (1M Tris-HCl, pH 7.5).

[0375]Neutralized DNA is desalted and concentrated by AMPure beads with ratio 1:1 (beads:sample), elute in 20ul 1×TE buffer.

[0376]Post-Capture amplification

[0377]PCR mix contains: (per reaction)

Captured DNA	20 ul
Water	65 ul
DMSO	2.5 ul
5X Phusion Buffer	10 ul
10mM dNTPs	1 ul
Index PE primer II	0.25 ul
PE primer I	0.25 ul
HotStart Phusion	1 ul

[0378]mix well

[0379]Amplification conditions:

Step1: 1 cycle	98 °C	1 minute
Step 2: 14 cycles of	98 °C	10 seconds
	65 °C	30 seconds
	72 °C	30 seconds
Step 3: 1 cycle	72 °C	5 minutes
Step 4: 1 cycle	4 °C	hold

[0380]The PCR is done in two wells for each sample, 50ul each. Then the amplified PCR product was purified using AMPure beads with ratio 1:1 (beads:sample), elute in 30ul 1×TE buffer.

[0381]Use Qubit to quantify yield. You will have ~20ng/ul in general.

REFERENCE

Benjamini, Y. and T. P. Speed (2012). "Summarizing and correcting the GC content bias in high-throughput sequencing." Nucleic Acids Res 40(10): e72.

Clark, M. J., et al. (2011). "Performance comparison of exome DNA sequencing technologies." Nat Biotechnol 29(10): 908-914.

McKenna, A., et al. (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res 20(9): 1297-1303.

Meienberg, J., et al. (2015). "New insights into the performance of human whole-exome capture platforms." Nucleic Acids Res 43(11): e76.

Tsiatis, A. C., et al. (2010). "Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications." J Mol Diagn 12(4): 425-432.

Untergasser, A., et al. (2012). "Primer3--new capabilities and interfaces." Nucleic Acids Res 40(15): e115.

Van Allen, E. M., et al. (2014). "Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine." Nat Med 20(6): 682-688.

CLAIMS

1. A method for enriching at least one target nucleic acid sequence from a biological sample, comprising:

preparing at least one pair of RNA probe sets, each pair comprising a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a duplex segment in each of the at least one target nucleic acid sequence, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with an immobilization portion configured to allow immobilization onto a solid support; and

capturing each strand of the at least one target nucleic acid sequence from the biological sample through hybridization of both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets respectively with the two antiparallel strands of the duplex segment in the each of the at least one target nucleic acid sequence and through immobilization onto the solid support via the immobilization portion labelled onto each RNA probe in any of the first RNA probe set and the second RNA probe set.

2. The method of Claim 1, wherein one or more of the at least one pair of RNA probe sets is prepared through chemical synthesis, and the preparing at least one pair of RNA probe sets comprises:

performing chemical synthesis reactions to thereby obtain the one or more of the at least one pair of RNA probe sets.

3. The method of Claim 2, wherein each RNA probe in each of the one or more of the at least one pair of RNA probe sets is labelled with the immobilization portion during the chemical synthesis reactions.

4. The method of Claim 3, wherein the immobilization portion is covalently attached onto the solid support during the chemical synthesis reactions.

5. The method of Claim 2, wherein each RNA probe in each of the one or more of the at least one pair of RNA probe sets is labelled with the immobilization portion after the chemical synthesis reactions, and the preparing at least one pair of RNA probe sets further comprises:

performing labelling reactions such that each RNA probe in each of the one or more of the at least one pair of RNA probe sets is labelled with the immobilization portion.

6. The method of Claim 1, wherein one or more of the at least one pair of RNA probe sets is prepared through transcription, and the preparing at least one pair of RNA probe sets comprises:

performing transcription reactions to thereby obtain the one or more of the at least one pair of RNA probe sets.

7. The method of Claim 6, wherein the performing transcription reactions to thereby obtain the one or more of the at least one pair of RNA probe sets comprises:

performing transcription reactions such that each RNA probe in any of the one or more of the at least one pair of RNA probe sets is labelled with the immobilization portion during each of the transcription reactions.

8. The method of Claim 7, wherein each of the transcription reactions is performed in presence of NTPs labelled with the immobilization portion, where the NTPs comprises at least one of ATPs, UTPs, GTPs, and CTPs.

9. The method of Claim 8, wherein the NTPs labelled with the immobilization portion comprise biotin-labelled UTPs.

10. The method of Claim 9, wherein the biotin-labelled UTPs have a relative molar percentage of 2%-100% in all UTPs present in each of the transcription reactions.

11. The method of Claim 6, wherein each RNA probe in any of the one or more of the at least one pair of RNA probe sets is labelled with the immobilization portion after each of the transcription reactions, and the performing transcription reactions to thereby obtain the one or more of the at least one pair of RNA probe sets comprises:

performing the transcription reactions; and

performing a labelling.

12. The method of Claim 6, wherein the performing the transcription reactions comprises:

providing a plurality of DNA vectors, comprising at least one pair of DNA vectors, each pair comprising a first DNA vector and a second DNA vector configured to respectively allow transcription of a first RNA molecule and a second RNA molecule respectively targeting two antiparallel strands of a duplex segment in each of the one or more of the at least one target nucleic acid sequence; and

performing the transcription reactions over the plurality of DNA vectors.

13. The method of Claim 12, wherein each of the plurality of DNA vectors comprises a promoter, selected from one of a T3 promoter, a T7 promoter, or a SP6 promoter.

14. The method of Claim 12, wherein in the performing transcription reactions over the plurality of DNA vectors, at least one of the transcription reactions is performed *in vitro*.

15. The method of Claim 12, wherein in the performing transcription reactions over the plurality of DNA vectors, at least one of the transcription reactions is performed *in vivo*.

16. The method of Claim 12, wherein the performing transcription reactions over the plurality of DNA vectors comprises:

pooling the plurality of DNA vectors to obtain at least two DNA vector pools, such that the first DNA vector and the second DNA vector in the each pair of DNA vectors are not in a same DNA vector pool; and

performing a transcription reaction over each of the at least two DNA vector pools respectively to obtain RNA molecules corresponding to the each of the at least two DNA vector pools.

17. The method of Claim 16, further comprising, after the performing a transcription reaction over each of the at least two DNA vector pools respectively and prior to the performing a labelling:

performing a fragmentation reaction to the RNA molecules corresponding to the each of the at least two DNA vector pools.

18. The method of Claim 12, wherein the performing transcription reactions over the plurality of DNA vectors comprises:

performing a transcription reaction over each of the plurality of DNA vectors to thereby obtain an RNA molecule corresponding thereto.

19. The method of Claim 11, further comprising, after the performing a transcription reaction over each of the plurality of DNA vectors and prior to the performing a labelling:

pooling the RNA molecule corresponding to each of the plurality of DNA vectors to obtain at least two RNA pools, such that a pair of RNA molecules respectively targeting two antiparallel strands of a duplex segment in any one of the one or more of the at least one target nucleic acid sequence are not in a same RNA pool; and

performing a fragmentation reaction to each of the at least two RNA pools respectively.

20. The method of Claim 11, further comprising, after the performing a transcription reaction over each of the plurality of DNA vectors and prior to the performing a labelling:

performing a fragmentation reaction to the RNA molecule corresponding to each of the plurality of DNA vectors respectively to obtain fragmented RNA molecules corresponding to each of the plurality of DNA vectors; and

pooling the fragmented RNA molecules corresponding to each of the plurality of DNA vectors such that a pair of fragmented RNA molecules respectively targeting two antiparallel strands of a duplex segment in any one of the one or more of the at least one target nucleic acid sequence are not in a same RNA probe set.

21. The method of Claim 11, wherein the performing a labelling comprises:

performing ligation reactions such that an immobilization portion-labelled nucleotide is ligated to one terminus, or to a middle, of each RNA probe in each of the at least one pair of RNA probe sets.

22. The method of Claim 21, wherein in each of the ligation reactions, a 5' phosphate terminus of a biotin-labeled nucleotide is ligated to a 3' hydroxyl terminus of each RNA probe in each of the at least one pair of RNA probe sets.

23. The method of Claim 22, wherein each of the ligation reactions is performed by means of an RNA ligase.

24. The method of Claim 23, wherein the RNA ligase comprises at least one of T4 RNA ligase or CircLigase RNA ligase.

25. The method of Claim 1, wherein the preparing at least one pair of RNA probe sets comprises:

performing direct transcription on the solid support to thereby obtain the one or more of the at least one pair of RNA probe sets.

26. The method of Claim 1, wherein the capturing each strand of the at least one target nucleic acid sequence from the biological sample comprises:

contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample; and

immobilizing the at least one target nucleic acid sequence on the solid support.

27. The method of Claim 26, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets has a length of about 100 - 150 nt, wherein:

the contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample is performed at a temperature of about 62-70°C and for about 6-24 hours.

28. The method of Claim 27, wherein the contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample is performed at a temperature of about 67.5°C.

29. The method of Claim 26, wherein the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets respectively target a different portion of the duplex segment in the each of the at least one target nucleic acid sequence, and the contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample comprises:

contacting both of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a single hybridization reaction.

30. The method of Claim 26, wherein the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets respectively target a substantially same portion of the duplex segment in the each of the at least one target nucleic acid sequence, and the contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample comprises at least one round of:

contacting one of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a first hybridization reaction; and

contacting another of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a second hybridization reaction.

31. The method of Claim 26, wherein the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets respectively target a substantially same portion of the duplex segment in the each of the at least one target nucleic acid sequence, and the contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample comprises at least one round of:

separately contacting the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a third hybridization reaction and a fourth hybridization reaction, respectively; and

combining the third hybridization reaction and the fourth hybridization reaction to thereby allow a fifth hybridization reaction to proceed.

32. The method of Claim 26, wherein one or more of the at least one target nucleic acid sequence in the biological sample are each in a polynucleotide containing at least one un-targeted sequence, wherein the method further comprises, prior to the contacting both the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in the biological sample:

contacting at least one blocking oligo with the at least one target nucleic acid sequence such that the at least one blocking oligo respectively hybridizes with, and thereby blocks, at least one strand of each of the at least one un-targeted sequence in the polynucleotide.

33. The method of Claim 32, wherein the at least one un-targeted sequence in the polynucleotide comprises a first adaptor sequence and a second adaptor sequence flanking each of the one or more of the at least one target nucleic acid sequence, wherein:

the at least one blocking oligo is configured to respectively block one strand of the first adaptor sequence and one strand of the second adaptor sequence in the polynucleotide.

34. The method of Claim 33, wherein the at least one blocking oligo is configured to respectively block both two antiparallel strands of the first adaptor sequence and both two antiparallel strands of the second adaptor sequence in the polynucleotide.

35. The method of Claim 33, wherein the at least one blocking oligo comprises a first blocking oligo set and a second blocking oligo set, each comprising one or more blocking oligo, configured to respectively block two antiparallel strands of one of the first adaptor sequence and the second

adaptor sequence in the polynucleotide, wherein the contacting at least one blocking oligo with the at least one target nucleic acid sequence comprises:

contacting one of the first blocking oligo set and the second blocking oligo set with the at least one target nucleic acid sequence; and

contacting another of the first blocking oligo set and the second blocking oligo set with the at least one target nucleic acid sequence.

36. The method of Claim 1, wherein the capturing each strand of the at least one target nucleic acid sequence from the biological sample comprises:

conjugating the at least one pair of RNA probe sets on the solid support via the immobilization portion labelled onto each RNA probe in any of the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets to thereby obtain at least one pair of solid support-conjugated RNA probe sets, each pair comprising a solid support-conjugated first RNA probe set and a solid support-conjugated second RNA probe set; and

contacting both the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence.

37. The method of Claim 36, wherein the contacting both the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence comprises:

contacting one of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in a sixth hybridization reaction; and

contacting another of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in a seventh hybridization reaction.

38. The method of Claim 36, wherein the contacting both the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence comprises:

separately contacting the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in an eighth hybridization reaction and a ninth hybridization reaction, respectively; and

combining the eighth hybridization reaction and the ninth hybridization reaction to thereby allow a tenth hybridization reaction to proceed.

39. The method of Claim 36, wherein the contacting both the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence comprises:

contacting both of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in the each of the at least one pair of solid support-conjugated RNA probe sets with the at least one target nucleic acid sequence in a single hybridization reaction.

40. The method of Claim 1, wherein the immobilization portion is configured to be able to form a stable non-covalent binding with a coupling partner conjugated onto surface of the solid support.

41. The method of Claim 40, wherein:

the immobilization portion comprises a biotin moiety; and

the coupling partner conjugated onto surface of the solid support comprises at least one of streptavidin, avidin, or an anti-biotin antibody.

42. The method of Claim 1, wherein the immobilization portion is configured to be able to form a covalent connection with a coupling partner conjugated onto surface of the solid support.

43. The method of Claim 1, wherein the solid support comprises at least one of a magnetic bead, a filter, a resin bead, a nanosphere, a plastic surface, a microtiter plate, a glass surface, a slide, a membrane, a microfluidic channel, a chip, or a matrix.

44. The method of Claim 1, further comprising:

eluting out the at least one target nucleic acid sequence from the solid support.

45. A kit for enriching at least one target nucleic acid sequence from a biological sample utilizing the method according to any one of Claims 1-44, comprising:

at least one pair of RNA probe sets, each pair comprising a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a duplex segment in each of the at least one target nucleic acid sequence, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with an immobilization portion; and

a solid support labelled with a coupling partner on a surface thereof, wherein the coupling partner is configured to be able to form a secure coupling to the immobilization portion to thereby allow immobilization of each RNA probe in any of the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets onto the solid support.

46. The kit of Claim 45, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets has a length of about 100 - 150 nt,

47. The kit of Claim 45, wherein the first RNA probe set and the second RNA probe set in the each of the at least one pair of RNA probe sets are configured to respectively target different portions of the duplex segment in the each of the at least one target nucleic acid sequence.

48. The kit of Claim 45, wherein the immobilization portion comprises a biotin moiety.

49. The kit of Claim 48, wherein the solid support is labelled with at least one of streptavidin, avidin, or an anti-biotin antibody.

50. The kit of Claim 45, wherein the solid support comprises at least one of a magnetic bead, a filter, a resin bead, a nanosphere, a plastic surface, a microtiter plate, a glass surface, a slide, a membrane, a microfluidic channel, a chip or a matrix.

51. The kit of Claim 45, further comprising an apparatus having a working surface as the solid support, wherein the first RNA probe set and the second RNA probe set in each pair are respectively conjugated onto the working surface, arranged such that each RNA probe in the solid support-conjugated first RNA probe set does not substantially interact with each RNA probe in the solid support-conjugated second RNA probe set.

52. The kit of Claim 51, wherein the apparatus is one of a column, a microfluidic channel, or a chip.

53. The kit of Claim 51, wherein the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in each pair are respectively arranged at at least one pair of two different regions of the working surface of the apparatus.

54. The kit of Claim 53, wherein the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in each pair are alternately arranged at more than one pairs of two different regions of the working surface of the apparatus.

55. The kit of Claim 51, wherein the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set in each pair are mixedly arranged on the working surface of the apparatus, configured such that each RNA probe from the first RNA probe set has a relatively large distance to each RNA probe from the second RNA probe set to thereby substantially prevent an interaction therebetween.

56. The kit of Claim 51, wherein the apparatus is configured to allow the biological sample to flow through the working surface for more than one round.

57. The kit of Claim 45, wherein one or more of the at least one target nucleic acid sequence are each in a polynucleotide comprising at least one un-targeted sequences, wherein the kit further comprises:

at least one blocking oligo, configured to respectively hybridize with, and to thereby block, at least one strand of each of the at least one un-targeted sequence in the polynucleotide.

58. The kit of Claim 57, wherein the at least one un-targeted sequence in the polynucleotide comprises a first adaptor sequence and a second adaptor sequence flanking each of the one or more of the at least one target nucleic acid sequence, wherein:

the at least one blocking oligo is configured to respectively block one strand of the first adaptor sequence and one strand of the second adaptor sequence in the polynucleotide.

59. The kit of Claim 58, wherein the at least one blocking oligo is configured to respectively block both two antiparallel strands of the first adaptor sequence and both two antiparallel strands of the second adaptor sequence in the polynucleotide.

60. The kit of Claim 59, wherein the at least one blocking oligo comprises a first blocking oligo set and a second blocking oligo set, configured to respectively block two antiparallel strands of one of the first adaptor sequence and the second adaptor sequence in the polynucleotide.

61. The kit of Claim 60, wherein the first blocking oligo set and the second blocking oligo set are configured to target two different portions within the one of the first adaptor sequence and the second adaptor sequence in the polynucleotide.

62. A kit for enriching at least one target nucleic acid sequence from a biological sample utilizing the method according to any one of Claims 1-42, comprising:

a plurality of DNA vectors, comprising at least one pair of DNA vectors, each pair comprising a first DNA vector and a second DNA vector configured, via transcription thereover, to respectively obtain a first RNA probe set and a second RNA probe set targeting respectively two antiparallel strands of a duplex segment in each of the at least one target nucleic acid sequence;

NTPs comprising each of ATPs, UTPs, GTPs, and CTPs;

immobilization portion-labelled NTPs, wherein NTPs comprises at least one of ATPs, UTPs, GTPs, and CTPs; and

a solid support labelled with a coupling partner on a surface thereof, wherein the coupling partner is configured to be able to form a secure coupling to the immobilization portion.

63. The kit of Claim 62, wherein the immobilization portion comprises a biotin moiety.

64. The kit of Claim 63, wherein the NTPs labelled with the immobilization portion comprise biotin-labelled UTPs.

65. The kit of Claim 64, wherein the biotin-labelled UTPs have a relative molar percentage of 2%-100% among all UTPs in the kit.

66. The kit of Claim 63, wherein the solid support is labelled with at least one of streptavidin, avidin, or an anti-biotin antibody.

67. The kit of Claim 62, wherein the solid support comprises at least one of a magnetic bead, a filter, a resin bead, a nanosphere, a plastic surface, a microtiter plate, a glass surface, a slide, a membrane, a microfluidic channel, a chip or a matrix.

68. The kit of Claim 62, further comprising an RNA ligase, configured to ligate a 3' hydroxyl terminus of each RNA probe in any of the first RNA probe set and the second RNA probe set generated from each of the at least one pair of DNA vectors with one of the immobilization portion -labeled NTPs.

69. The method of Claim 68, wherein the RNA ligase comprises at least one of T4 RNA ligase or CircLigase RNA ligase.

70. The kit of Claim 62, wherein each of the plurality of DNA vectors comprises:

a DNA template, comprising a sequence corresponding to one of two antiparallel strands of a duplex segment in each of the at least one target nucleic acid sequence; and

a promoter, configured to initiate a transcription reaction of the DNA template in a presence of an RNA polymerase compatible with the promoter.

71. The kit of Claim 70, wherein the promoter is selected from one of a T3 promoter, a T7 promoter, a SP6 promoter, wherein:

the kit further comprises a T3 RNA polymerase, a T7 RNA polymerase, or a SP6 RNA polymerase, corresponding to the promoter in each of the plurality of DNA vectors.

72. The kit of Claim 70, further comprising cells or viruses containing the RNA polymerase compatible with the promoter in each of the plurality of DNA vectors.

73. The kit of Claim 72, further comprising cells containing an RNA polymerase compatible with the promoter in each of the plurality of DNA vectors, wherein the cells comprise at least one of a bacterial cell line, a yeast cell line, or a mammalian cell line.

74. The kit of Claim 62, wherein each of the plurality of DNA vectors is a double-stranded DNA vector or a single-stranded DNA vector.

75. The kit of Claim 62, wherein one or more of the at least one target nucleic acid sequence are each in a polynucleotide comprising at least one un-targeted sequences, wherein the kit further comprises:

at least one blocking oligo, configured to respectively hybridize with, and to thereby block, at least one strand of each of the at least one un-targeted sequence in the polynucleotide.

76. The kit of Claim 75, wherein the at least one un-targeted sequence in the polynucleotide comprises a first adaptor sequence and a second adaptor sequence flanking each of the one or more of the at least one target nucleic acid sequence, wherein:

the at least one blocking oligo is configured to respectively block one strand of the first adaptor sequence and one strand of the second adaptor sequence in the polynucleotide.

77. The kit of Claim 76, wherein the at least one blocking oligo is configured to respectively block two antiparallel strands of the first adaptor sequence and both two antiparallel strands of the second adaptor sequence in the polynucleotide.

78. The kit of Claim 77, wherein the at least one blocking oligo comprises a first blocking oligo set and a second blocking oligo set, configured to respectively block two antiparallel strands of one of the first adaptor sequence and the second adaptor sequence in the polynucleotide.

79. The kit of Claim 78, wherein the first blocking oligo set and the second blocking oligo set are configured to target two different portions within the one of the first adaptor sequence and the second adaptor sequence in the polynucleotide.

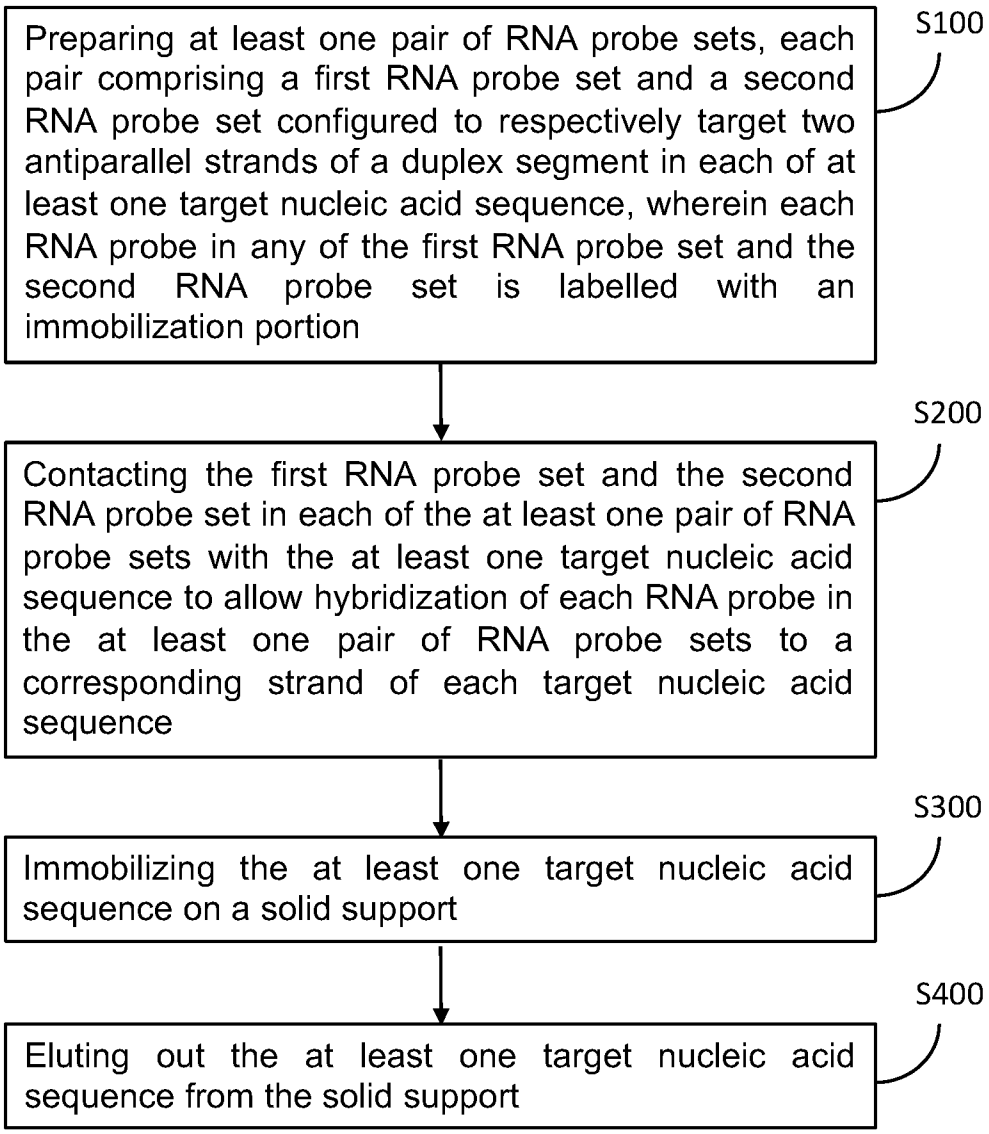
Preparing at least one pair of RNA probe sets, each pair comprising a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a duplex segment in each of at least one target nucleic acid sequence, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with an immobilization portion
S100
Contacting the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence to allow hybridization of each RNA probe in the at least one pair of RNA probe sets to a corresponding strand of each target nucleic acid sequence
S200
Immobilizing the at least one target nucleic acid sequence on a solid support
S300
Eluting out the at least one target nucleic acid sequence from the solid support
S400

FIG. 1A

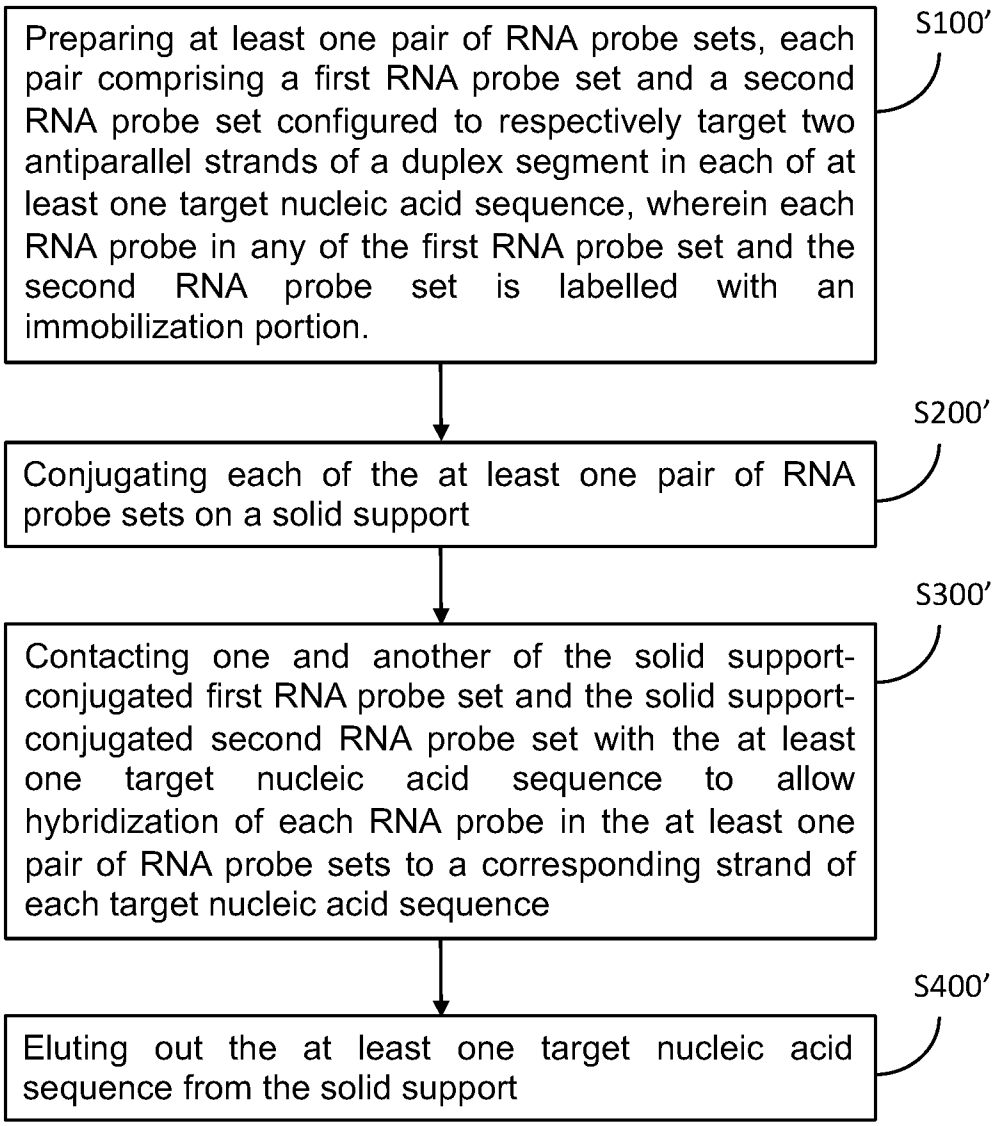
Preparing at least one pair of RNA probe sets, each pair comprising a first RNA probe set and a second RNA probe set configured to respectively target two antiparallel strands of a duplex segment in each of at least one target nucleic acid sequence, wherein each RNA probe in any of the first RNA probe set and the second RNA probe set is labelled with an immobilization portion.
S100'
Conjugating each of the at least one pair of RNA probe sets on a solid support
S200'
Contacting one and another of the solid support-conjugated first RNA probe set and the solid support-conjugated second RNA probe set with the at least one target nucleic acid sequence to allow hybridization of each RNA probe in the at least one pair of RNA probe sets to a corresponding strand of each target nucleic acid sequence
S300'
Eluting out the at least one target nucleic acid sequence from the solid support
S400'

FIG. 1B

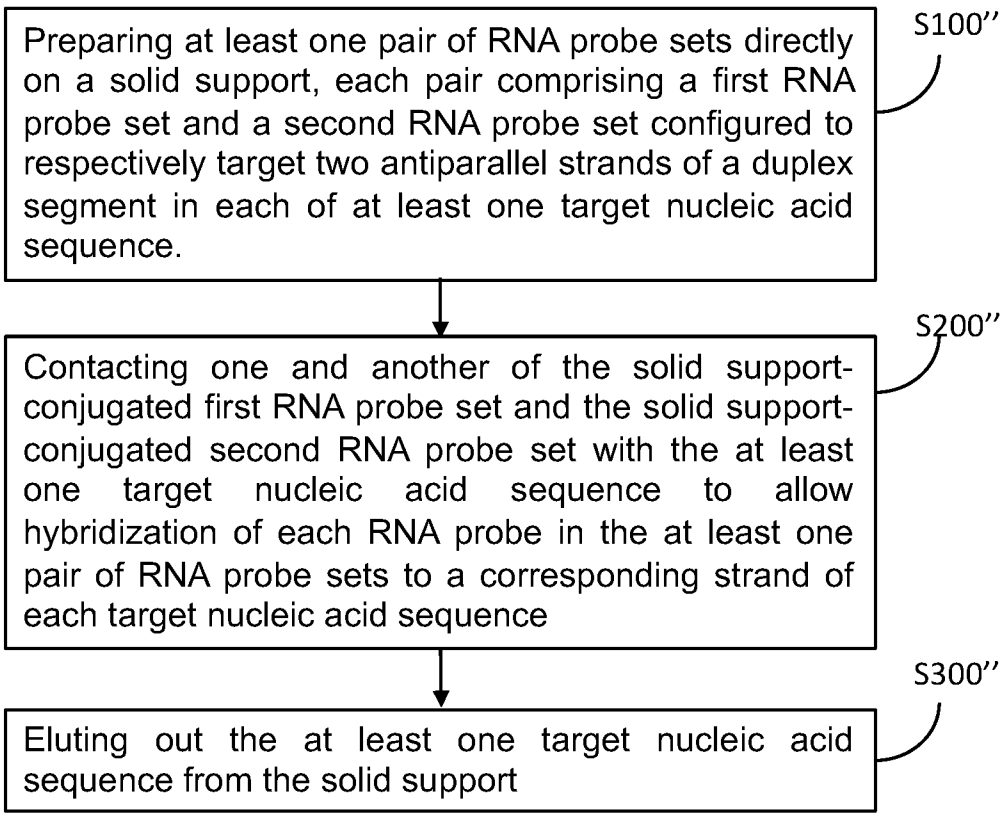

FIG. 1C

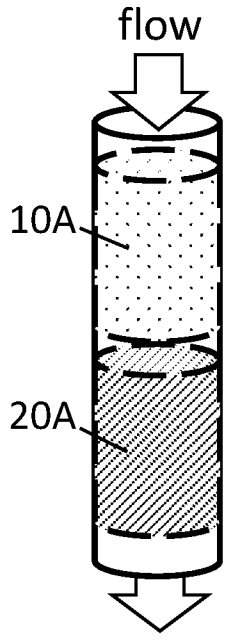

FIG. 1D

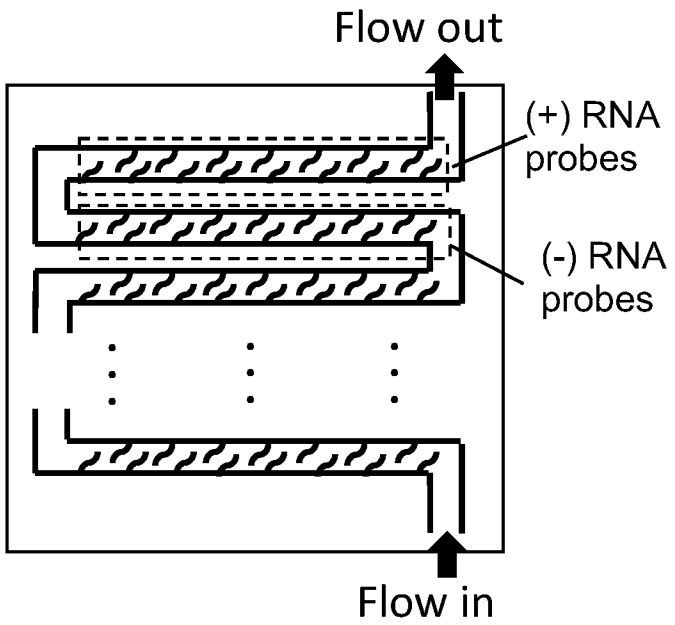

FIG. 1E

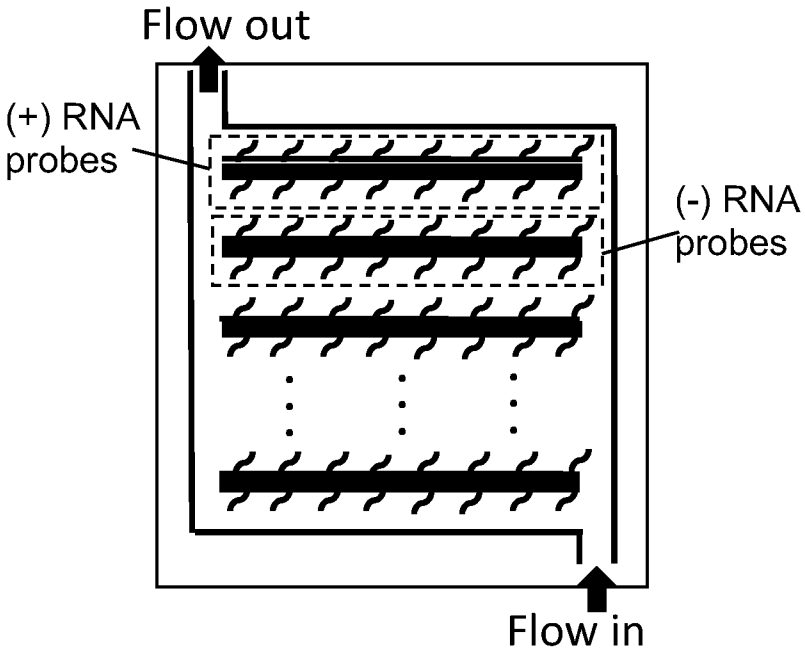
Flow out
(+) RNA probes
(-) RNA probes
Flow in

FIG. 1F

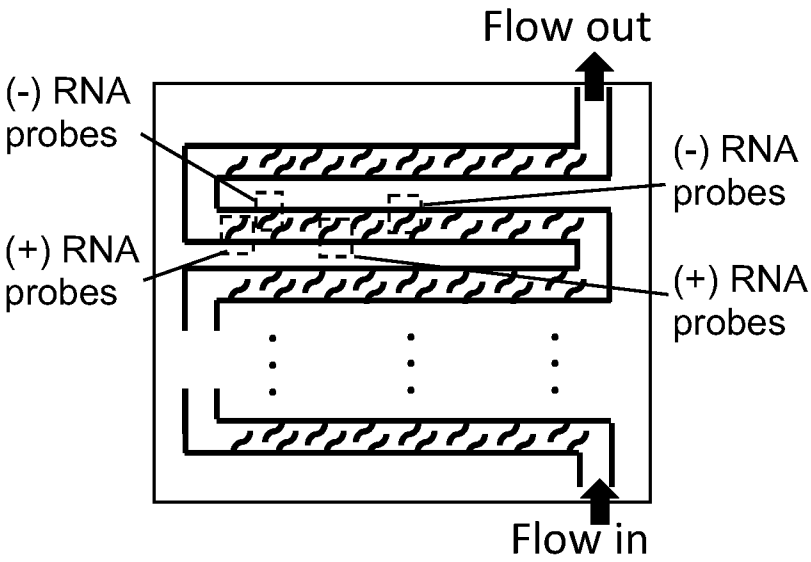
Flow out
(-) RNA probes
(-) RNA probes
(+) RNA probes
(+) RNA probes
Flow in

FIG. 1G

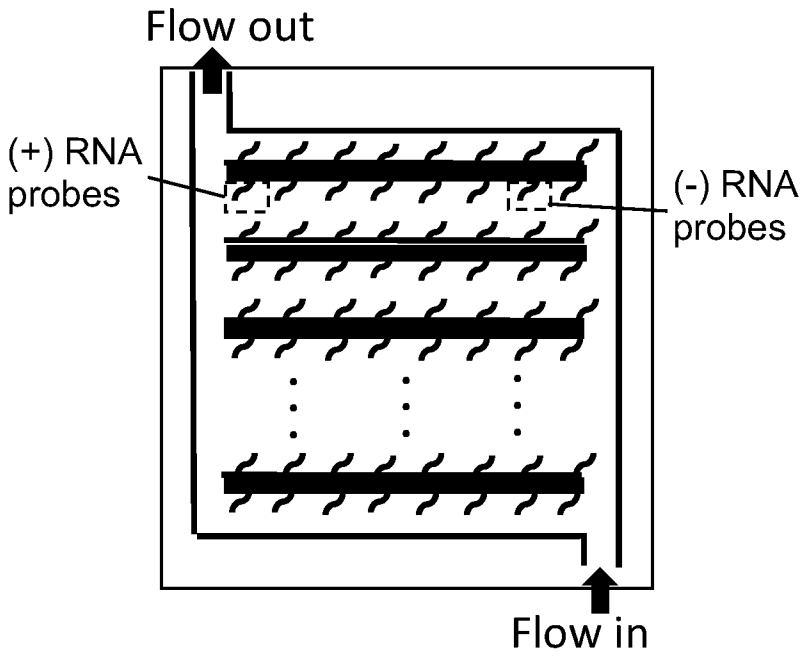

FIG. 1H

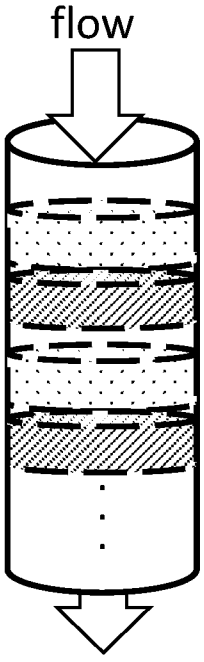

FIG. 1I

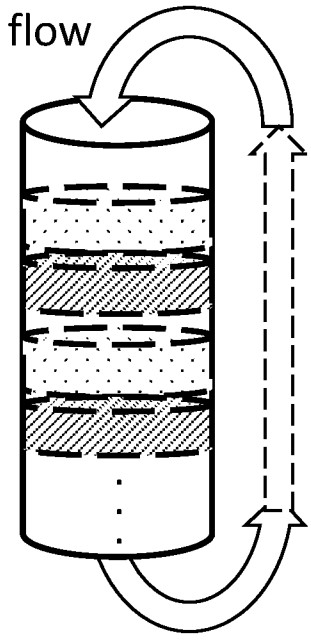

FIG. 1J

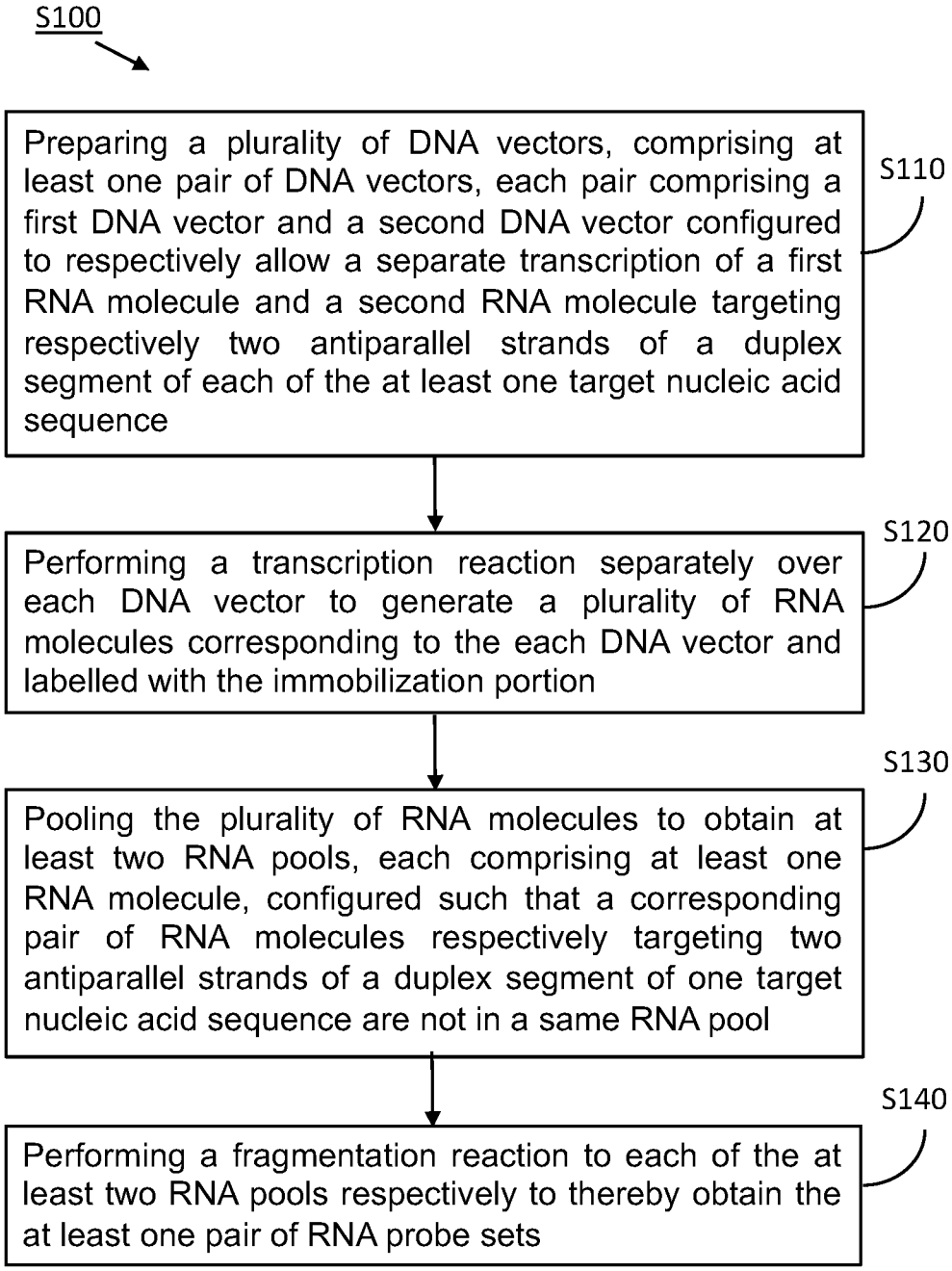
S100
Preparing a plurality of DNA vectors, comprising at least one pair of DNA vectors, each pair comprising a first DNA vector and a second DNA vector configured to respectively allow a separate transcription of a first RNA molecule and a second RNA molecule targeting respectively two antiparallel strands of a duplex segment of each of the at least one target nucleic acid sequence
S110
Performing a transcription reaction separately over each DNA vector to generate a plurality of RNA molecules corresponding to the each DNA vector and labelled with the immobilization portion
S120
Pooling the plurality of RNA molecules to obtain at least two RNA pools, each comprising at least one RNA molecule, configured such that a corresponding pair of RNA molecules respectively targeting two antiparallel strands of a duplex segment of one target nucleic acid sequence are not in a same RNA pool
S130
Performing a fragmentation reaction to each of the at least two RNA pools respectively to thereby obtain the at least one pair of RNA probe sets
S140

FIG. 2A

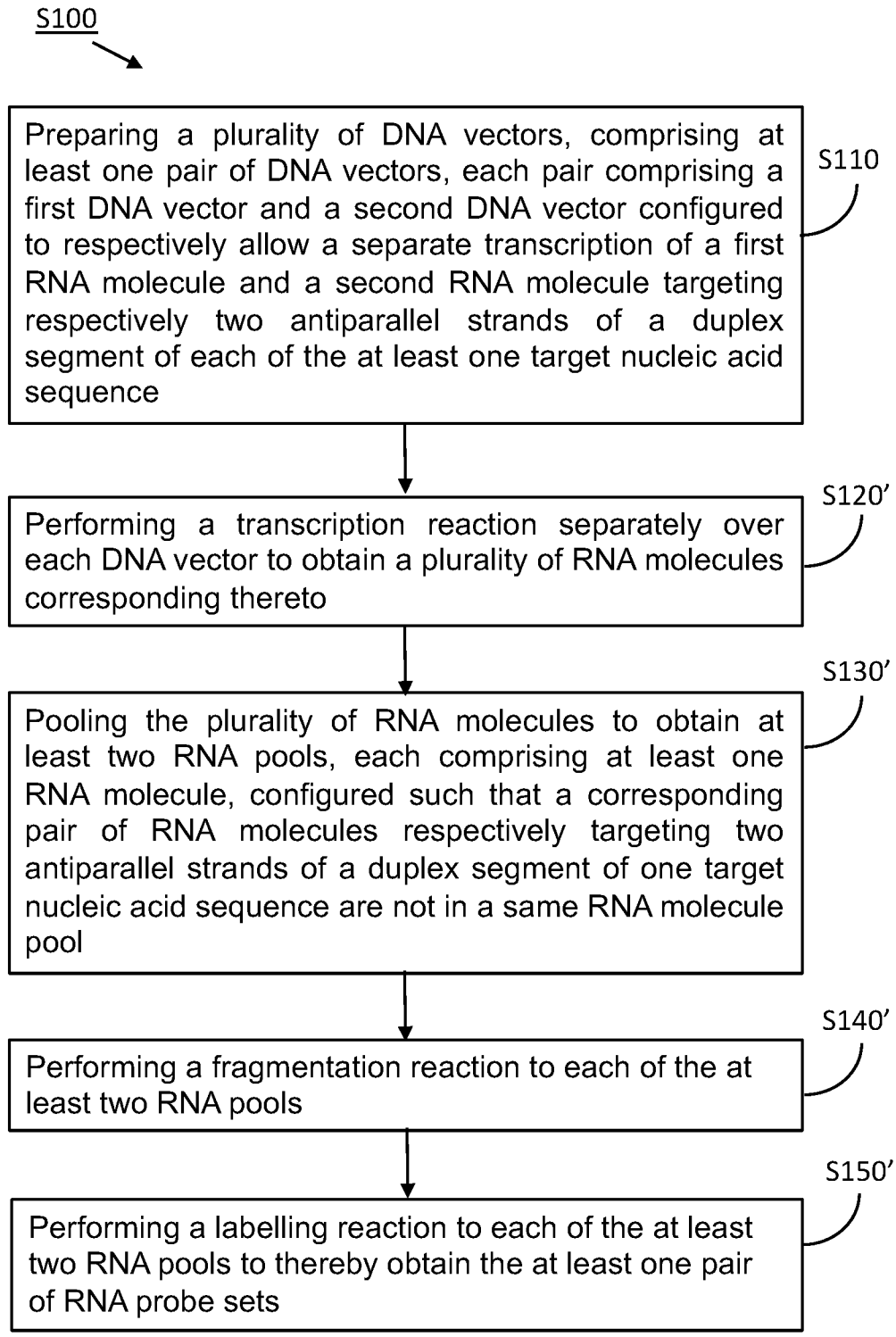
S100
Preparing a plurality of DNA vectors, comprising at least one pair of DNA vectors, each pair comprising a first DNA vector and a second DNA vector configured to respectively allow a separate transcription of a first RNA molecule and a second RNA molecule targeting respectively two antiparallel strands of a duplex segment of each of the at least one target nucleic acid sequence
S110
Performing a transcription reaction separately over each DNA vector to obtain a plurality of RNA molecules corresponding thereto
S120'
Pooling the plurality of RNA molecules to obtain at least two RNA pools, each comprising at least one RNA molecule, configured such that a corresponding pair of RNA molecules respectively targeting two antiparallel strands of a duplex segment of one target nucleic acid sequence are not in a same RNA molecule pool
S130'
Performing a fragmentation reaction to each of the at least two RNA pools
S140'
Performing a labelling reaction to each of the at least two RNA pools to thereby obtain the at least one pair of RNA probe sets
S150'

FIG. 2B

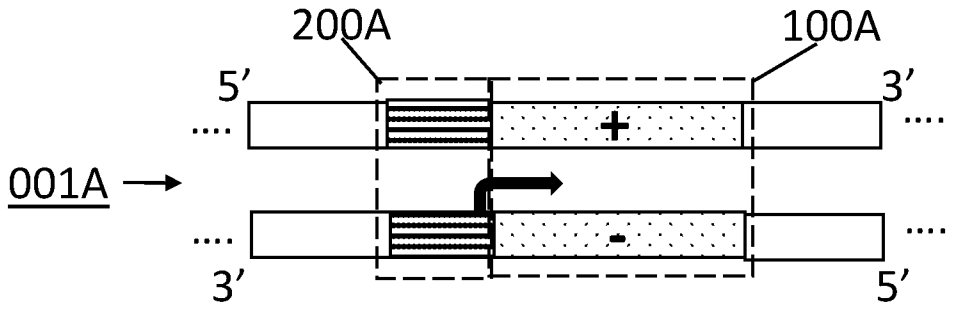

FIG. 3A

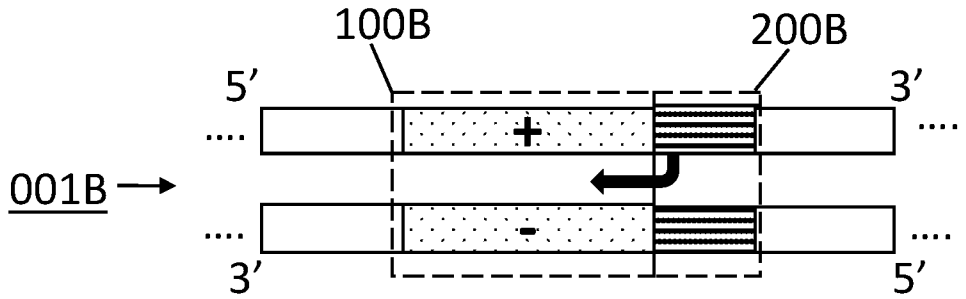

FIG. 3B

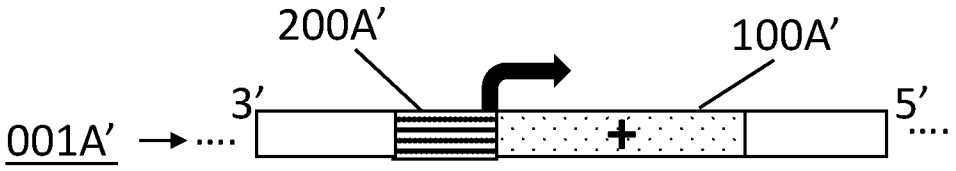

FIG. 3C

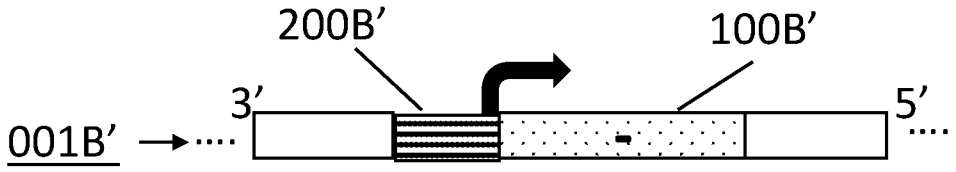

FIG. 3D

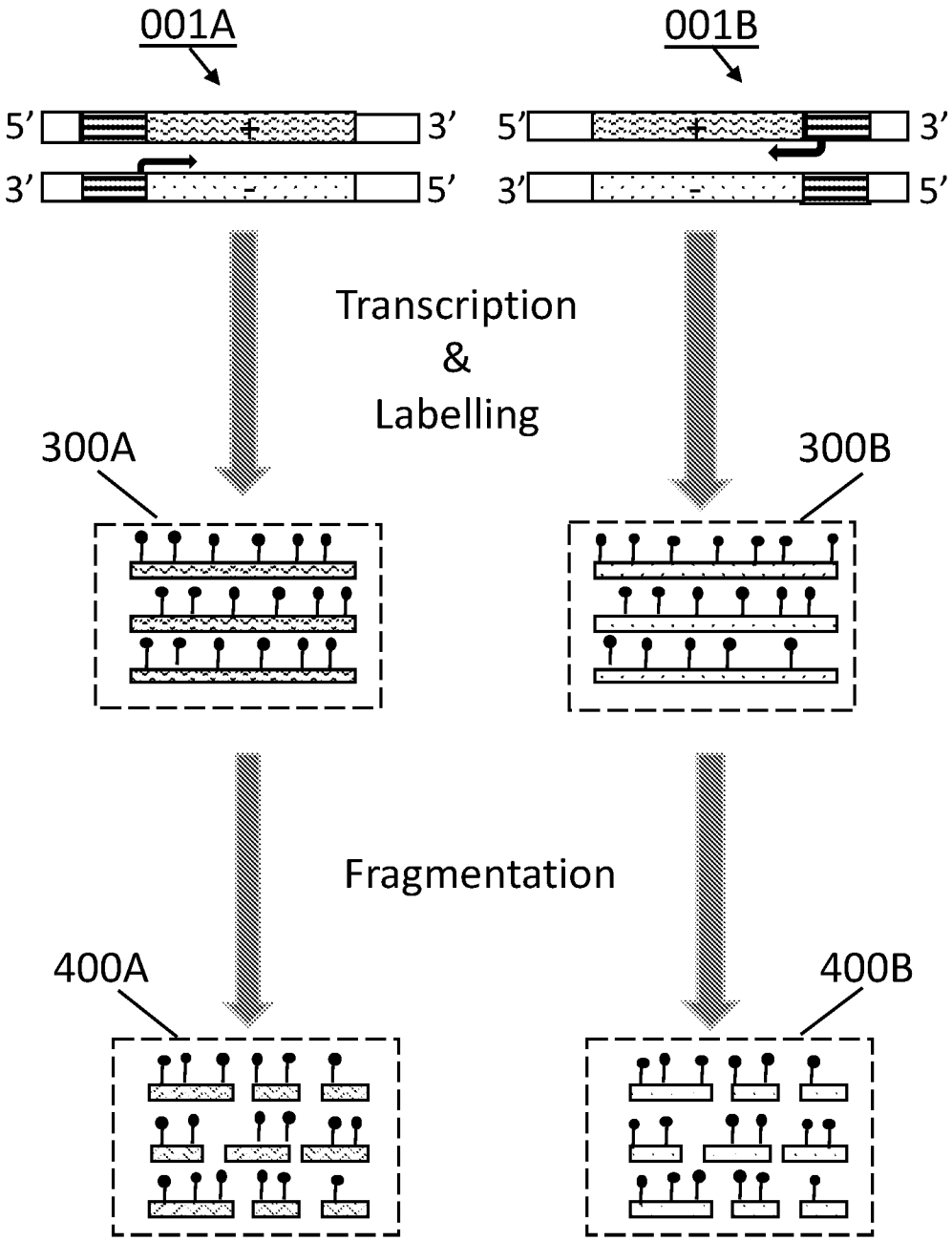
001A
001B
5'
3'
3'
5'
5'
3'
3'
5'
Transcription
&
Labelling
300A
300B
Fragmentation
400A
400B

FIG. 4A

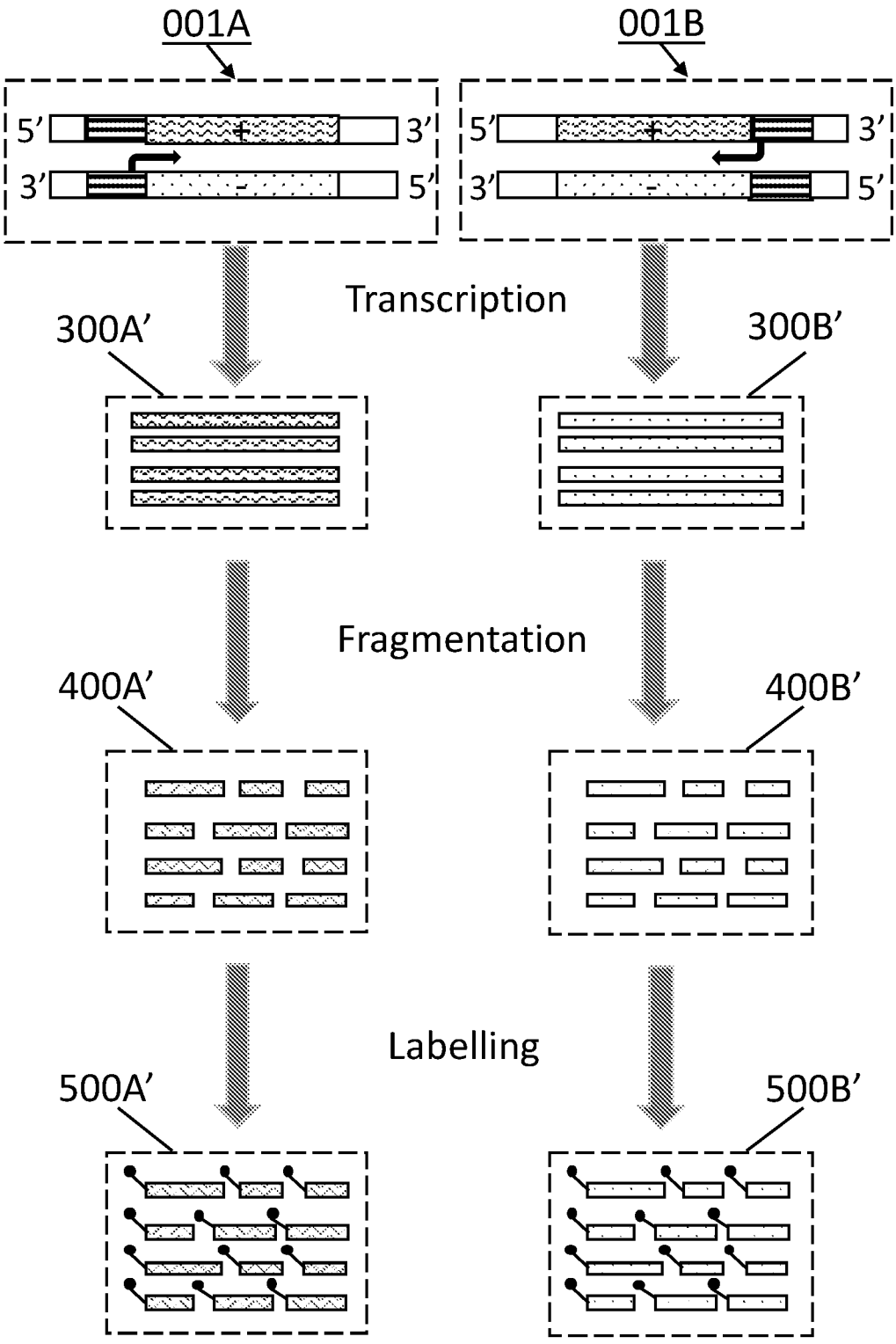
001A
001B
5′
3′
3′
5′
+
-
Transcription
300A′
300B′
Fragmentation
400A′
400B′
Labelling
500A′
500B′

FIG. 4B

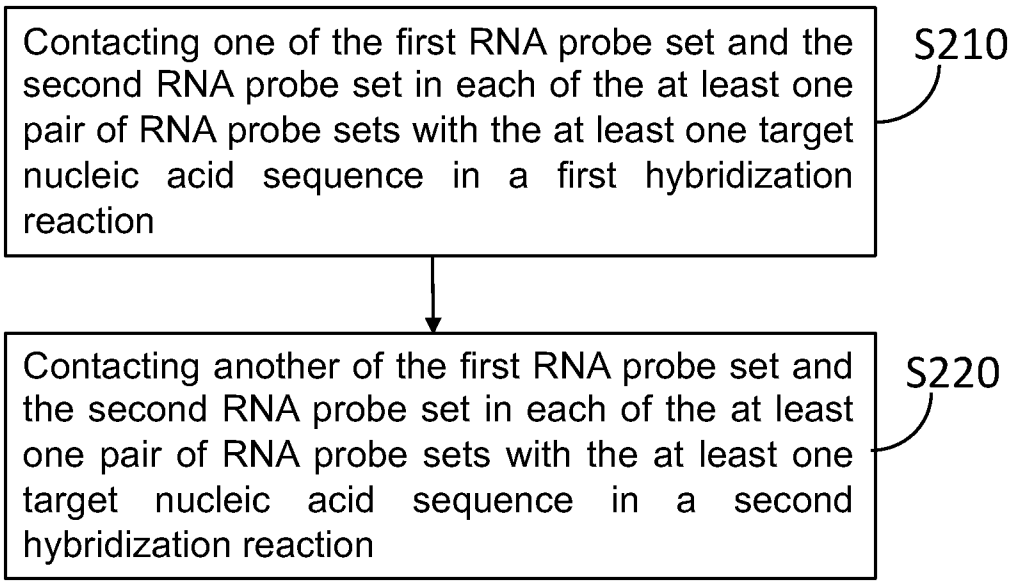
Contacting one of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a first hybridization reaction
S210
Contacting another of the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a second hybridization reaction
S220

FIG. 5A

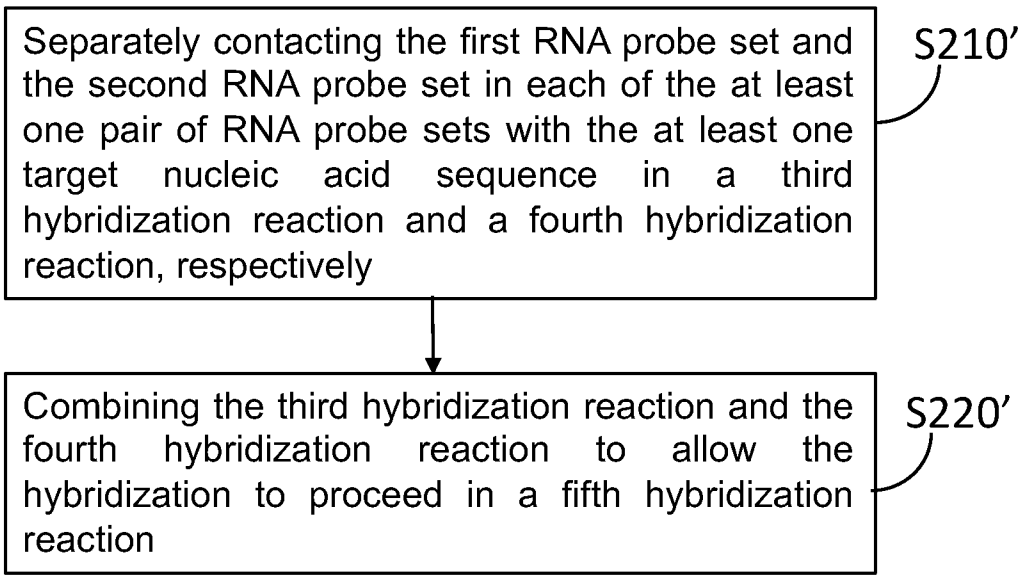
Separately contacting the first RNA probe set and the second RNA probe set in each of the at least one pair of RNA probe sets with the at least one target nucleic acid sequence in a third hybridization reaction and a fourth hybridization reaction, respectively
S210'
Combining the third hybridization reaction and the fourth hybridization reaction to allow the hybridization to proceed in a fifth hybridization reaction
S220'

FIG. 5B

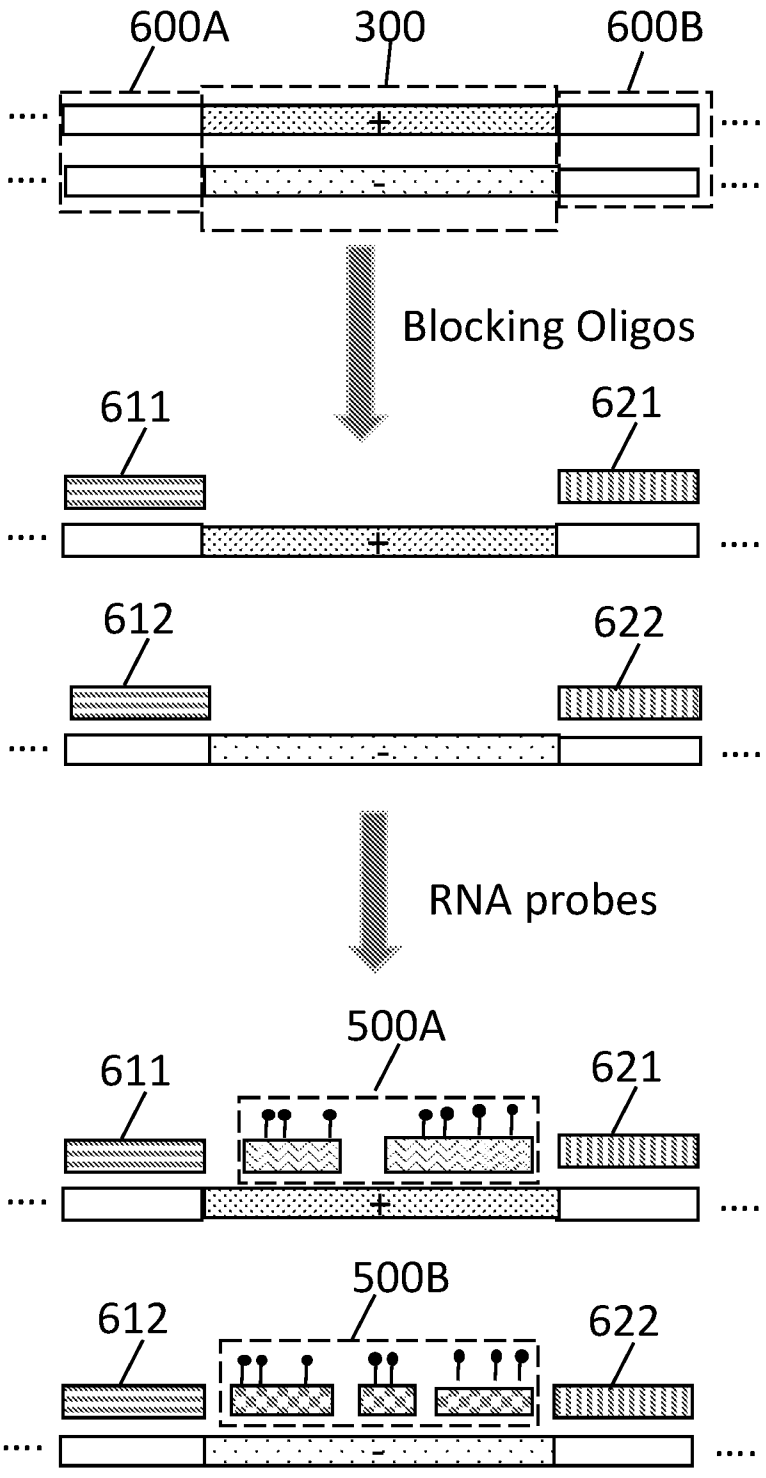
600A
300
600B
Blocking Oligos
611
621
612
622
RNA probes
500A
611
621
500B
612
622

FIG. 6A

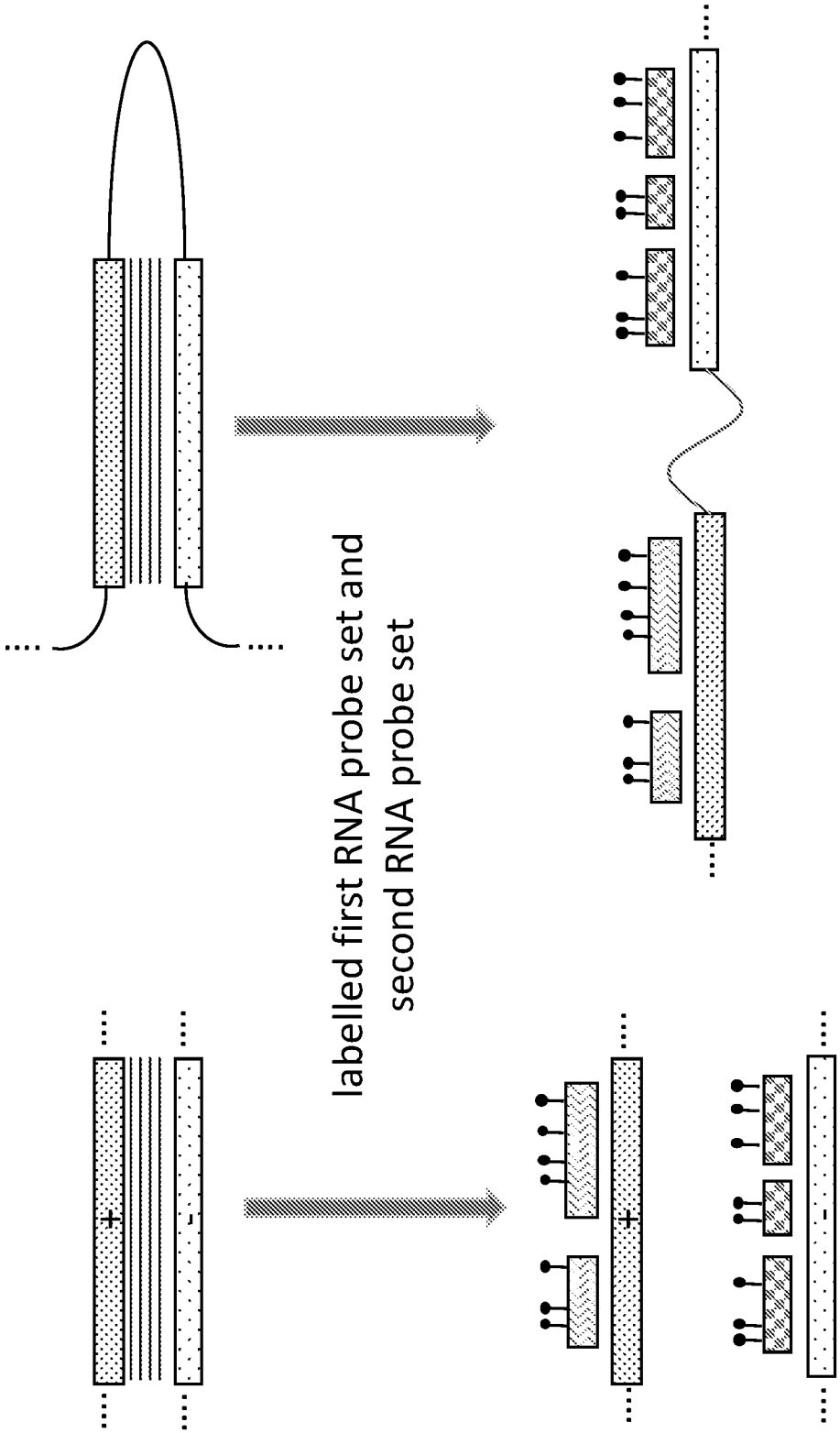

FIG. 6B

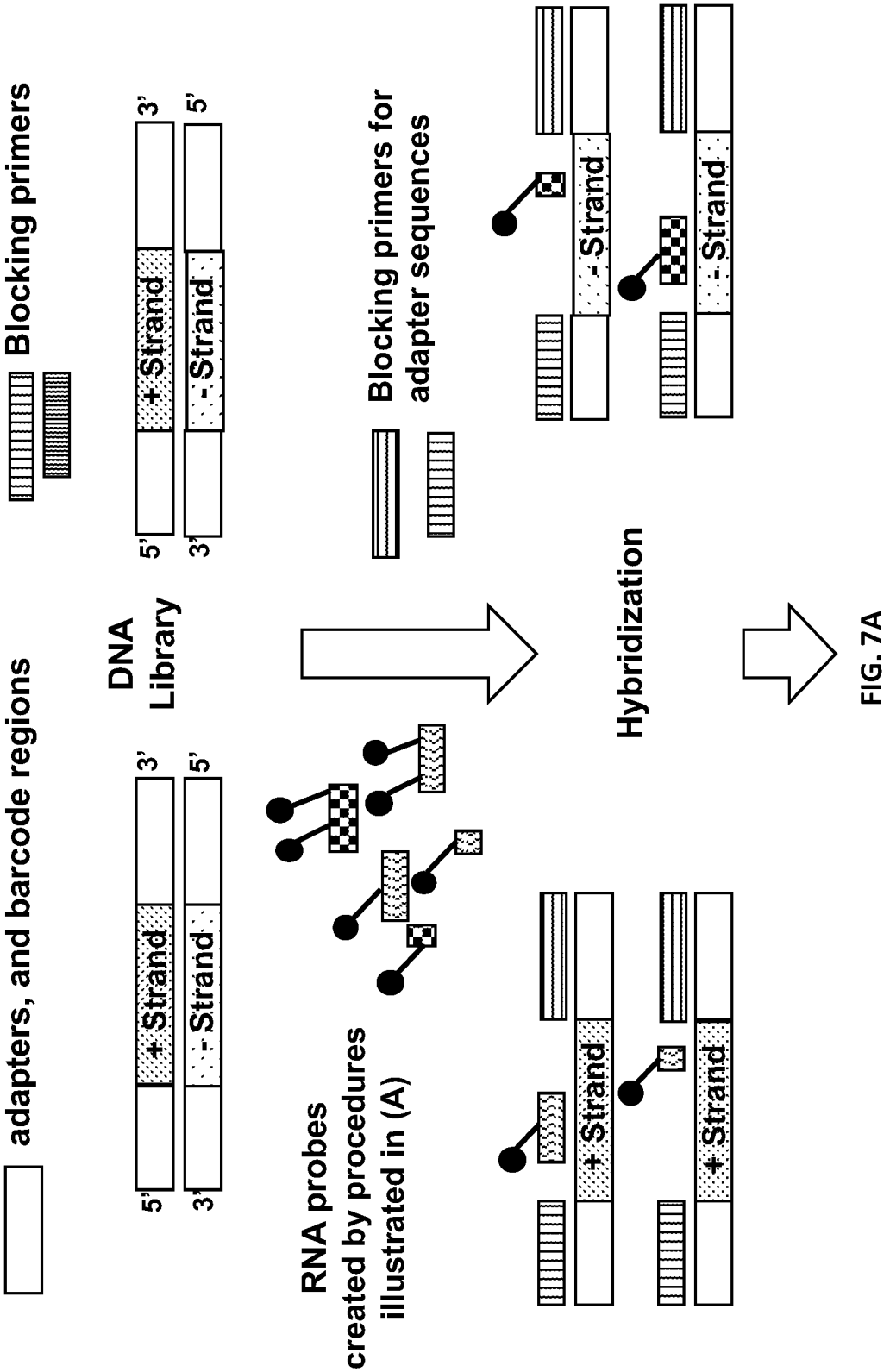
adapters, and barcode regions
Blocking primers
DNA Library
+ Strand
- Strand
5'
3'
RNA probes created by procedures illustrated in (A)
Blocking primers for adapter sequences
Hybridization

FIG. 7A

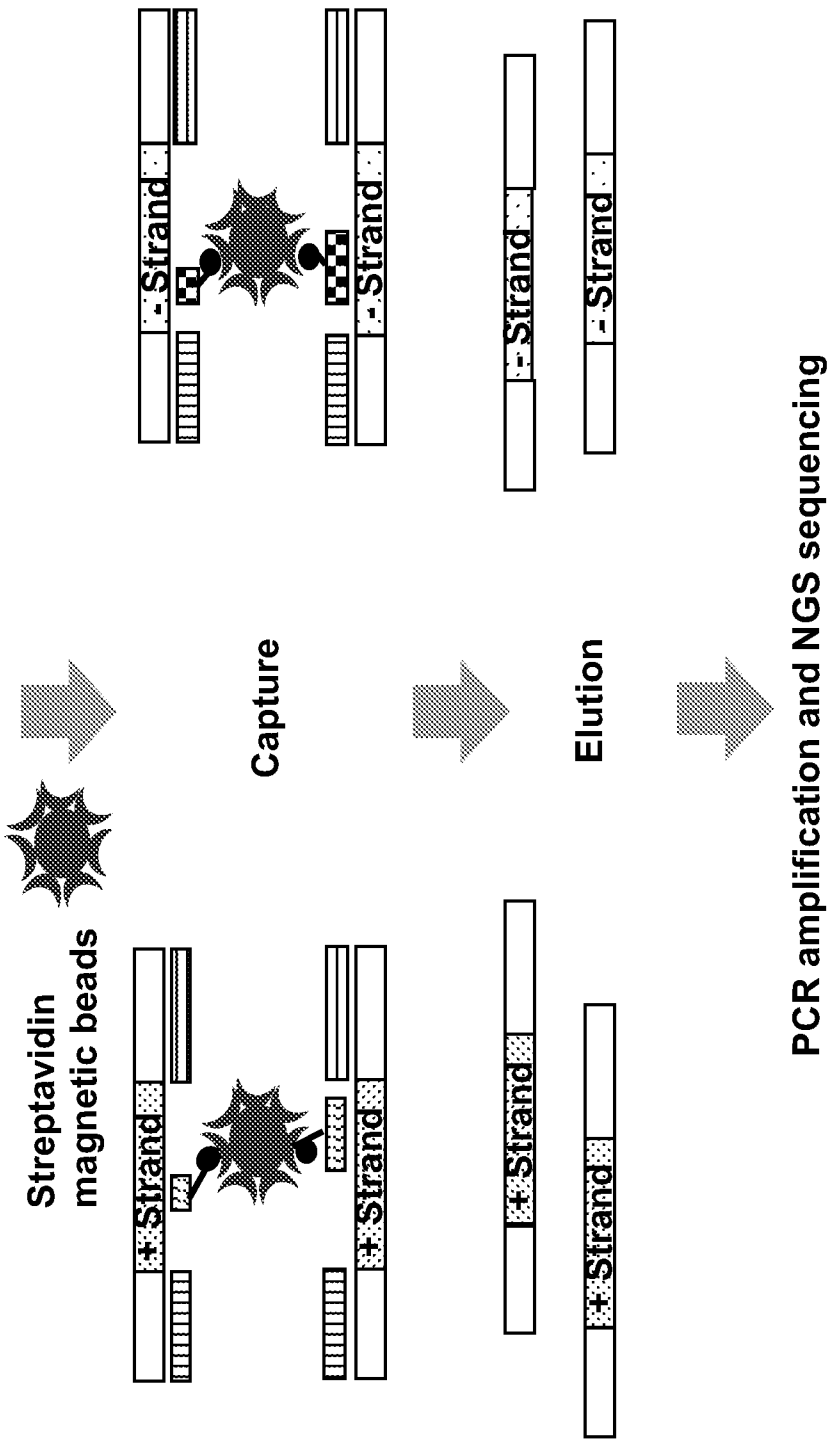
Streptavidin magnetic beads
+ Strand
+ Strand
Capture
- Strand
- Strand
Elution
+ Strand
+ Strand
PCR amplification and NGS sequencing
- Strand
- Strand

FIG. 7B

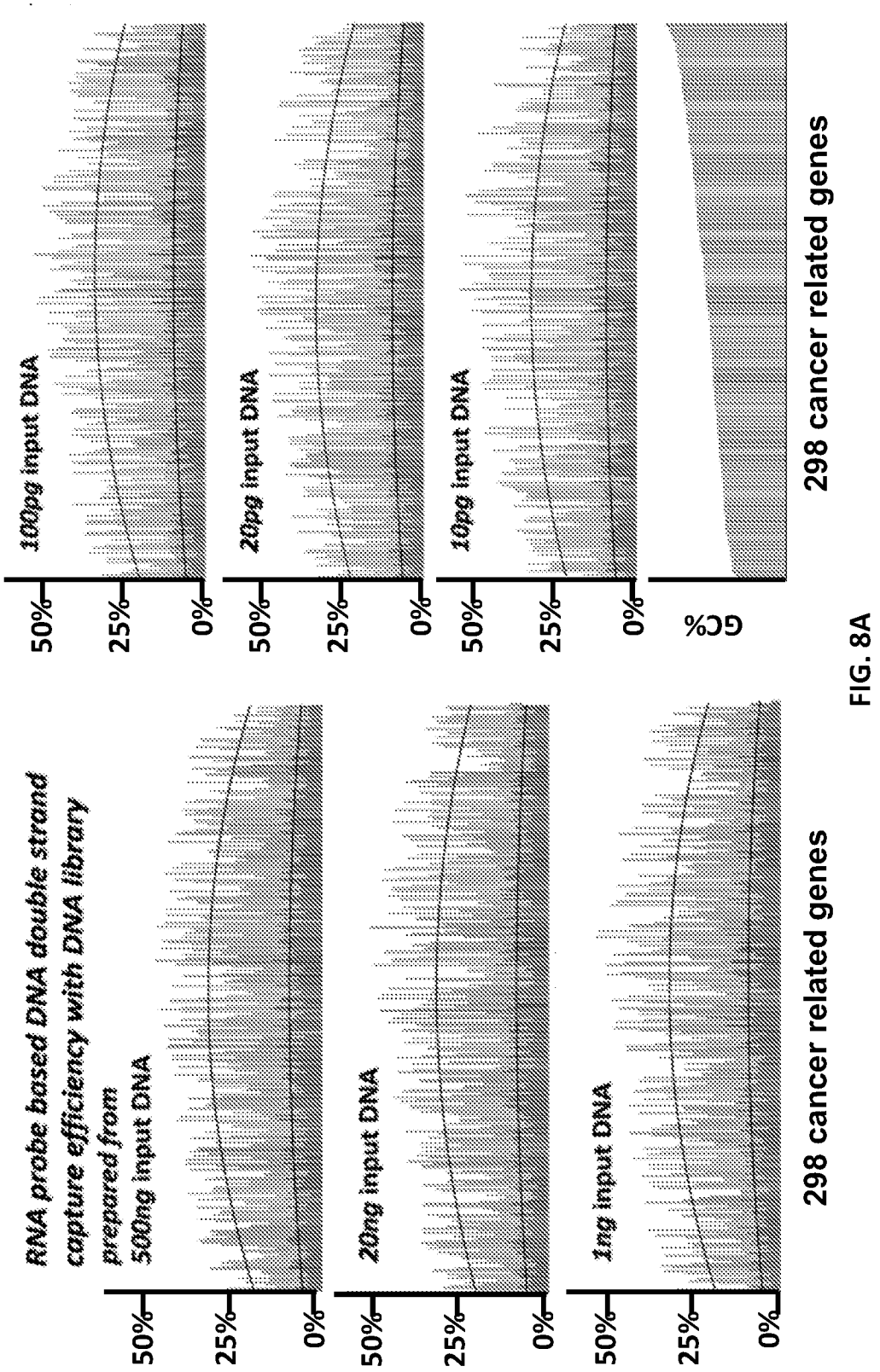
RNA probe based DNA double strand capture efficiency with DNA library prepared from
500ng input DNA
20ng input DNA
1ng input DNA
100pg input DNA
20pg input DNA
10pg input DNA
GC%
50%
25%
0%
298 cancer related genes

FIG. 8A

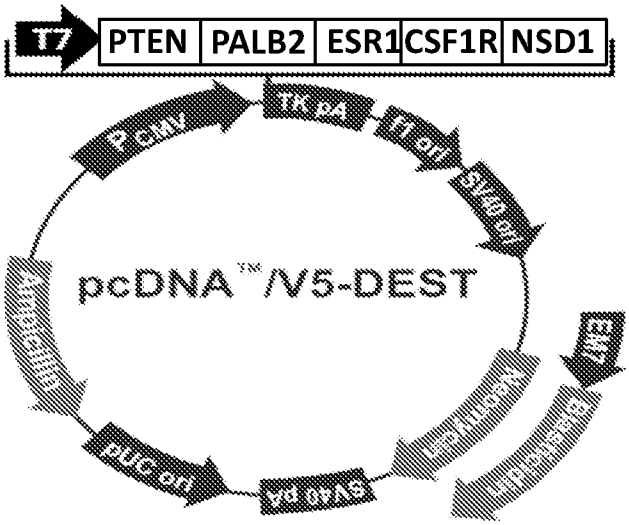
T7
PTEN
PALB2
ESR1
CSF1R
NSD1
pcDNA™/V5-DEST

FIG. 8B

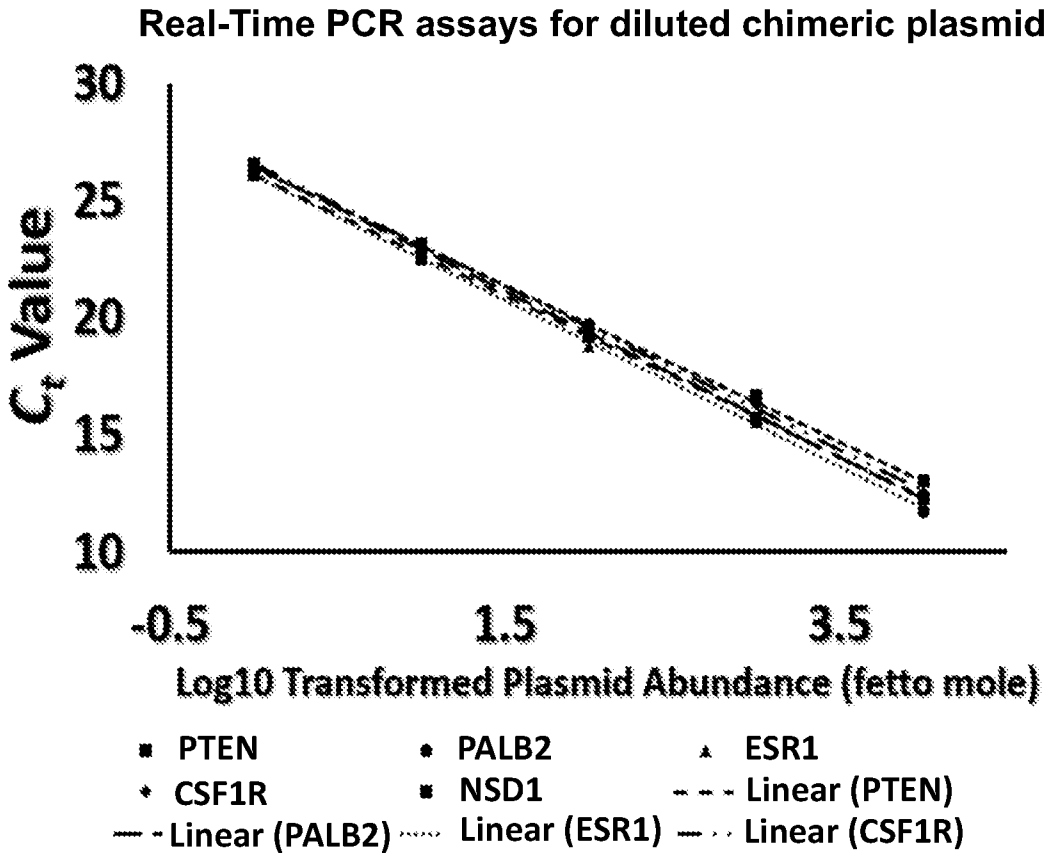
Real-Time PCR assays for diluted chimeric plasmid
30
25
20
15
10
C_t Value
-0.5
1.5
3.5
Log10 Transformed Plasmid Abundance (fetto mole)
PTEN
PALB2
ESR1
CSF1R
NSD1
Linear (PTEN)
Linear (PALB2)
Linear (ESR1)
Linear (CSF1R)

FIG. 8C

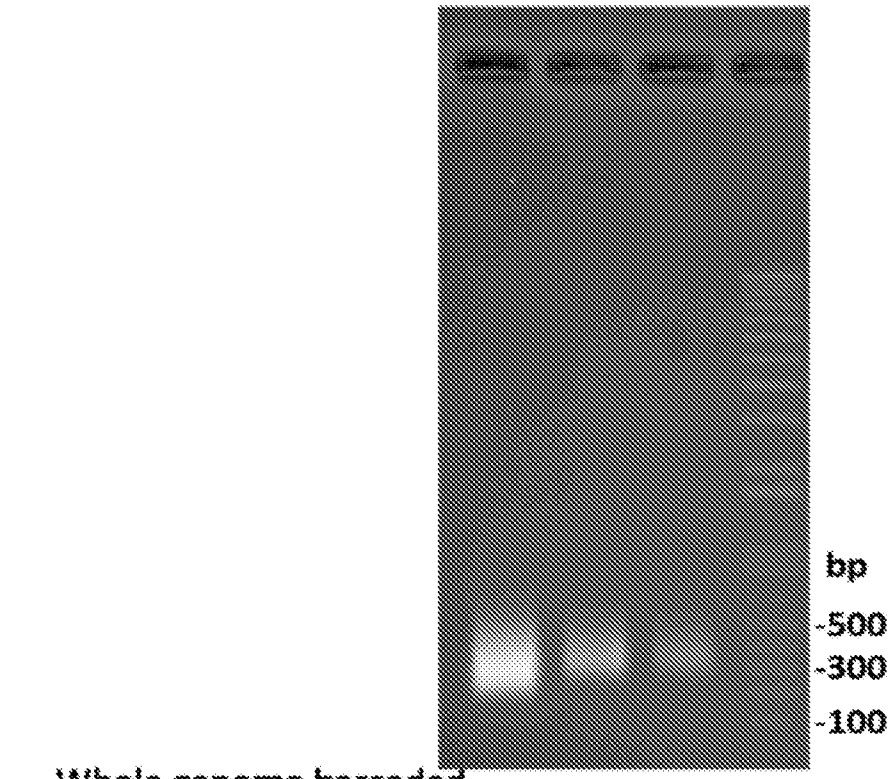
bp
~500
~300
~100
Whole genome barcoded single stranded library + + +
Double stranded RNA-based whole exome capture - + -
Single stranded RNA-based whole exome capture - - +

FIG. 8D

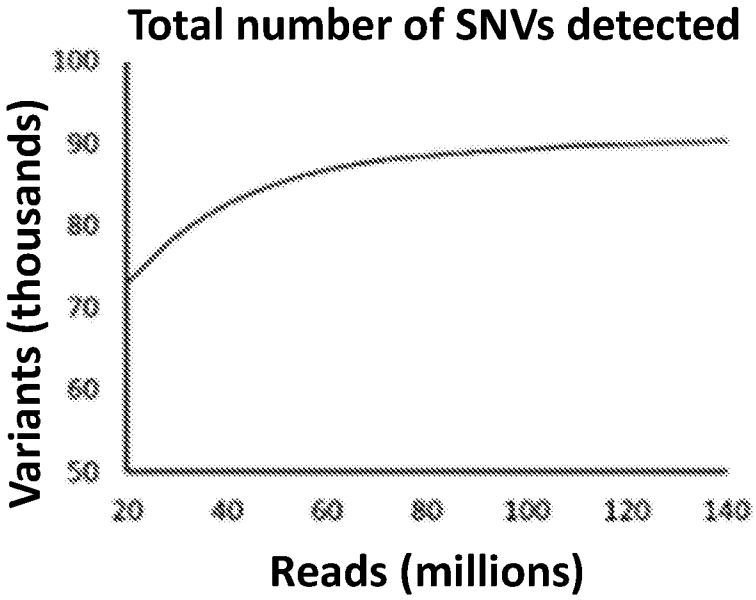
Total number of SNVs detected
Variants (thousands)
100
90
80
70
60
50
20
40
60
80
100
120
140
Reads (millions)

FIG. 9A

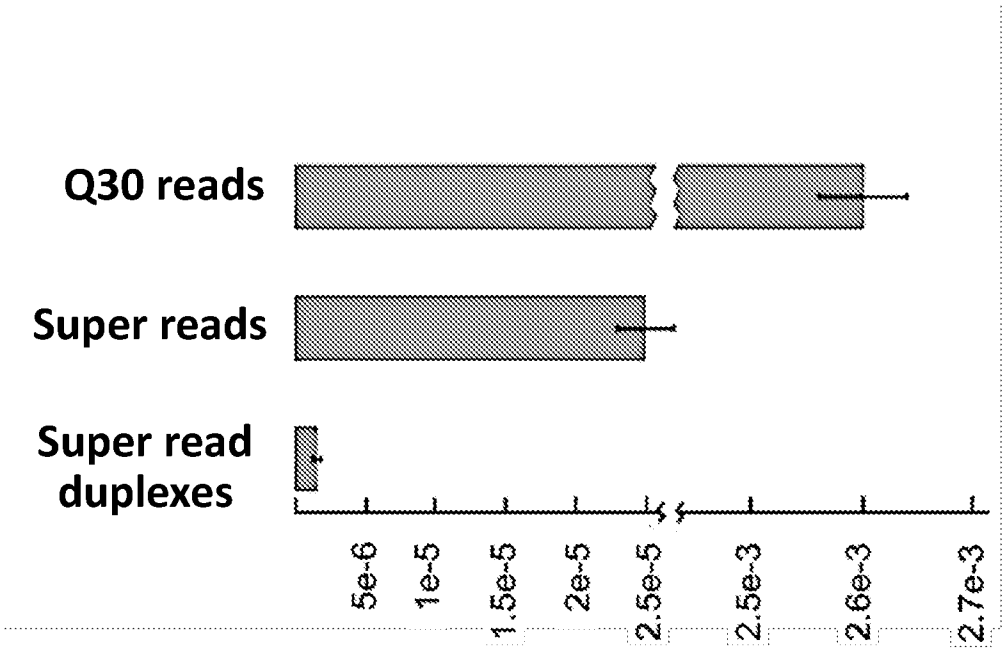
Q30 reads
Super reads
Super read duplexes
5e-6
1e-5
1.5e-5
2e-5
2.5e-5
2.5e-3
2.6e-3
2.7e-3

FIG. 9B

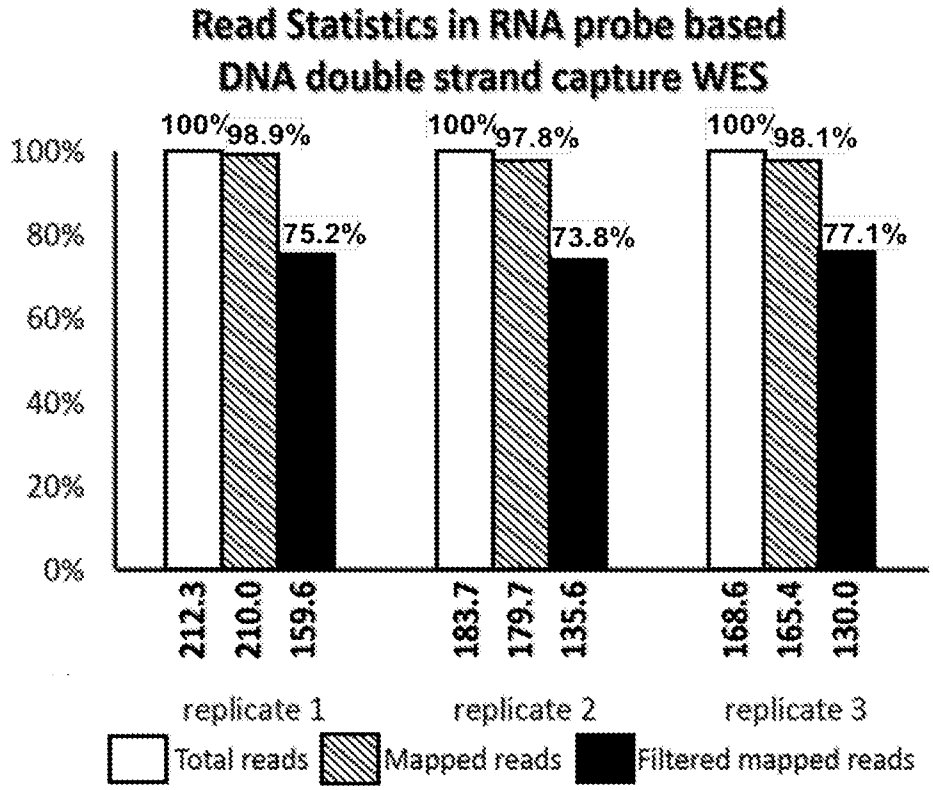

FIG. 10A

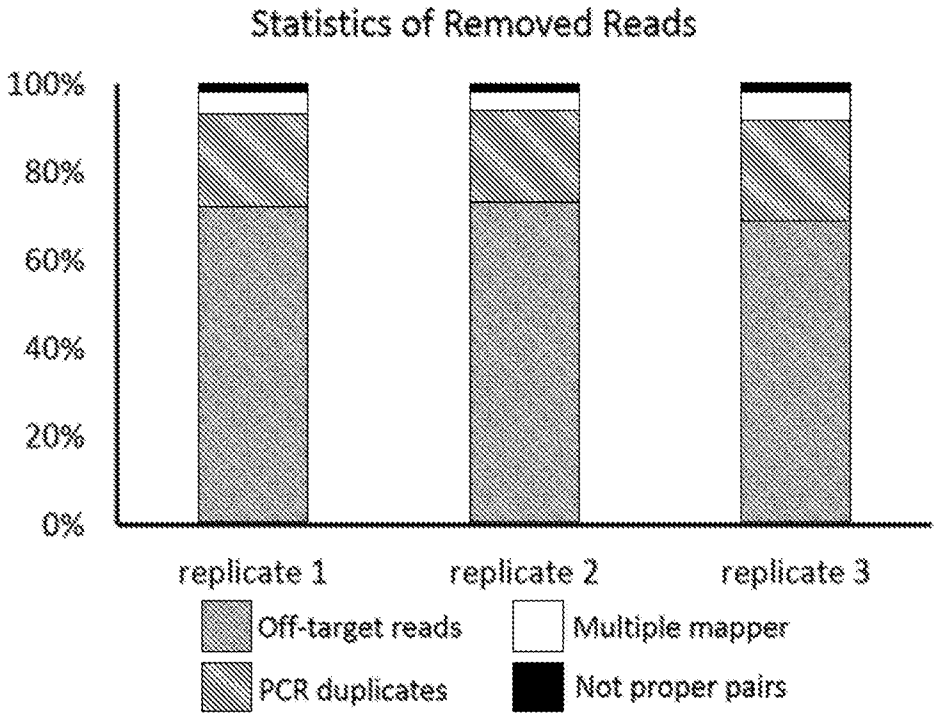

FIG. 10B

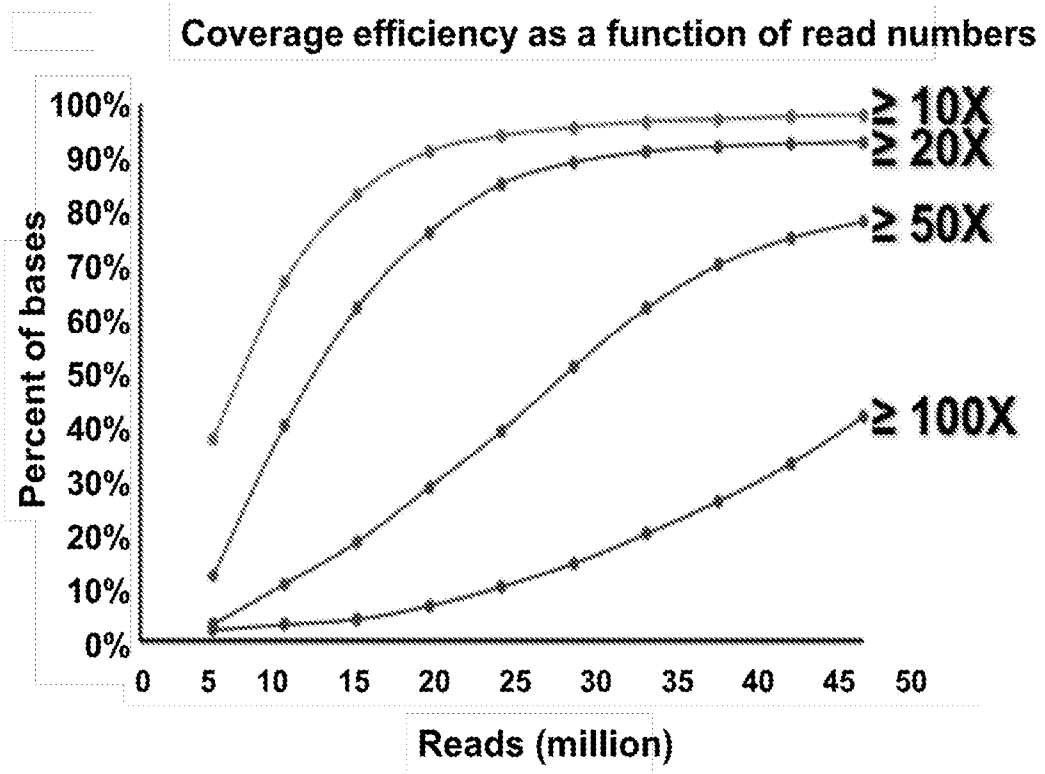

FIG. 10C

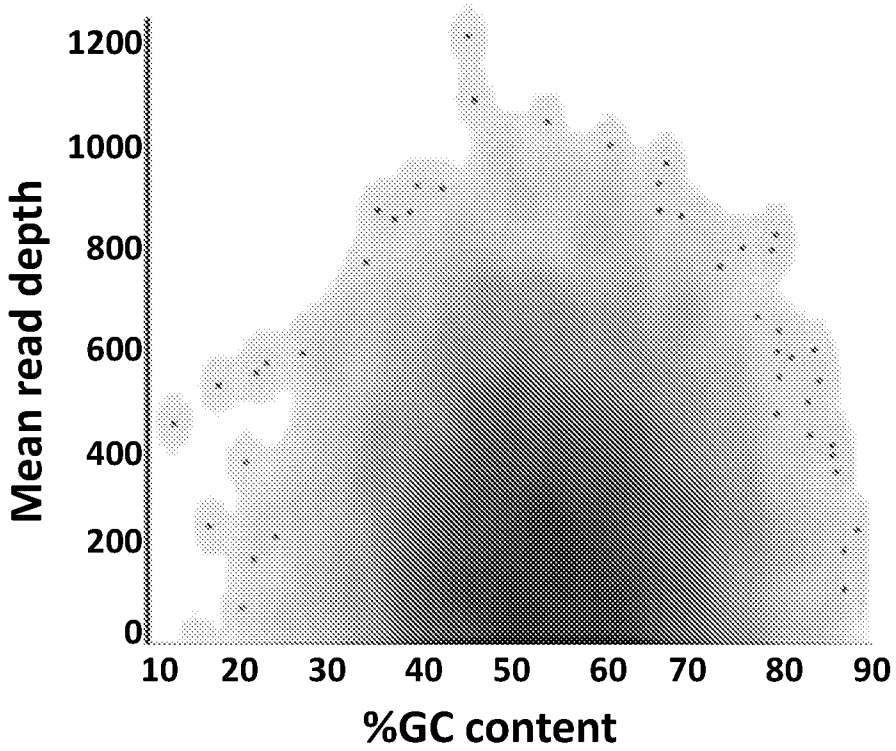

FIG. 11A

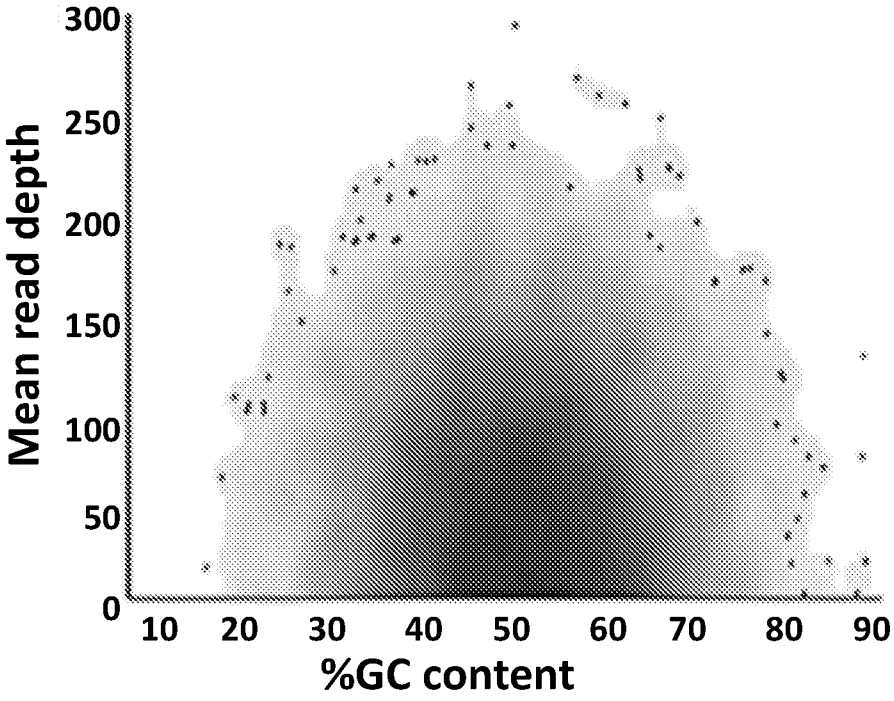

FIG. 11B

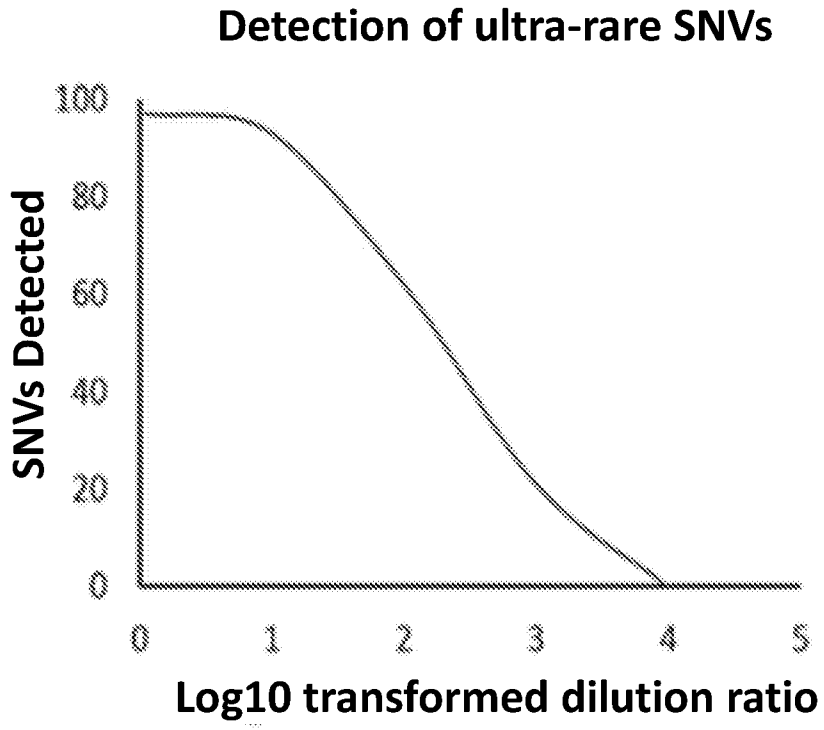

FIG. 12

Gene Symbol	EXON ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
PTEN	3	1	305	609	1.85	27.3%
FANCL	3	2	306	610	1.99	30.1%
FANCL	13	3	307	611	1.95	35.0%
FANCM	12	4	308	612	1.78	30.3%
NRAS	6	5	309	613	1.90	31.7%
XPA	3	6	310	614	1.94	32.4%
XPA	5	7	311	615	1.79	34.9%
BRCA2	9	8	312	616	1.85	32.8%
FANCF	1	9	313	617	1.92	32.8%
FIP1L1	6	10	314	618	1.97	32.9%
BRIP1	10	11	315	619	1.76	33.0%
MLH1	17	12	316	620	1.92	33.1%
STAG2	9	13	317	621	1.77	33.3%
XPO1	5	14	318	622	1.83	33.3%
RB1	14	15	319	623	1.78	33.6%
NBN	16	16	320	624	1.78	33.6%
NBN	10	17	321	625	1.94	41.7%
MDM2	4	18	322	626	1.78	33.8%
ATR	37	19	323	627	1.80	33.9%
FBXO11	2	20	324	628	1.91	34.0%
AKAP9	20	21	325	629	1.80	34.1%
SDHD	4	22	326	630	1.96	34.7%

FIG. 13A

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
PIK3CA	17	23	327	631	1.89	34.9%
CDC73	12	24	328	632	1.92	35.2%
ACSL3	10	25	329	633	1.96	35.4%
WRN	16	26	330	634	1.94	35.4%
HSP90AA1	6	27	331	635	1.77	36.0%
CDK12	7	28	332	636	1.76	36.0%
POT1	6	29	333	637	1.84	36.1%
KDM6A	7	30	334	638	1.98	36.4%
KRAS	2	31	335	639	1.90	36.5%
BLM	3	32	336	640	1.87	36.6%
GOLGA5	9	33	337	641	1.91	36.6%
PRKAR1A	11	34	338	642	1.91	36.7%
TMPRSS2	3	35	339	643	1.89	36.8%
CLTC	24	36	340	644	1.95	36.9%
FAS	8	37	341	645	1.99	36.9%
MSH2	9	38	342	646	1.91	37.0%
BUB1B	3	39	343	647	1.75	37.0%
NF1	24	40	344	648	1.79	37.3%
TOP1	7	41	345	649	1.96	37.3%
NCOR1	17	42	346	650	1.87	37.3%
POLH	8	43	347	651	1.75	37.3%
SMAD2	8	44	348	652	1.87	37.5%

FIG. 13B

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
FUBP1	3	45	349	653	1.83	37.5%
RAF1	13	46	350	654	1.80	37.6%
SETD2	7	47	351	655	1.98	37.7%
DICER1	17	48	352	656	1.77	37.7%
FANCC	7	49	353	657	1.89	37.7%
GPC3	4	50	354	658	1.98	37.7%
DDX5	11	51	355	659	1.79	38.0%
FAM46C	2	52	356	660	1.96	38.1%
KAT6B	9	53	357	661	2.00	38.1%
RAD51C	7	54	358	662	1.91	38.4%
PPP6C	6	55	359	663	1.91	38.4%
IDH1	4	56	360	664	1.82	38.4%
PMS2	9	57	361	665	1.96	38.5%
ATRX	4	58	362	666	1.86	38.7%
ASXL1	2	59	363	667	1.85	38.7%
ASXL1	10	60	364	668	1.84	54.7%
PBRM1	8	61	365	669	2.00	38.8%
MET	16	62	366	670	1.96	38.9%
EWSR1	12	63	367	671	1.89	39.0%
ATM	38	64	368	672	1.84	39.3%
PALB2	6	65	369	673	1.79	39.3%
SF3B1	5	66	370	674	1.93	39.3%

FIG. 13C

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
FANCI	16	67	371	675	1.98	39.7%
TET2	5	68	372	676	1.92	39.8%
TSHR	6	69	373	677	1.97	40.0%
WHSC1L1	19	70	374	678	1.76	40.2%
NFE2L2	4	71	375	679	2.00	40.3%
BCOR	4	72	376	680	1.91	40.4%
ETV5	10	73	377	681	1.77	40.5%
CREBBP	23	74	378	682	1.82	40.6%
SDHAF2	3	75	379	683	1.82	40.7%
NPM1	11	76	380	684	1.86	40.7%
SPOP	10	77	381	685	1.93	40.7%
TCF12	20	78	382	686	1.97	40.7%
ARID1B	3	79	383	687	1.96	40.8%
BRAF	14	80	384	688	1.89	41.0%
FH	2	81	385	689	1.85	41.0%
ERCC5	3	82	386	690	1.78	41.2%
FANCD2	8	83	387	691	1.78	41.3%
MTOR	52	84	388	692	1.81	41.3%
CASP8	7	85	389	693	1.76	41.5%
NUP98	3	86	390	694	1.90	41.7%
MDM4	10	87	391	695	1.99	41.7%
ACVR1	7	88	392	696	1.83	41.8%

FIG. 13D

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
ARID2	18	89	393	697	1.95	41.8%
MECOM	6	90	394	698	1.76	42.0%
RET	16	91	395	699	2.00	42.0%
BCL2	1	92	396	700	1.99	42.0%
FUS	9	93	397	701	1.94	42.1%
TCF7L2	4	94	398	702	1.94	42.4%
ATIC	5	95	399	703	1.80	42.6%
PMS1	2	96	400	704	1.79	42.6%
CHD4	34	97	401	705	1.77	42.7%
PTPN11	8	98	402	706	1.88	42.8%
ABL2	3	99	403	707	1.78	42.9%
MAP2K4	9	100	404	708	1.91	42.9%
CRTC3	5	101	405	709	1.96	43.0%
BRCA1	3	102	406	710	1.82	43.2%
LCP1	2	103	407	711	1.87	43.3%
FCRL4	10	104	408	712	1.83	43.3%
RAC1	4	105	409	713	1.77	43.5%
SMARCA4	14	106	410	714	1.76	43.5%
HSP90AB1	6	107	411	715	2.00	43.5%
TBL1XR1	3	108	412	716	1.88	43.6%
VHL	2	109	413	717	1.93	43.7%
NOTCH2	12	110	414	718	1.94	43.7%

FIG. 13E

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
CTNNB1	15	111	415	719	1.86	43.7%
POLE	3	112	416	720	1.95	43.8%
SDHC	3	113	417	721	1.97	43.9%
KIT	20	114	418	722	1.76	44.1%
ETV6	7	115	419	723	1.85	44.1%
CTCF	6	116	420	724	1.98	44.1%
MAP2K1	9	117	421	725	1.97	44.2%
PSIP1	16	118	422	726	1.80	44.3%
RAD21	13	119	423	727	1.85	44.6%
KMT2A	31	120	424	728	1.88	44.7%
CBL	14	121	425	729	1.90	44.7%
MYB	4	122	426	730	1.87	44.7%
SYK	8	123	427	731	1.91	44.7%
ETV1	14	124	428	732	1.92	44.7%
BMPR1A	12	125	429	733	1.96	44.7%
TBX3	4	126	430	734	1.96	44.9%
CBFB	3	127	431	735	1.90	44.9%
DDR2	11	128	432	736	1.79	45.0%
NF2	5	129	433	737	2.00	45.0%
MSH6	7	130	434	738	1.88	45.1%
TSC1	2	131	435	739	1.86	45.3%
AFF4	12	132	436	740	1.91	45.3%

FIG. 13F

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
FOXP1	24	133	437	741	1.75	45.3%
TNFAIP3	4	134	438	742	1.96	45.6%
EP300	19	135	439	743	1.96	45.8%
MAP3K1	18	136	440	744	1.80	46.0%
U2AF1	4	137	441	745	1.98	46.0%
ERCC3	4	138	442	746	1.82	46.0%
CYLD	2	139	443	747	1.87	46.1%
IKZF1	4	140	444	748	1.96	46.1%
STK11	2	141	445	749	1.83	46.2%
PHOX2B	1	142	446	750	1.97	46.3%
SBDS	2	143	447	751	1.88	46.3%
SFPQ	7	144	448	752	1.86	46.6%
WHSC1	10	145	449	753	1.85	46.6%
PRRX1	4	146	450	754	1.80	46.7%
GNAQ	4	147	451	755	1.76	46.8%
ERCC4	10	148	452	756	1.91	47.2%
FLT3	16	149	453	757	1.75	47.2%
MAX	3	150	454	758	1.92	47.5%
EGFR	5	151	455	759	1.80	47.7%
RPL22	3	152	456	760	1.80	47.7%
SMAD4	7	153	457	761	1.94	48.0%
KDR	20	154	458	762	1.82	48.0%

FIG. 13G

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
CDKN1B	2	155	459	763	1.76	48.1%
MKL1	6	156	460	764	1.99	48.1%
BCL11A	3	157	461	765	1.77	48.4%
JAK2	2	158	462	766	1.86	48.6%
ERBB4	5	159	463	767	1.87	48.7%
ERBB3	24	160	464	768	1.85	48.7%
PDGFRA	2	161	465	769	1.77	48.7%
ABL1	6	162	466	770	1.96	48.8%
TGFBR2	6	163	467	771	1.92	49.0%
EZH2	13	164	468	772	1.88	49.0%
ATF1	3	165	469	773	1.99	49.1%
XPC	8	166	470	774	1.96	49.3%
DNAJB1	2	167	471	775	1.80	49.4%
USP6	2	168	472	776	1.79	49.4%
FGFR2	5	169	473	777	1.98	49.4%
RHOA	4	170	474	778	1.96	49.5%
EXT2	9	171	475	779	1.89	49.7%
CDK4	5	172	476	780	1.92	49.7%
BTK	12	173	477	781	1.77	50.0%
PIK3R1	3	174	478	782	1.83	50.0%
CUX1	8	175	479	783	1.77	50.0%
BCR	6	176	480	784	1.89	50.0%

FIG. 13H

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
DNMT3A	10	177	481	785	1.94	50.0%
COL1A1	6	178	482	786	1.94	50.3%
ESR1	10	179	483	787	1.80	50.5%
ALK	26	180	484	788	1.76	51.0%
ARHGAP26	3	181	485	789	1.95	51.4%
ERCC1	6	182	486	790	1.94	51.6%
PTCH1	25	183	487	791	1.80	51.7%
AR	7	184	488	792	1.90	51.7%
NCOA3	16	185	489	793	1.84	51.7%
IGF2R	20	186	490	794	1.87	52.0%
IGF1R	15	187	491	795	1.92	52.2%
TERT	10	188	492	796	1.76	52.2%
ROS1	4	189	493	797	1.95	52.3%
RUNX1	6	190	494	798	1.87	52.5%
XRCC1	16	191	495	799	1.79	52.6%
MLLT4	11	192	496	800	1.85	53.2%
ARID1A	17	193	497	801	1.96	53.2%
SDHB	6	194	498	802	1.87	53.2%
KLF6	3	195	499	803	1.82	53.3%
SMAD3	7	196	500	804	1.96	53.3%
PAX5	7	197	501	805	1.85	53.3%
SMARCB1	3	198	502	806	1.99	53.4%

FIG. 13I

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
PRSS1	1	199	503	807	1.80	53.8%
ELF4	7	200	504	808	1.86	54.0%
FANCA	13	201	505	809	1.90	54.0%
CALR	4	202	506	810	1.94	54.1%
EXT1	8	203	507	811	1.98	54.1%
TP53	6	204	508	812	1.93	54.3%
PDGFRB	21	205	509	813	1.83	54.3%
FAT1	20	206	510	814	1.79	54.6%
KDM5C	2	207	511	815	1.83	54.6%
KDM5C	24	208	512	816	1.80	55.6%
WT1	7	209	513	817	1.76	54.6%
IDH2	3	210	514	818	1.88	54.7%
WAS	8	211	515	819	1.95	54.7%
MUTYH	18	212	516	820	1.79	54.8%
MYD88	4	213	517	821	1.80	54.9%
PLCG1	20	214	518	822	1.87	54.9%
AURKA	3	215	519	823	1.79	54.9%
MAML1	4	216	520	824	1.86	55.1%
SH2B3	6	217	521	825	1.91	55.2%
FGFR1	5	218	522	826	1.92	55.5%
PPP2R1A	2	219	523	827	1.93	55.6%
MED12	44	220	524	828	1.84	56.0%

FIG. 13J

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
NDRG1	2	221	525	829	1.84	56.0%
FANCE	8	222	526	830	1.91	56.0%
DDB2	5	223	527	831	1.75	56.1%
ERCC2	16	224	528	832	1.98	56.1%
SRC	6	225	529	833	1.81	56.2%
AXIN2	9	226	530	834	1.80	56.2%
NOTCH4	26	227	531	835	1.93	56.4%
GATA3	5	228	532	836	1.98	56.4%
EIF4A2	1	229	533	837	1.97	56.5%
PRKACA	2	230	534	838	1.83	56.7%
JAK1	3	231	535	839	1.89	57.1%
MYH9	5	232	536	840	1.85	57.1%
CHEK2	1	233	537	841	2.00	57.3%
FANCG	9	234	538	842	2.00	57.3%
SUFU	5	235	539	843	1.89	57.6%
CEBPA	1	236	540	844	1.97	57.8%
SRGAP3	19	237	541	845	1.81	58.0%
NCOR2	9	238	542	846	1.99	58.1%
FLCN	3	239	543	847	1.98	58.2%
CIITA	10	240	544	848	1.85	58.7%
RBM10	5	241	545	849	1.99	58.7%
ETV4	2	242	546	850	1.98	58.7%

FIG. 13K

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
CCND1	4	243	547	851	1.87	58.8%
ERBB2	10	244	548	852	1.82	58.8%
MEN1	9	245	549	853	1.98	58.9%
CCNE1	4	246	550	854	1.84	59.1%
CDK6	2	247	551	855	1.77	59.3%
MYH11	1	248	552	856	1.97	59.4%
HLA-A	5	249	553	857	1.99	59.4%
RNF43	5	250	554	858	1.87	59.7%
HLF	2	251	555	859	1.76	60.0%
AXIN1	4	252	556	860	1.92	60.0%
FLT4	22	253	557	861	1.79	60.0%
SMO	7	254	558	862	1.98	60.2%
ACSL6	10	255	559	863	1.99	60.7%
AKT1	13	256	560	864	1.94	60.7%
H3F3B	3	257	561	865	1.80	60.8%
RARA	1	258	562	866	1.83	60.9%
FBXW7	2	259	563	867	1.79	61.2%
TACC3	5	260	564	868	1.87	61.3%
FGFR3	4	261	565	869	1.98	61.7%
TSC2	2	262	566	870	1.96	62.0%
MYCN	1	263	567	871	1.84	62.0%
CSF1R	17	264	568	872	1.92	62.0%

FIG. 13L

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
APC	2	265	569	873	1.88	62.2%
NTRK1	5	266	570	874	1.91	62.3%
BAP1	8	267	571	875	1.94	62.4%
AKT2	7	268	572	876	1.78	62.7%
POLD1	14	269	573	877	1.84	62.7%
RECQL4	10	270	574	878	1.86	62.9%
GNAS	9	271	575	879	1.85	62.9%
DAXX	1	272	576	880	1.79	62.9%
GATA2	6	273	577	881	1.77	63.1%
PHF6	1	274	578	882	1.82	63.2%
MPL	11	275	579	883	1.96	63.3%
MYC	2	276	580	884	1.93	63.3%
CNOT3	14	277	581	885	1.81	63.6%
CIC	14	278	582	886	1.93	63.9%
FGFR4	14	279	583	887	1.90	64.2%
NOTCH1	2	280	584	888	1.86	64.9%
FOXL2	1	281	585	889	1.78	65.3%
GNA11	4	282	586	890	1.98	65.3%
GATA1	1	283	587	891	1.95	65.4%
KLF4	2	284	588	892	1.95	65.5%
FANCB	1	285	589	893	1.86	65.7%
THRAP3	1	286	590	894	1.80	65.7%

FIG. 13M

Gene Symbol	Exon ID	Forward Primer (SEQ ID NO.)	Reverse Primer (SEQ ID NO.)	Amplicon Seq (SEQ ID NO.)	Amplif Eff Constant	GC% of Amplicon
RPL5	1	287	591	895	1.90	65.8%
JAK3	1	288	592	896	1.87	66.7%
NOTCH3	2	289	593	897	1.91	66.7%
CDKN2A	1	290	594	898	1.90	66.8%
H3F3A	1	291	595	899	1.84	66.9%
HNF1A	8	292	596	900	1.96	67.4%
NKX2-1	1	293	597	901	1.88	67.5%
FOXA1	2	294	598	902	1.82	68.4%
CDH1	2	295	599	903	1.93	68.7%
CDKN2B	1	296	600	904	1.77	68.7%
ZFHX3	1	297	601	905	1.79	70.0%
HRAS	5	298	602	906	1.79	70.0%
MYCL1	2	299	603	907	1.91	70.2%
PER1	1	300	604	908	1.80	71.1%
PER1	3	301	605	909	1.83	60.1%
SUZ12	1	302	606	910	1.79	71.6%
CDKN2C	1	303	607	911	1.77	72.8%
NSD1	2	304	608	912	1.93	74.1%

FIG. 13N

	Normal Tissue	Tumor Tissue
Initial mapped reads	30.7 billion	25.7 billion
Average raw coverage	240.5 X	173.9 X
Unique read family (URF)	625 million	577 million
Super read duplexes	272 million	231 million
Initial mapped reads per super read family	49.1	44.5
Initial mapped reads per super read duplex	112.9	111.3
Super reads per super read duplex	2.3	2.5

FIG. 14

Gene Symbol	Transcript Accession	Nucleotide (genomic)	Amino Acid (protein)	Mutant Allele Fraction	Mutant Allele Fraction in 1 to 1000 spiked normal sample	Mutation Type	Consequence	Confirmed	Detected again in 1:1,000 dilution
FOCAD	CCDS34993	chr9:20865973 A->G	p.K2104E	29.41%	0.029%	Substitution	Non-synonymous	YES	YES
KRAS	CCDS8703.1	chr12:25289551C->A	p.G12V	27.69%	0.028%	Substitution	Non-synonymous	YES	YES
TRIM59	CCDS3190	chr3:160156368 ->T	NA	23.08%	0.023%	Insertion	Frameshift	YES	NO
GOLGA6L2	ENST00000312015	chr15:23685010 C->T	p.E1633K	23.08%	0.023%	Substitution	Non-synonymous	YES	YES
CAMKK2	CCDS53837	chr12:121678328 ->T	NA	22.58%	0.023%	Insertion	Frameshift	YES	NO
TTC39A	ENST00000262676	chr1:51767913 C->	NA	21.74%	0.022%	Deletion	Frameshift	YES	NO
VEGFC	CCDS43285	chr4:177605084 A->C	p.M1256R	21.43%	0.021%	Substitution	Non-synonymous	YES	NO
NBPF15	CCDS932	chr1:148594474 C->T	p.S1847L	19.54%	0.020%	Substitution	Non-synonymous	YES	YES
ALG9	ENST00000428306	chr11:111742145 C->G	NA	16.96%	0.017%	NA	Splicesite donor	YES	YES
OAS1	ENST00000377508	chr12:113369718 C->A	NA	16.95%	0.017%	NA	Splicesite acceptor	YES	NO
NFE2L2	CCDS42782	chr2:178097188 C->T	p.D526N	16.46%	0.016%	Substitution	Non-synonymous	YES	YES
OAS1	ENST00000377508	chr12:113369716 A->C	NA	15.87%	0.016%	NA	Splicesite acceptor	YES	NO
E2F8	ENST00000396159	chr11:119263371 T->	NA	15.63%	0.016%	NA	Splicesite donor	YES	NO
FUZ	ENST00000421740	chr19:50314458 G->A	p.S650F	15.15%	0.015%	Substitution	Non-synonymous	YES	NO
FUZ	ENST00000421740	chr19:50314459 A->G	p.S649P	15.07%	0.015%	Substitution	Non-synonymous	YES	NO
ADAM21	CCDS9804	chr14:70924632 G->A	p.S416N	14.67%	0.015%	Substitution	Non-synonymous	YES	YES
COL27A1	CCDS6802	chr9:116918267 GCG->	NA	14.04%	0.014%	Deletion	Frameshift	YES	NO

FIG. 15A

USP19	CCDS43090	chr3:49155556 A->	NA	14.04%	0.014%	NA	Splicesite acceptor	YES	NO
FAM182A	ENST00000246000	chr20:26061967 G->T	p.A319S	13.72%	0.014%	Substitution	Non-synonymous	YES	YES
VLDLR	ENST00000397921	chr9:2622173 C->G	p.T155S	12.28%	0.012%	Substitution	Non-synonymous	YES	NO
MMEL1	CCDS30569	chr1:2560819 CAG->	NA	11.94%	0.012%	Deletion	Frameshift	YES	NO
LSM12	ENST00000411445	chr17:42141311 ->A	NA	11.32%	0.011%	Insertion	Frameshift	YES	YES
MFSD4B	CCDS5090	chr6:111587361 T->	NA	10.94%	0.011%	Deletion	Frameshift	YES	YES
NADK	CCDS30565	chr1:1684348 ->CCT	NA	10.85%	0.011%	Insertion	Frameshift	YES	YES
PTPRG	CCDS2895	chr3:62267290 A->G	p.E3818G	10.71%	0.011%	Substitution	Non-synonymous	YES	NO
CTNNB1	CCDS2694.1	chr3:41241445C->A	NA	10.53%	0.011%	NA	Splicesite acceptor	YES	YES
ZMYND8	ENST00000372023	chr20:45867410 A->	NA	10.09%	0.010%	Deletion	Frameshift	YES	NO
TP53	CCDS11118.1	chr17:7520075A->C	p.F113V	9.39%	0.009%	Substitution	Non-synonymous	YES	YES
DCD	ENST00000419084	chr12:55039821 G->A	p.P211S	8.46%	0.008%	Substitution	Non-synonymous	YES	NO
KCNB2	CCDS6209	chr8:73480397 ->A	NA	8.34%	0.008%	Insertion	Frameshift	YES	YES
ALG9	ENST00000428306	chr11:111742141 G->C	p.P761R	7.98%	0.008%	Substitution	Non-synonymous	YES	NO
PRRT2	CCDS10654	chr16:29825016 C->	NA	7.80%	0.008%	Deletion	Frameshift	NO	YES
PYGM	CCDS8079	chr11:64526120 C->A	p.M300I	7.55%	0.008%	Substitution	Non-synonymous	YES	YES
TP53	CCDS11118.1	chr17:7518898A->G	NA	6.97%	0.007%	NA	Splice site donor	YES	YES
MYO9B	CCDS46010	chr19:17294680 ->A	NA	6.56%	0.007%	Substitution	Splice site donor	YES	YES
TUBBP5	ENST00000290377	chr9:141070139 C->T	p.L37F	6.49%	0.006%	Substitution	Non-synonymous	YES	NO
FOXC1	CCDS4473	chr6:1612018 CGG->	NA	5.96%	0.006%	Deletion	Frameshift	NO	NO

FIG. 15B

GALNT9	ENST00000424720	chr12:132824394 C->A	p.M1146I	5.44%	0.005%	Substitution	Non-synonymous	YES	YES
RP11-830F9.6	ENST00000378347	chr16:89017605 A->G	p.D1079G	5.29%	0.005%	Substitution	Non-synonymous	YES	YES
SRD5A2	ENST00000233139	chr2:31805883 C->G	p.K87N	5.06%	0.005%	Substitution	Non-synonymous	YES	YES
SMG7	CCDS1355	chr1:183515267 ->A	NA	4.98%	0.005%	Insertion	Frameshift	N/A	N/A
LZTS3	CCDS13049	chr20:3146451 G->A	p.R1015W	4.86%	0.005%	Substitution	Non-synonymous	N/A	N/A
TNFSF9	CCDS12169	chr19:6531149 GCT->	NA	4.74%	0.005%	Deletion	Frameshift	N/A	N/A
NOP9	CCDS9624	chr14:24769850 ->GAG	NA	4.63%	0.005%	Insertion	Frameshift	N/A	N/A
ABCA2	CCDS43909	chr9:139917492 G->A	p.A176V	4.28%	0.004%	Substitution	Non-synonymous	N/A	N/A
SHC1	CCDS44233	chr1:154938890 ->C	NA	4.14%	0.004%	Insertion	Frameshift	N/A	N/A
RNF43	CCDS11607.1	chr17:53792509C->A	p.E318X	3.99%	0.004%	Substitution	Nonsense	N/A	N/A
RNF43	CCDS11607.1	chr17:53790631C->A	p.S502I	3.66%	0.004%	Substitution	Non-synonymous	N/A	N/A
SMAD4	CCDS11950.1	chr18:46836279C->A	p.Y195X	3.45%	0.003%	Substitution	Nonsense	N/A	N/A
IGSF3	CCDS30814	chr1:117158745 A->C	p.D378E	3.42%	0.003%	Substitution	Non-synonymous	N/A	N/A
FOLR3	ENST00000456237	chr11:71850159 A->T	p.E449V	3.27%	0.003%	Substitution	Non-synonymous	N/A	N/A
CTRC	CCDS156	chr1:15771154 C->T	p.H547Y	3.20%	0.003%	Substitution	Non-synonymous	N/A	N/A
NRAP	CCDS7579	chr10:115348730 T->G	NA	3.08%	0.003%	Substitution	Non-synonymous	N/A	N/A
RNF43	CCDS11607.1	chr17:53790810G->T	p.D442E	3.07%	0.003%	Substitution	Non-synonymous	N/A	N/A
SH3BP5	CCDS2625.2	chr3:153733633 TCC->	NA	3.04%	0.003%	Deletion	Frameshift	N/A	N/A
WIZ	ENST00000389282	chr19:15549933 C->A	p.E1728D	2.96%	0.003%	Substitution	Non-synonymous	N/A	N/A
HDC	CCDS10134	chr15:50557790 C->A	p.G31W	2.85%	0.003%	Substitution	Non-synonymous	N/A	N/A

FIG. 15C

UBQLNL	CCDS31385	chr11:55366683 C->T	p.R989Q	2.82%	0.003%	Substitution	Non-synonymous	N/A	N/A
CDKN2A	CCDS6511.1	chr9:21961019C->T	p.A169T	2.69%	0.003%	Substitution	Non-synonymous	N/A	N/A
AHDC1	CCDS30662	chr1:27877505 G->	NA	2.40%	0.002%	Deletion	Frameshift	N/A	N/A
PTPRS	CCDS45930	chr19:5208036 G->A	p.T5675M	2.34%	0.002%	Substitution	Non-synonymous	N/A	N/A
TMEM63A	CCDS31042.1	chr1:226034840 CTG->	NA	2.13%	0.002%	Deletion	Frameshift	N/A	N/A
NOTCH4	CCDS34420.1	chr6:32163554 ->C	NA	2.02%	0.002%	Insertion	Frameshift	N/A	N/A
AZGP1	CCDS5680.1	chr7:99573573 T->C	p.Q71R	1.94%	0.002%	Substitution	Non-synonymous	N/A	N/A
TBC1D29	CCDS32606.1	chr17:28890361 G->A	p.R371Q	1.88%	0.002%	Substitution	Non-synonymous	N/A	N/A
GIGYF2	CCDS46542.1	chr2:233697764 GCA->	NA	1.84%	0.002%	Deletion	Frameshift	N/A	N/A
KIF7	CCDS32325.2	chr15:90171904 G->	NA	1.81%	0.002%	Deletion	Frameshift	N/A	N/A
TP53I13	CCDS42289.1	chr17:27899242 C->T	p.S596F	1.79%	0.002%	Substitution	Non-synonymous	N/A	N/A
PCNX3	CCDS44650.1	chr11:65385491 C->A	p.L658M	1.72%	0.002%	Substitution	Non-synonymous	N/A	N/A
EIF3J	CCDS10111.1	chr15:44829395 GGC->	NA	1.71%	0.002%	Deletion	Frameshift	N/A	N/A
SFSWAP	CCDS9273.1	chr12:132281734 AGA->	NA	1.60%	0.002%	Deletion	Frameshift	N/A	N/A
TMC4	CCDS12882.1	chr19:54675747 TCC->	NA	1.55%	0.002%	Deletion	Frameshift	N/A	N/A
IQSEC2	CCDS48130.1	chrX:53264131 TGG->	NA	1.43%	0.001%	Deletion	Frameshift	N/A	N/A
C6orf132	CCDS47428.1	chr6:42072710 CTC->	NA	1.36%	0.001%	Deletion	Frameshift	N/A	N/A
ARX	CCDS14215.1	chrX:25025356 GGC->	NA	1.34%	0.001%	Deletion	Frameshift	N/A	N/A
NHS	ENST00000380057	chrX:17705855 C->T	p.L22F	1.24%	0.001%	Substitution	Non-synonymous	N/A	N/A
RAB40C	CCDS10413.1	chr16:677580 ->C	NA	1.22%	0.001%	Insertion	Frameshift	N/A	N/A

FIG. 15D

C16orf91	CCDS32360.1	chr16:1476278 C->A	p.Q345H	1.06%	0.001%	Substitution	Non-synonymous	N/A	N/A
EME2	CCDS32364.1	chr16:1824298 TGC->	NA	1.04%	0.001%	Deletion	Frameshift	N/A	N/A
ATP13A2	CCDS175.1	chr1:17318887 G->	NA	0.94%	0.001%	Deletion	Frameshift	N/A	N/A
ZIC3	CCDS14663.1	chrX:136648985 CGC->	NA	0.74%	0.001%	Deletion	Frameshift	N/A	N/A
NAA30	CCDS32088.1	chr14:57858199 AGG->	NA	0.61%	0.001%	Deletion	Frameshift	N/A	N/A
NDUFS3	CCDS7941.1	chr11:47600646 GGC->	NA	0.58%	0.001%	Deletion	Frameshift	N/A	N/A
CCKBR	CCDS7761.1	chr11:6292451 ->T	NA	0.49%	0.000%	Insertion	Frameshift	N/A	N/A
RUSC2	CCDS35008.1	chr9:35560102 GCT->	NA	0.46%	0.000%	Deletion	Frameshift	N/A	N/A
NOS2	CCDS11223.1	chr17:26099407 G->A	p.A1631V	0.42%	0.000%	Substitution	Non-synonymous	N/A	N/A
WASL	CCDS34743.1	chr7:123388755 GCG->	NA	0.38%	0.000%	Deletion	Frameshift	N/A	N/A
SNRNP35	CCDS45005.1	chr12:123950763 GA->	NA	0.31%	0.000%	Deletion	Frameshift	N/A	N/A
NBPF14	CCDS30836.1	chr1:148009468 C->G	p.E1839D	0.29%	0.000%	Substitution	Non-synonymous	N/A	N/A
FAM104B	CCDS35305.1	chrX:55172647 G->A	p.P221L	0.27%	0.000%	Substitution	Non-synonymous	N/A	N/A
TFEB	CCDS4858.1	chr6:41658830 TGC->	NA	0.25%	0.000%	Deletion	Frameshift	N/A	N/A
SEPT9	CCDS45790.1	chr17:75478417 G->A	p.G913R	0.22%	0.000%	Substitution	Non-synonymous	N/A	N/A
ROR2	CCDS6691.1	chr9:94486026 TCC->	NA	0.15%	0.000%	Deletion	Frameshift	N/A	N/A
PKD2	CCDS3627.1	chr4:88929174 GAG->	NA	0.12%	0.000%	Deletion	Frameshift	N/A	N/A
GLIS2	CCDS10511.1	chr16:4384871 G->	NA	0.08%	0.000%	Deletion	Frameshift	N/A	N/A
HECTD4	CCDS44978.1	chr12:112622808 T->	NA	0.02%	0.000%	Deletion	Frameshift	N/A	N/A
MUC4	NM_018406	chr3:195508284 C->A	p.Q10167H	0.01%	0.000%	Substitution	Non-synonymous	N/A	N/A

FIG. 15E

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/019788

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 15/10; C12Q 1/68 (2018.01)
CPC - C12N 15/1006; C12N 15/1065; C12Q 1/6816; C12Q 1/6837; C12Q 1/6876; C12Q 2600/16 (2018.05)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 435/6.11; 506/4; 506/16; 536/24.3 (keyword delimited)

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2007/0281317 A1 (BECKER et al) 06 December 2007 (06.12.2007) entire document	1, 2, 6, 12-15, 18, 26, 29, 36, 39, 40, 43, 45-47, 50, 62, 67, 70, 71 --- 3-5, 7-11, 16, 17, 19-25, 27, 28, 30-35, 37, 38, 41, 42, 44, 48, 49, 51-61, 63-66, 68, 69, 72-79
Y	US 2016/0040218 A1 (THE BROAD INSTITUTE, INC. et al) 11 February 2016 (11.02.2016) entire document	3-5, 7-11, 17, 19-25, 27, 28, 41, 42, 44, 48, 49, 63-66, 68, 69
Y	US 5,219,726 A (EVANS) 15 June 1993 (15.06.1993) entire document	16, 17
Y	WO 1993/020236 A1 (APPLIED BIOSYSTEMS, INC.) 14 October 1993 (14.10.1993) entire document	30
Y	WO 2016/065192 A1 (ROKA BIOSCIENCE, INC.) 28 April 2016 (28.04.2016) entire document	31, 37, 38, 51-56
Y	US 2014/0287468 A1 (DIRECTED GENOMICS, LLC) 25 September 2014 (25.09.2014) entire document	32-35, 57-61, 75-79
Y	US 6,458,530 B1 (MORRIS et al) 01 October 2002 (01.10.2002) entire document	54, 55

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 May 2018	06 JUN 2018
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/019788

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 7,326,561 B2 (GOODMAN et al) 05 February 2008 (05.02.2008) entire document	56
Y	US 2008/0020371 A1 (GERMAN et al) 24 January 2008 (24.01.2008) entire document	72-74