

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

01 October 2020 (01.10.2020)



(10) International Publication Number

WO 2020/193769 A1

(51) International Patent Classification:

C12Q 1/6806 (2018.01) C12Q 1/6874 (2018.01)

Published:

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

(21) International Application Number:

PCT/EP2020/058791

(22) International Filing Date:

27 March 2020 (27.03.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

PCT/EP2019/057777

27 March 2019 (27.03.2019) EP

19200404.2 30 September 2019 (30.09.2019) EP

(71) Applicant: **DIAGENODE S.A.** [BE/BE]; Liège Science Parc, Rue Bois Saint-Jean 3, 4102 SERAING (BE).

(72) Inventors: **KROONEN, Jerome**; Rue Sainte-Marie 16/12, 4000 Liège (BE). **SABATEL, Celine**; Tige sur le Mont 21, 4550 Villers-le-Temple (BE). **BROCARD, Gilles**; Avenue de la Grotte, 83, 4130 Esneux (BE). **DURIEUX, Florence**; Clos St Roch, 10, 4683 Vivegnis (BE).

(74) Agent: **PRONOVEM**; c/o Office Van Malderen, Avenue Josse Goffin, 158, 1082 Brussels (BE).

(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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

(54) Title: A HIGH THROUGHPUT SEQUENCING METHOD AND KIT

(57) Abstract: A high throughput sequencing method and kit The present invention is in the field of diagnostic and sequencing technologies and is related to a high throughput sequencing method and a kit comprising tools for performing this method, that combine a capture and amplification by switching detection step, preferably the so-called "Capture and Amplification by Tailing and Switching" (CATS) and sequencing technology, preferably the so-called "Nanoballs sequencing" technologies.



WO 2020/193769 A1

5

A high throughput sequencing method and kit**Field of the invention**

[0001] The present invention is in the field of diagnostic and sequencing technologies and is related to a high throughput sequencing method and a kit comprising tools for performing this method, that combine a capture and amplification by switching detection step , preferably the so-called "Capture and Amplification by Tailing and Switching" (CATS) and sequencing technology, preferably the so-called "Nanoballs sequencing" technologies.

Background of the invention

15 [0002] A Capture and Amplification by Switching technology, especially the so-called "Capture and Amplification by Tailing and Switching"(CATS) technology is a ligase-free method to produce DNA libraries for a further sequencing from RNA or DNA and is described in the international patent application WO2015/173402-A1.

20 [0003] There is a need to improve RNA sequencing (or RNA-Seq) which is using the next generation sequencing (NGS) technologies to reveal the presence and quantity of RNA in a biological sample at a given moment, analyzing the continuously changing cellular transcriptome. The Capture and amplification by switching protocols, especially CATS protocol are more efficient for RNA-Seq library creation than protocols using ligase by incorporating adaptors during cDNA synthesis in a single reaction tube. In particular, the CATS technology allows optimal sequencing of sensitive, degraded, cell free RNA (cfRNAs) sequence, plasma derived RNA sequences, non-coding RNA (ncRAS) sequences such as miRNA sequences or long non-coding RNA (lncRNA sequences), exosomal RNA sequences, rare and low input RNA sample, that are efficient markers of different diseases, such as cancers.

25 [0004] Improved sequencing protocols, especially the "Nanoball sequencing" technology disclosed by Drmanac et al (Science 327: 5961, page 78-81 (2010)) require a fragmentation of genomic DNA, wherein individual fragments are used to

produce circular DNAs, in which platform specific oligonucleotides adapters separate genomic DNA sequence.

[0005] The obtained circular DNAs are amplified to generate advantageously single-stranded concatemers (DNA nanoballs (DNBs) that have a size of about 300 nanometers) that can be immobilized on a substrate at a specific location and that remain separated from each other, because of their negatively charges upon the patterned substrate containing up to 3 billion spots each spot containing one (and only one) DNA nanoball.

Aims of the invention

[0006] The present invention aims to provide a new detection and sequencing method and tools for performing such method that do not present the drawbacks of the method and kit of the state of the art.

[0007] A first aim of the present invention is to obtain a method and tools for performing this method that improve the nucleic acids libraries production and sequencing, especially of sensitive, degraded, chemically modified, cell free nucleic acid sequences, especially all kind of RNA sequences (coding or non-coding RNA sequences, miRNAs, MiscRNAs, piRNAs, rRNAs, siRNAs, snRNAs, snoRNAs, TRNAs,...), regardless of a spike-in possibly obtained from a single cell.

[0008] A further aim of the invention is to obtain such method and tools for performing this method that are easy to use, with minimal hands-on time; that are also robust and present an improved sensitivity and excellent reproducibility.

Definitions.

[0009] All literature and similar material cited in the application, including, but no limited to patents, patent applications, scientific articles, books and web pages are expressly incorporated by reference in their entirety to the description of the present invention.

[0010] Unless otherwise defined, all terms used in disclosing the invention, including technical and scientific terms, have the meaning as commonly understood by one of ordinary skilled in the art in the invention field.

[0011] A used in this description and claims, the singular forms “a”, “an”, and “the” include singular and plural referents, unless the content of the description clearly dictates otherwise.

[0012] The terms “comprising”, “comprises” and “composed of” are synonymous to “including” or “containing” and are inclusive and not open ended and do not exclude any additional, non-recited members, elements or methods steps.

[0013] The terms “one or more” or “at least one” are clear per se and encompasses a reference to any of these members, which means any two or more of the members and up to all members .

[0014] The term “about” as used herein, when referring to a measurable value, such as an amount of a compound, dose, time and the like is meant to encompass 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5% or even 0;1% of the specified amount or value.

[0015] As used in the description and claims, the terms “nucleic acid(s)” includes polymeric and oligomeric macromolecules, made of DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) known as nucleotides, comprising bases selected from the group consisting of Adenine (A), Thymine (T), Cytosine (C), Guanine (G) and Uracil (U).

[0016] The terms “single stranded nucleic acids” (ss nucleic acid) refer to a nucleic acid consisting of only one polynucleotide or oligonucleotide strand. In contrast a “double stranded nucleic acid “ (ds nucleic acid) consist of two polynucleotide or oligonucleotide strands wherein the majority of the nucleotides are paired according to known pairing rules.

[0017] The terms “genetic amplification” is a biochemical technology used in molecular biology for many years to amplify by primer sequences, a single or few copies of a piece or portion of DNA by replication and copy across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence. The most known genetic amplification technology is the so called “polymerase Chain Reaction or PCR” as described in US patents 4, 683, 195-B2 and 4, 683, 202- B2 using primers sequences and the heat stable DNA polymerase, such as the Taq polymerase obtained from Bacterium Thermus aquatic allowing thermal cycling.

[0018] The term “primer” refers to a oligonucleotide sequence, usually comprising between about 12 nucleotides and about 25 nucleotides, hybridizing specifically to a target sequence of interest and which functions as a substrate onto which nucleotides can be polymerized by a polymerase.

[0019] The terms “Template Switch Oligo” or TSO, refer to an oligo that hybridizes to untemplated C nucleotides added by a reverse transcriptase during reverse transcription.

Summary of the invention

[0020] The present invention is related to a high throughput (detection and) sequencing method of a nucleic acid strand sequence as well as tools (preferably

included into a kit) for performing this method, this (detection and sequencing) method comprising at least (or consisting of the steps of, preferably the consecutive steps of :

- providing a sample, especially liquid or solid biopsies, such as a blood sample, preferably the plasma, a tissue sample, a fossil, a single cell sample or even targeted compartments of cells (nucleus, endoplasmic reticulum, ...) , this sample comprising a native single stranded nucleic acid sequence or native double stranded nucleic acid sequence,
- possibly fragmenting the native single stranded nucleic acid sequence or native double stranded nucleic acid sequence, into smaller nucleic acid sequence fragments
- possibly denaturing the native double strand nucleic acid sequence(s)
- possibly end-repairing the native nucleic acid sequences
- possibly adding at least 5 consecutive nucleotides to the 3-terminus of the native single stranded or native double stranded nucleic acid sequence or their fragments,
- hybridizing a priming oligonucleotide sequence complementary to the added nucleotide sequence and synthesizing a cDNA sequence with a template dependent DNA polymerase to obtain a double stranded nucleic acid sequence,
- hybridizing a template switching oligonucleotide (TSO) to this generated double stranded nucleic acid sequence,
- extending the 3' end of the cDNA strand to synthesize a double stranded nucleic acid sequence, wherein one strand of the nucleic acid sequence comprises the priming oligonucleotide and a cDNA sequence that is complementary to the single stranded nucleic acid sequence and to the template switching oligonucleotide sequence. When the reverse transcriptase reaches the 5'-end of the nucleic acid sequence, it switches the template and continue DNA synthesis over the template-switching oligonucleotide (TSO). The TSO containing three 3'-terminal ribonucleotides x (rX) facilitates the template switching and carry adaptor sequence,
- possibly adding a splint oligo sequence that hybridize to adapter DNA sequences which are ligated to form a circle and adding an exonuclease to remove all remaining single stranded and double stranded DNA products to collect only circular DNA template,
- obtaining DNA nanoballs (DNBs) by performing a rolling circle replication of the synthesized stranded nucleic acid sequence,
- fixing the DNA nanoballs (DNBs) on a patterned array flow cell,

- performing a base sequencing, preferably this base sequencing is obtained by a method selected from the group consisting of synthesis, ligase base sequencing or pyrosequencing, and
- obtaining an identification of each nucleotide of the native nucleic acid sequence, preferably by nanopore sequencing or imaging, more preferably upon a high resolution CCD camera.

[0021] In the method of the invention, the synthesized double stranded nucleic acid sequences present a length preferably comprised between about 200 and about 500 nucleotides.

[0022] According to the invention, the native single stranded nucleic acid sequence or native double stranded nucleic acid sequence is preferably selected from the group consisting of fragmented and/or bisulfite-converted DNA sequence, mRNA sequence, miRNA sequence small RNA sequence, piRNA sequence, bisulfite-converted RNA or a mixture thereof.

[0023] In the method according to the invention, the at least 5 consecutive identical nucleotides are preferably selected from the group consisting of ribo, deoxy-ribonucleotides or didesoxy-ribonucleotides of A, T, C, G or U, that are preferably added by an enzyme selected from the group consisting of a poly(A)-polymerase, poly(U)-polymerase, poly(G)-polymerase, terminal transferase, DNA ligase, RNA ligase and the dinucleotides and the trinucleotides RNA ligases.

[0024] Another aspect of the invention concerns an apparatus or a sequencing kit for performing the method of the invention, this kit or apparatus comprising (or consisting of) the following reagents present in suitable vials

- a reagent capable of adding nucleotides to the 3-terminus of a single stranded nucleic acid,
- reagents for a genetic amplification, preferably reagents for performing a PCR amplification a reverse transcriptase enzyme,
- a priming oligonucleotide,
- a template switching oligonucleotide,
- a rolling circle replication enzyme, preferably the Phi 29 DNA polymerase
- possibly reagents for cyclization,
- possibly a patterned flow cell and
- possibly a template independent DNA or RNA polymerase and a blocking nucleotide, such as 3d-NTP, 3-Me-NTP and ddNTP, and
- possibly (written) instructions for performing the method steps of the invention.

[0025] In the method, apparatus and kit according to the invention, the priming oligonucleotide preferably comprises the nucleotide sequence disclosed in claims 9 to 12 and claims 19 and 20 of WO2015/173402 incorporated herein by reference.

[0026] Advantageously, in the method, apparatus and kit according to the invention, the rolling cycle amplification is obtained by addition of a sufficient amount of the Phi 29 DNA polymerase, this enzyme allowing a production of concatemers or DNA nanoballs (DNBs) into a long single stranded DNA sequence comprising several head-to-tail copies of the circular template, wherein the resulting nanoparticle self assembles into a tight ball of DNA.

[0027] This polymerase replicates the looped DNA and when it finishes one circle, it does not stop-it, continues the replication by peeling off its -previously copied DNA. This copying process continues over and over, forming the DNA nanoball this large mass of repeating DNA to be sequenced all connected together.

[0028] Preferably, in the method, apparatus and kit according to the invention the patterned array flow cell is a silicon wafer coated with silicon dioxide, titanium, hexamethyldisilazane (HDMS) and a photoresist material and each DNA nanoball selectively binds to the positively-charged aminosilane according to the pattern.

[0029] Advantageously, in the method of the invention, the ligase base sequencing is obtained by adding dNTP incorporated by polymerase, each dNTP being preferably conjugated to a particular label or comprises a modification that allows their future detection through a binding with one more labeled antibody(ies) (CoolNGS ® technology improved in sensitivity and less costly for obtaining more accurate and longer reads), preferably a label being a fluorophore or dye and possibly containing a termination blocking addition extension, wherein unincorporated dNTPs are washed, wherein image is captured, wherein dye and terminator are preferably cleaved and wherein these steps are repeated until sequencing is complete.

[0030] The CoolNGS technology is based the use of multiple fluorescent dye molecules attached to the antibodies providing a higher signal-to-noise ratio and reduced consumption of expensive materials, together with incorporating natural bases with no interference between sequencing cycles.

[0031] In addition, in the method of the invention, the added fluorophore is excited with a laser that excites specific wavelength of light and the emission of fluorescence from each DNA nanoball is captured on high resolution CCD camera and wherein the color of each DNA nanoball corresponding to a base to the interrogative position and wherein the computer records the base position information.

[0032] A last aspect of the invention concerns the use of the apparatus, the kit or the method according to anyone of the preceding claims. The preferred use is

proposed for sequencing or expression analysis, for cloning labelling, for the identification of genes or mutation, in detection of human or animal disease or forensic science, for the analysis of infectious diseases and genomes of viruses, bacteria, fungi, animals or plant, including their derived cells, for the characterization of plants, fruits, breeding checks detection of plants or fruits diseases.

[0033] The present invention will be described hereafter in the following examples presented as non-limiting preferred embodiments of the present invention

[0034] **Examples**

[0035] The following table 1 represents an outline of the experiments conducted to validate the method of the invention, with capture and amplification by switching detection, such as the CATS small RNA-seq construct (from Diagenode) on a DNBSEQ-G400 (CoolMps system for "Nanoballs sequencing" (from MGI)).

[0036] In the table 1, L02 = 15 % spike-in; L03 = 0% spike-in and L05 = 15 % spike-in . The libraries have been sequenced in three different sequencing lanes to address technical variability in the sequencing.

[0037] The applicant has obtained the averaged per base sequence distribution of the samples sequenced in lane 03. This distribution displays a typical Capture and amplification by switching detection construct, being CATS small RNA-seq construct with a short insert short insert size in conformity with the nature of the RNAs sequenced (small non coding RNAs) and also displays the expected poly(A) tail synthesized during library preparation after the small RNA reads.

[0038] The N content is non-null, but low enough not to cause problems later on during data analysis. The template Switch motif (Template switching oligonucleotide TSO) is absent from the first (1-3) sequencing cycles as the sequencing was done in dark cycling mode for those cycles.

[0039] The applicant has obtained also the averaged quality distribution of the DNBs sequenced in lane 03. As the vast majority of the DNBs (> 85%) across lane 03 obtained present a quality score above 30 (lower part of the figure 2) which makes the sequencing of CATS small RNA libraries on the DNBSEQ-G400 system an efficient and high quality sequencing system.

[0040] With the method of the invention, the applicant has selected reads allocated per sample (#index n°) in the different sequencing lanes and mean Q30% for the samples in the different sequencing lanes. The obtained results show that the libraries are able to be sequenced normally, regardless of a spike-in and produce high quality reads (Q30 > 85%).

[0041] Furthermore, the relative proportion of mapped reads (%) out of the trimmed reads were obtained for the different samples across the different lanes. Most of the

reads after filtering and trimming are mapping (STAR) to a reference genome (hg19) to an expected percentage for a CATS small RNA library. The sequencing method was performed in the different lanes, regardless of a spike-in content, do not impact the mapping stats. This means that the sequencing method and system according to the invention is reproducible across lanes.

[0042] The complete diversity biotyping at TPM higher or equal to 2 of the libraries sequenced in lane 03 was obtained by using the Ensembl annotations. most of the library contents are annotated as non-coding RNA, even though a certain fraction is coming from protein coding transcripts, constituting products of degradation, that are captured during library preparation. This biotyping representation is unexpectedly totally in accordance with libraries representation obtained by the state of the art methods and systems, especially the so-called Illumina (ILMN) sequencing method and system.

[0043] The small non-coding RNA diversity biotyping at TPM higher or equal to 2 of the libraries sequenced in lane 03 was obtained by using the Ensembl annotations. The noncoding RNAs spanning a wide diversity of small non-coding RNAs ranging from miRNAs to snoRNAs identified by the method and system of the invention. Therefore, the claimed method and system according to the invention are as efficient as the known methods and systems of the state of the art, especially the so-called Illumina (ILMN) sequencing method and system.

Claims

1. A high throughput sequencing method of a nucleic acid strand sequence comprising the steps of
 - providing a sample, single cell sample, comprising a native single stranded nucleic acid sequence or native double stranded nucleic acid sequence,
 - possibly fragmenting the native single stranded nucleic acid sequence or native double stranded nucleic acid sequence, into smaller nucleic acid sequence fragments
 - possibly denaturing the native double strand nucleic acid sequence(s)
 - possibly end-repairing the native nucleic acid sequences,
 - hybridizing a priming oligonucleotide sequence complementary to the added nucleotide sequence and synthesizing a cDNA sequence with a template dependent DNA polymerase to obtain a double stranded nucleic acid sequence
 - hybridizing a template switching oligonucleotide to the said generated double stranded nucleic acid sequence,
 - extending the 3' end of the cDNA strand to synthesize a double stranded nucleic acid sequence, wherein one strand of the nucleic acid sequence comprises the priming oligonucleotide and a cDNA sequence that is complementary to the single stranded nucleic acid sequence and to the template switching oligonucleotide sequence,
 - adding a splint oligo sequence that hybridize to adapter DNA sequences which are ligated to form a circle and adding an exonuclease to remove all remaining single stranded and double stranded DNA products to collect only circular DNA template,
 - obtaining DNA nanoballs (DNBs) by performing a rolling circle replication of the synthesized double stranded nucleic acid sequence,
 - fixing the DNA nanoballs (DNBs) on a patterned array flow cell,
 - performing a base sequencing and

obtaining an identification, preferably of each nucleotide of the native nucleic acid sequence, by nanopore sequencing or imaging preferably upon a high resolution CCD camera.
2. The method of claim 1, which comprises the step of adding at least 5 consecutive nucleotides, to the 3-terminus of the native single stranded or native double stranded nucleic acid sequence or their fragments, before the hybridizing step and wherein

the at least 5 consecutive nucleotides are selected from the group consisting of ribo-, desoxy-ribonucleotides or didesoxy-ribonucleotides of A, T, C, G or U.

3. The method of claim 2, wherein the nucleotides are added by an enzyme selected from the group consisting of a poly(A)-polymerase, poly(U)-polymerase, poly(G)-polymerase, terminal transferase, DNA ligase, RNA ligase and the dinucleotides and the trinucleotides RNA ligases

4. The method according to any one of the preceding claims 1 to 3, the synthesized double stranded nucleic acid sequences present a length comprised between 200 and 500 nucleotides.

5. The method according to any one of the preceding claims 1 to 4, wherein the native single stranded nucleic acid sequence or native double stranded nucleic acid sequence is selected from the group consisting of fragmented and/or bisulfite-converted DNA sequence, mRNA sequence, miRNA sequence small RNA sequence, piRNA sequence, bisulfite-converted RNA or a mixture thereof.

6. The method according to any of the preceding claims 1 to 5, wherein the rolling cycle amplification is obtained by addition of a sufficient amount of the Phi 29 DNA polymerase.

7. The method according to any of the preceding claims 1 to 6, wherein the rolling circle replication allows the production of concatemers or DNA nanoballs (DNBs) into a long single stranded DNA sequence comprising several head-to-tail copies of the circular template, wherein the resulting nanoparticle self assembles into a tight ball of DNA.

8. The method according to any one of the preceding claims 1 to 7, wherein the patterned array flow cell is a silicon wafer coated with silicon dioxide, titanium, hexamethyldisilazane (HDMS) and a photoresist material.

9. The method according to any one of the preceding claims 1 to 8, wherein each DNA nanoball selectively binds to the positively charged aminosilane according to a pattern.

10. The method according to any one of the preceding claims 1 to 9, wherein the ligase base sequencing is obtained by adding dNTP incorporated by polymerase, each dNTP either being modified to be recognized by one or more labelled antibody(ies) or being conjugated to a particular label, preferably a label being a fluorophore and containing a termination blocking addition extension, wherein unincorporated dNTPs are washed, wherein image is captured, wherein dye and terminator are preferably cleaved and wherein these steps are repeated until sequencing is complete.

11. The method according to the claim 10, wherein the added fluorophore is excited with a laser that emits specific wavelength of light and the emission of fluorescence from each DNA nanoball is captured on high resolution CCD camera, wherein the color of each DNA nanoball corresponding to a base to the interrogative position and wherein the computer records the base position information.

12. A sequencing kit comprising

- a reagent capable of adding nucleotides to the 3-terminus of a single stranded nucleic acid,
- an end-repair enzyme,
- reagents for a genetic amplification, preferably a PCR genetic amplification,
- a reverse transcriptase enzyme,
- a priming oligonucleotide,
- a template switching oligonucleotide and
- a rolling circle replication enzyme

13. The kit of claim 12, further comprising a patterned flow cell.

14. The kit of claim 13, wherein the patterned flow cell is a silicon wafer coated with silicon dioxide, titanium, hexamethyldisilazane (HDMS) and a photoresist material

15. The kit according to any one of the preceding claim 12 to 14, wherein the reagent is a template independent DNA or RNA polymerase and a blocking nucleotide, such as 3d-NTP, 3-Me-NTP and ddNT.

16. The kit according to any one of the preceding claims 12 to 15, wherein the rolling circle replication enzyme is the Phi 29 DNA polymerase.

17. Use of the kit or the method according to anyone of the preceding claims, for sequencing or expression analysis, cloning labelling, for the identification of genes or mutation, in personalized medicine, therapy monitoring, prediction, prognosis, early detection of human or animal disease or forensic science, analysis of infectious diseases and genomes of viruses, bacteria, fungi, animals or plant, including their derived cells, characterization of plants, fruits, breeding checks detection of plants or fruits diseases.

Cod e	small_25pg_ a_L2F15_i1	small_25pg_ b_L2F15_i2	small_250pg_ a_L2F15_i3	small_250pg_ b_L2F15_i4	small_1ng_ a_L2F15_i13	small_1ng_ b_L2F15_i14	total_250pg_ a_L2F15_i15	total_250pg_ b_L2F15_i16	total_2500pg_ a_L2F15_i97	total_2500pg_ b_L2F15_i98	total_10ng_ a_L2F15_i99	total_10ng_ b_L2F15_i100
Tem plate	Isolated small RNA fraction (17-30 nt) from HepG2 total RNA						Total RNA fraction of HepG2 cells					
Input for librar y prep	25pg		250pg		1ng		250pg		2,5ng		10ng	
Tech nical repli cate	1	2	1	2	1	2	1	2	1	2	1	2
Raw reads allocated per sample												
L02	5,5E+07	3,2E+07	3,2E+07	3,3E+07	2,2E+07	2,8E+07	4,6E+07	2,7E+07	2,3E+07	3,4E+07	4,1E+07	3,0E+07
L03	6,0E+07	3,4E+07	3,6E+07	3,7E+07	2,6E+07	3,0E+07	5,1E+07	3,0E+07	2,8E+07	3,6E+07	4,5E+07	3,4E+07
L04	5,3E+07	3,0E+07	3,0E+07	3,1E+07	2,0E+07	2,6E+07	4,3E+07	2,5E+07	2,1E+07	3,2E+07	3,9E+07	2,7E+07

Table1. Outline of the experiment conducted to validate the CATS small RNA-seq construct on the DNBSEQ G400 (CoolMPS). It can be noticed that the libraries have been sequenced in three different sequencing lanes to address technical variability in the sequencing. L02 = 15% spike-in; L03 = 0% spike-in; L04 = 15% spike-in.

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/058791

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6806 C12Q1/6874
ADD.

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

B. FIELDS SEARCHED

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

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2014/242581 A1 (JOHNSON MARK T [US]) 28 August 2014 (2014-08-28) [0005], [0018], fig. 5 and [0067], [0068], [0131], [0137].	1-17
Y	& ZHU Y Y ET AL: "REVERSE TRANSCRIPTASE TEMPLATE SWITCHING: A SMART APPROACH FOR FULL-LENGTH CDNA LIBRARY CONSTRUCTION", BIOTECHNIQUES, INFORMA HEALTHCARE, US, vol. 30, no. 4, 1 April 2001 (2001-04-01), pages 892-897, XP001121210, ISSN: 0736-6205 Fig. 1 and p. 894-895 "Principle of the SMART library construction" ----- -/--	1-17



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

19 June 2020

Date of mailing of the international search report

30/06/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Lapopin, Laurence

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/058791

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LONGQI LIU ET AL: "An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos", NATURE COMMUNICATIONS, vol. 10, no. 1, 21 January 2019 (2019-01-21), pages 1-11, XP055672600, DOI: 10.1038/s41467-018-08244-0 p. 9 and "material and methods" -----	1-17
Y	T. W. NILSEN: "Splinted Ligation Method to Detect Small RNAs", COLD SPRING HARBOR PROTOCOLS, vol. 2013, no. 1, 1 January 2013 (2013-01-01), XP055705878, DOI: 10.1101/pdb.prot072611 abstract and fig. p. 55 -----	1
A	WO 2018/005811 A1 (GRAIL INC [US]) 4 January 2018 (2018-01-04) [0005], [0061] -----	1-17

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2020/058791

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2014242581 A1	28-08-2014	EP 2958574 A1	30-12-2015
		US 2014242581 A1	28-08-2014
		US 2016186262 A1	30-06-2016
		US 2017044610 A1	16-02-2017
		US 2018195123 A1	12-07-2018
		WO 2014116881 A1	31-07-2014

WO 2018005811 A1	04-01-2018	AU 2017290237 A1	06-12-2018
		CA 3024984 A1	04-01-2018
		CN 109563541 A	02-04-2019
		EP 3478856 A1	08-05-2019
		US 2018002749 A1	04-01-2018
		US 2019062831 A1	28-02-2019
		WO 2018005811 A1	04-01-2018
