



- (51) **International Patent Classification:**  
C12N 15/10 (2006.01) C12Q 1/68 (2006.01)
- (21) **International Application Number:**  
PCT/EP2016/066876
- (22) **International Filing Date:**  
15 July 2016 (15.07.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**  
15177361.1 17 July 2015 (17.07.2015) EP  
16164608.8 11 April 2016 (11.04.2016) EP
- (71) **Applicant:** MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER WISSENSCHAFTEN E.V. [DE/DE]; Hofgartenstraße 8, 80539 München (DE).
- (72) **Inventors:** ILIK, Ibrahim Avsar; Untere Waldstr. 8, 79194 Gundelfingen (DE). AKHTAR, Asifa; Johann-Sebastian-Bach-Str. 47, 79104 Freiburg (DE). ILIK, Tugce Aktas; Untere Waldstr. 8, 79194 Gundelfingen (DE).
- (74) **Agent:** VOSSIUS & PARTNER; Siebertstraße 3, 81675 München (DE).

- (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, 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).

**Published:**

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

(54) **Title:** CLONING OF SINGLE-STRANDED NUCLEIC ACID

(57) **Abstract:** The present invention relates to an oligonucleotide comprising (a) a double-stranded portion, which double-stranded portion is DNA and 9 to 30 base in length; (b) a loop connecting the 3' end of the first strand of said double-stranded portion with the 5' end of the second strand of said double-stranded portion, said loop comprising, in 5' to 3' direction: (ba) a first DNA portion which is 4 to 20 nucleotides in length; (bb) a non-nucleic acid spacer which (i) does not interfere with the formation of a stem-loop by said oligonucleotide; and (ii) causes polymerases to cease; and (be) a second DNA portion which is 4 to 20 nucleotides in length; wherein said first DNA portion and said second DNA portion are not complementary to each other; (c) a single-stranded overhang at its 5' end, said overhang being 5 to 40 nucleotides in length, wherein (ca) the bond connecting said double-stranded portion with the nucleotide of said overhang which is directly adjacent to said double-stranded portion is cleavable under alkaline conditions; and (cb) said overhang optionally comprises a barcode sequence, said barcode sequence preferably being 5 to 10 nucleotides in length; and optionally (d) within one, more or all of (a), (b) and (c), one, more or all of the following: (da) one or more modified nucleotides; (db) one or more sequences conferring compatibility with nucleic acid sequencing kits, such compatibility being preferably the presence of regions within said oligonucleotide which are complementary to primers comprised in said sequence kits; and (dc) one or more random bases.



5

### Cloning of single-stranded nucleic acid

10 The present invention relates to an oligonucleotide comprising (a) a double-stranded portion, which double-stranded portion is DNA and 9 to 30 base in length; (b) a loop connecting the 3' end of the first strand of said double-stranded portion with the 5' end of the second strand of said double-stranded portion, said loop comprising, in 5' to 3' direction: (ba) a first DNA portion which is 4 to 20 nucleotides in length; (bb) a non-nucleic acid spacer which (i) does not  
15 interfere with the formation of a stem-loop by said oligonucleotide; and (ii) causes polymerases to cease; and (bc) a second DNA portion which is 4 to 20 nucleotides in length; wherein said first DNA portion and said second DNA portion are not complementary to each other; (c) a single-stranded overhang at its 5' end, said overhang being 5 to 40 nucleotides in length, wherein (ca) the bond connecting said double-stranded portion with the nucleotide of said  
20 overhang which is directly adjacent to said double-stranded portion is cleavable under alkaline conditions; and (cb) said overhang optionally comprises a barcode sequence, said barcode sequence preferably being 5 to 10 nucleotides in length; and optionally (d) within one, more or all of (a), (b) and (c), one, more or all of the following: (da) one or more modified nucleotides; (db) one or more sequences conferring compatibility with nucleic acid sequencing kits, such  
25 compatibility being preferably the presence of regions within said oligonucleotide which are complementary to primers comprised in said sequence kits; and (dc) one or more random bases.

In this specification, a number of documents including patent applications and manufacturer's  
30 manuals is cited. The disclosure of these documents, while not considered relevant for the patentability of this invention, is herewith incorporated by reference in its entirety. More specifically, all referenced documents are incorporated by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

35

Various methods to study protein-DNA interaction in living cells are available. These methods generally use formaldehyde cross-linking to fix DNA-protein interactions *in vivo*, after which the cells are sonicated to release pieces of chromatin which contain DNA fragments of a size in the range of ~150-500 bp. DNA is then isolated, cloned using commercially available kits and

sequenced. This is now considered a standard practice and results of numerous such experiments are publicly available.

5 Studying protein-RNA interactions is more challenging. Even though there are a few protocols, these methods are not widely applied because they are labor-intensive and require the use of radioactivity.

10 In more detail, an established protocol typically involves: (i) cross-linking of the protein of interest to RNA with UV-C or a chemical cross-linker such as formaldehyde; (ii) immunoprecipitating the protein with a specific antibody; (iii) isolating bound RNA; (iv) cloning of the ssRNA with a commercially available kit; (v) sequencing; and (vi) analyzing the sequences determined.

15 The resulting data is usually extremely convoluted due to the presence of contaminating RNA such as tRNAs, snRNAs, ribosomal RNAs etc. Secondly, the sequenced fragments will have a broad profile, i.e. most RNA-binding proteins will recognize 6-12 nts, but the fragments that are sequenced will be ~100-200 nt without any positional information as to where the binding of the RNA-binding proteins might have happened in the cell.

20 More recent protocols (CLIP, HITS-CLIP, PAR-CLIP, iCLIP, CRAC, CLASH) circumvent these shortcomings. At the end of the protocol, often relatively clean data (much less contamination from tRNAs, snRNAs, ribosomal RNAs) are obtained. Also, the position of the cross-linking site may be determined. Drawbacks include that these protocols are difficult, especially to the extent they rely on the use of radioactivity ( $\gamma$ -<sup>32</sup>P ATP is used to mark bound RNAs); and  
25 generally they take about a week to be completed.

WO2012/103154 describes means and methods of manipulating nucleic acid molecules, especially by amplification and adapter ligation. The document describes RNA/DNA chimeric stem/loop adapter/primers which are utilized during target nucleic acid manipulation. The  
30 document does not address the issue of library generation from single-stranded nucleic acid. Moreover, the notion of an adapter comprising a non-nucleic acid linker is alien to this document.

35 Given the shortcomings of the prior art, the technical problem underlying the present invention can be seen in the provision of alternative or improved means and methods of sequencing single-stranded nucleic acids, of producing libraries from single-stranded nucleic acids and of analyzing protein-nucleic acid interactions.

The solution to these technical problems is provided by the aspects and embodiments disclosed below.

In a first aspect, the present invention provides an oligonucleotide comprising (a) a double-stranded portion, which double-stranded portion is DNA and 9 to 30 base in length; (b) a loop connecting the 3' end of the first strand of said double-stranded portion with the 5' end of the second strand of said double-stranded portion, said loop comprising, in 5' to 3' direction: (ba) a first DNA portion which is 4 to 20 nucleotides in length; (bb) a non-nucleic acid spacer which (i) does not interfere with the formation of a stem-loop by said oligonucleotide; and (ii) causes polymerases to cease; and (bc) a second DNA portion which is 4 to 20 nucleotides in length; wherein said first DNA portion and said second DNA portion are not complementary to each other; (c) a single-stranded overhang at its 5' end, said overhang being 5 to 40 nucleotides in length, wherein (ca) the bond connecting said double-stranded portion with the nucleotide of said overhang which is directly adjacent to said double-stranded portion is cleavable under alkaline conditions; and (cb) said overhang optionally comprises a barcode sequence, said barcode sequence preferably being 5 to 10 nucleotides in length; and optionally (d) within one, more or all of (a), (b) and (c), one, more or all of the following: (da) one or more modified nucleotides; (db) one or more sequences conferring compatibility with nucleic acid sequencing kits, such compatibility being preferably the presence of regions within said oligonucleotide which are complementary to primers comprised in said sequence kits; and (dc) one or more random bases.

The term "oligonucleotide" (sometimes briefly "oligo") has its art-established meaning. It refers to a polycondensate of nucleotides. Nucleotides may be either ribonucleotides, deoxyribonucleotides and modified nucleotides. Modified nucleotides, the presence of which is optional, may be locked nucleotides. The use of locked nucleotides is a means of stabilizing the double-stranded portion.

The oligonucleotide of the invention is also referred to as "adapter" herein.

A requirement of a certain part of the oligonucleotide of the invention to be DNA implies that this part consists exclusively of deoxyribonucleotides or of deoxyribonucleotides and modified nucleotides. The presence of ribonucleotides is excluded.

Similarly, to the extent a part of the oligonucleotide is required to be RNA (this applies to a preferred embodiment disclosed below), this implies that this part either consists exclusively of

ribonucleotides or of ribonucleotides and modified nucleotides. Also in this context, preference is given to modified nucleotides which are locked nucleotides.

5 Preferably, the presence of modified nucleotides does not exceed 25%, 20%, 15%, 10%, 5%, 2% or 1% of the positions in the entire molecule and/or within anyone of (a), (ba), (bc) and (c).

10 Further envisaged modified nucleotides are nucleotides with modified bases and/or modifications of the ribose and/or modifications of the phosphate. Preferred modifications of the ribose include 2'-modifications such as 2'-O-methyl. Preferred modifications of the phosphate include thiophosphate.

15 In view of the definitions given above, an oligonucleotide according to the invention may comprise up to 140 nucleotides. For the ease of referencing, also for molecules of that length the term "oligonucleotide" is used herein. The minimal length, counting only nucleotides and not the non-nucleic acid spacer, is 31 nucleotides. The exemplary oligonucleotide disclosed further below and depicted in the figures has a length of 50 nucleotides.

20 The oligonucleotide according to the invention has three compulsory components, designated (a), (b) and (c).

25 The double-stranded portion in accordance with (a) is DNA in the sense of the definition given above. It may have length of 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 base pairs. Preference is given to perfect complementarity. Alternatively, one or more mismatches, for example one, two or three mismatches may be present. In either case the total number of nucleotides comprised in the double-stranded portion (a) is an even number between 18 and 60. The presence of bulges, preferably not more than 1, 2 or 3, while not being preferred, is not excluded. Given that a bulge results in non-identity of the numbers of nucleotides comprised in the first and second strand, respectively, it follows that the double-stranded portion may comprise a total number of nucleotides which is any integer value from 30 18 to 60.

35 The double-stranded portion, when considered in isolation, can be viewed as consisting of a first and a second strand. The first strand is the strand bearing the 5' overhang in accordance with feature (c). The second strand is the complementary strand.

The loop in accordance with (b) contains both DNA and a non-nucleic acid spacer. The two DNA portions in accordance with (ba) and (bc) each and independently of each other have a

length of between 4 and 20 nucleotides, more preferably between 5 and 10 nucleotides. Any other integer value falling in these intervals, including 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18 and 19 is also envisaged. First and second DNA portions (ba) and (bc) may have equal or different length.

5

The non-nucleic acid spacer is defined in functional terms in the broadest embodiment. The chemistry is not particularly limited. The requirement "does not interfere with the formation of a stem loop by said oligonucleotide" implies a certain degree of flexibility. In particular, any structural implementation which would not permit base-pairing of the two strands of the double-stranded portion in accordance with (a) is excluded. Preferred implementations are the subject of embodiments disclosed below.

10

Different from the two strands forming the double-stranded portion in accordance with (a), the first and second DNA portions in accordance with (ba) and (bc) are not complementary to each other. Preferably, the term "not complementary" relates to the absence of any complementarity between first and second DNA portions in accordance with (ba) and (bc). This does not exclude, though, that in a less preferred embodiment there may be 1, 2 or 3 base pairs between first and second DNA portions in accordance with (ba) and (bc).

15

In a preferred embodiment, the following applies. For the purpose of this preferred embodiment, two hypothetical fully complementary double-stranded molecules are considered. The first molecule consists of (i) a strand which in turn consists of the first strand of said double-stranded portion (a) and the adjacent DNA portion (ba), and (ii) a strand which is perfectly complementary thereto. Analogously, the second hypothetical fully double-stranded molecule consists of (iii) a strand which in turn consists of the second DNA portion (bc) and the adjacent second strand of the double-stranded portion (a), and (iv) a strand which is perfectly complementary thereto. It is understood that while in these two hypothetical molecules, each of (ba) and (bc) have a complementary sequence, this does not apply to the oligonucleotide of the invention. The melting temperature as of each of the two hypothetical fully double-stranded molecules is determined or calculated. According to the mentioned preferred embodiment, the difference in melting temperature between the two hypothetical molecules shall not exceed 5°C. This is a means of optimizing amplification by PCR.

20

25

30

Constituent (c) has a single-stranded overhang which is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nucleotides in length. The only compulsory feature of said overhang is defined by feature (ca).

35

A preferred structural implementation of requirement (ca), namely the first nucleotide at the 3' end of the 5' overhang being a ribonucleotide, is described further below.

5 A purpose of the oligonucleotide of the invention is to serve as an adapter in the generation of a library, preferably a sequencing library. A sequencing library is a library designed for sequencing. State of the art adapters are typically designed for nucleic acid material which is double-stranded, especially double-stranded DNA. The oligonucleotide according to the present invention is equipped with structural features which makes it useful for the generation of libraries from single-stranded nucleic acid. The term "single-stranded nucleic acid" includes  
10 single-stranded DNA as well as single-stranded RNA, single-stranded RNA being preferred.

A common single-stranded RNA is mRNA. As such, the oligonucleotide of the present invention is particularly useful for experiments which aim at elucidating protein-single-stranded RNA interactions, especially protein-mRNA interactions. The term "elucidating" in this context  
15 includes determining the type of interaction, more specifically the sequence of the single-stranded RNA interacting with a given protein and/or the specific site(s) of interaction between said protein and said single-stranded RNA.

In the alternative, a library is to be generated from single-stranded DNA as starting material.  
20 Single-stranded DNA may originate from various sources, said sources including 5'-3'-exonuclease-treated dsDNA. To the extent DNA-protein interactions are to be investigated, protein and DNA may be cross-linked, for example by formaldehyde, prior to subjecting the mixture to 5'-3'-exonuclease treatment. Prior to exonuclease treatment, the DNA may be fragmented by sonication. Another source of single-stranded DNA is heat-denatured double-  
25 stranded DNA. For example, heating double-stranded DNA to 95 to 100 °C converts it into ssDNA which can be used as a starting material for the methods according to the present invention which methods make use of the oligonucleotide in accordance with the first aspect. A third possible origin of single-stranded DNA is naturally degraded DNA. Such material is generally found in decaying tissue or bone fragments. This is a starting material  
30 anthropologists and forensic scientists are typically confronted with.

Deviating from state of the art adapters, the oligonucleotide of the present invention is characterized *inter alia* by the following features: (i) The presence of a 5' overhang, i.e. feature (c) of the first aspect. This allows for reverse transcription of the construct formed by ligating  
35 single-stranded nucleic acid to the oligonucleotide of the invention (for further details see below). (ii) While the state of the art adapters typically comprise two separate single strands which associate non-covalently, the adapter of the present invention is one single-stranded

oligonucleotide (which contains a double-stranded portion as defined above). As a consequence, the oligonucleotide of the invention assumes the structure of a stem loop. The loop portion of the stem loop contains nucleic acid elements as well as (iii) a non-nucleic acid spacer. The non-nucleic acid spacer causes polymerases to cease. As a consequence, in the protocol making use of the oligonucleotide as disclosed further below, there is no requirement for linearization prior to amplification by PCR.

As compared to the cumbersome protocols known from the prior art for the purpose of elucidating protein single-stranded nucleic acid interactions, the oligonucleotide of the present invention enables a robust, quick and high-throughput capable procedure. This procedure is the subject of further aspects of the present invention which are disclosed below.

Turning to the optional features of the oligonucleotide in accordance with the first aspect, it is of note that for certain applications multiplexing is desirable. Multiplexing means that nucleic acid, in particular single-stranded nucleic acid, originating from a plurality of sources is pooled. The pooled material is then subjected to downstream processing, downstream processing typically including sequencing. This is a means of further increasing throughput. On the other hand, the need arises to determine, upon sequencing, from which particular source a sequence of interest was originally derived from. This is done by using barcode sequences, wherein each source is given a different barcode. Barcode sequences are short oligonucleotide sequences having a length of preferably 5, 6, 7, 8, 9 or 10 nucleotides. The barcode sequences, if present, typically occur at a known position with the oligonucleotide.

The oligonucleotide according to the first aspect may furthermore comprise, within one, more or all of (a), (b) and (c), one or more optional features which are disclosed above and will be explained further below. These optional features do not entail the addition of nucleotides or sequences. Rather, they define how one, more or all of (a), (b) and (c) may be implemented.

The above phrase "one, more or all of (a), (b) and (c)" indicates that, for example, one or more modified nucleotides may be present only within (a), only within (b), only within (c), within (a) and (b), within (a) and (c), within (b) and (c), or within all of (a), (b) and (c). The same applies *mutatis mutandis* for the compatibility conferring sequences in accordance with (db). For example, and this will become apparent also in view of preferred or exemplary embodiments disclosed below, the sequence of (a) plus (ba) may comprise or consist of a sequence which is complementary to an art-established primer. Similarly, the sequence of (a) plus (bb) may comprise or consist of a sequence which is complementary to another art-established primer.



A further optional feature is the presence of one or more sequences in said oligonucleotide which one or more sequences confer compatibility to nucleic acid sequencing kits. Nucleic acid sequencing kits are well-established on the market and can be obtained from a number of manufacturers such as Illumina (San Diego, US) and Life Technologies (Carlsbad (CA), US; product name: IonTorrent). More specifically, the particular sequences chosen to implement the requirements laid down in the first aspect may be designed in a way that commercially available primers, e.g. from Illumina, recognize and bind to them. Such primers in turn may contain further features, in particular further sequences which are not comprised in the oligonucleotide according to the invention nor have a counterpart in the oligonucleotide according to the invention, which further sequences provide for binding to the oligonucleotide lawn present on the surface, e.g., of Illumina flow cells.

The same applies *mutatis mutandis* for compatibility with the IonTorrent system. Also this can be done by choosing the sequences in accordance with the above items (a), (ba) and (bc) appropriately.

To explain the feature of compatibility with sequencing kits further, we direct the attention to the exemplary constructs displayed in the figures enclosed herewith. In particular, the sequences of (a) plus (ba) of the construct shown in the figures together comprise a sequence which is complementary to a primer sequence designated "TruSeq Read 1" of the Illumina system. This is the standard primer to sequence the 5' ends of libraries. However, this is not mandatory for sequencing or cluster generation in Illumina flow cells. One is in principle free to change (a) and (ba); this would simply entail the requirement to use a different primer for subsequent sequencing. Similarly, the sequences of (a) plus (bb) of the construct shown in the figures together comprise a sequence which is complementary to another standard Illumina sequencing primer ("Multiplex Read 2"). This primer is commonly used to sequences 3' end of libraries. "Complementary" in this context preferably means 100% complementarity, but allows for 1, 2 or 3 mismatches.

Compatibility with the commercially available sequencing kits does not entail a requirement to adapt the overhang in accordance with (c). This overhang is a special feature in accordance with the present invention which renders the oligonucleotide reverse-transcription capable.

A further optional feature are one or more, preferably 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10, random bases. Preferably, said random bases are located in the 5' overhang (c). The term "random bases" designates a base with undetermined chemical nature. More specifically, a random

base may be any one of A, C, G or T/U. A random base may occur within a ribonucleotide or a deoxyribonucleotide.

5 Random bases may be used for quality control, more specifically to differentiate between true PCR duplicates and sequence reads that seem to be PCR duplicates but actually are derived from independent ligation events.

A preferred position of a random base is the 5' end of the oligonucleotide according to the invention.

10 In a preferred embodiment one, more or all of said random bases may be semi-random bases. A semi-random base is a base which may be chosen from less than four of the naturally occurring bases in either DNA or RNA. Preferred semi-random bases are pyrimidines (Y) and purines (R). Semi-random bases Y are accordingly to be chosen from C and T/U, whereas  
15 semi-random bases R are to be chosen from A and G.

The alkaline conditions in accordance with (ca) preferably involve a pH above 12.

20 An exemplary oligonucleotide according to the invention is shown in part (A) of Figure 2. A schematic drawing illustrating key structural features is shown in Figure 1. As noted above, the single-stranded overhang (c) may be DNA or RNA. The two alternatives are shown in Figure 1. The nucleotide at the 3' end of the overhang which is the nucleotide of said overhang which is directly adjacent to the double-stranded portion (a) is a ribonucleotide in either case. The 5' nucleotide of the overhang (which is also the 5' end of the entire oligonucleotide) is  
25 phosphorylated. Positions 3 to 8 is an exemplary barcode sequence. The non-nucleic acid spacer in accordance with the present invention is shown as a curvy line in Figure 1 as well as in part (B) of Figure 2. In part (A) of Figure 2, it is indicated as "~18C~".

30 In a preferred embodiment, (a) the melting temperature of said double-stranded portion is between 37°C and 70°C, preferably between 60°C and 65°C; and/or (b) the number of mismatches between said first strand and said second strand is 0, 1, 2, or 3.

In a further preferred embodiment, said non-nucleic acid spacer is an oligo-ethyleneglycol spacer, preferably  $-(\text{CH}_2\text{CH}_2\text{O})_n-$ , n being an integer number between 3 and 10, preferably 6.  
35 In other words, the phosphate connecting two adjacent nucleotides is preferably replaced by  $-\text{O}-(\text{CH}_2\text{CH}_2\text{O})_n-\text{OPO}_3-$ .

In an alternative preferred embodiment, said non-nucleic acid spacer is an oligo-methylene spacer, preferably  $-(CH_2)_m-$ ,  $m$  being an integer number between 5 and 50, 18 being a preferred value of  $m$ .

5 Generally speaking, there is a host of structural implementations at the skilled person's disposal of a non-nucleic acid spacer which meets the requirements (i) and (ii) as specified in accordance with the first aspect. Examples include abasic nucleotides, such as 1',2'-dideoxyribose (also called dBase); triethylene glycol spacer such as Cholesterol TEG; photocleavable linkers; unnatural bases; natural bases with modifications that block  
10 polymerases such as thymidine glycol and cys-syn thymine dimers; left-handed versions of nucleotides; and thiol SS Serinol Dipod. Corresponding products are available from various manufacturers including Gene Link (Hawthorne, NY, USA). All these structural implementations of the non-nucleic acid linker in accordance with the present invention provide for sufficient flexibility that they do not interfere with the formation of a stem loop of said  
15 oligonucleotide and furthermore cause any polymerase to cease. Appropriate lengths and/or number of groups (such as the number of abasic nucleotides etc.) can easily be determined.

In a further preferred embodiment, said nucleotide of said overhang which is directly adjacent to said double-stranded portion comprises ribose.

20 In a further preferred embodiment, said overhang is RNA. To the extent modified nucleotides are comprised in said overhang, it is understood that it is also preferred that the backbone chemical bonds connecting any of said modified nucleotides with both upstream and downstream parts within said overhang are amenable to cleavage under those conditions  
25 where RNA is amenable to cleavage. Preferred cleavage conditions are disclosed herein below.

In the alternative, the overhang may be DNA or may contain both ribonucleotides and deoxyribonucleotides. Moreover, and as stated above, the option of modified nucleotides  
30 applies for the entire molecule. In any case, also when the overhang is made of DNA, feature (ca) of the first aspect in its broadest definition has to be met, wherein choosing said nucleotide of said overhang which is directly adjacent to said double-stranded portion to be a ribonucleotide is one option to implement feature (ca).

35 In a further preferred embodiment, (a) the 5' end of said oligonucleotide is amenable to ligation and/or comprises a 5' phosphate, a 5' hydroxy group, a 5' tri-phosphate or a 5' pre-adenylated phosphate; and/or (b) the 3' end of said oligonucleotide is capable of functioning as a primer

and/or comprises a 3' hydroxy or a 3' phosphate.

5 A pre-adenylated 5' end is a 5' end where a riboadenosine diphosphate is bound via a phosphate ester to the 5' hydroxy group of the 5'-terminal nucleotide. Pre-adenylation can be performed, for example, with a commercially available kit such as the kit from New England Biolabs (Ipswich, MA, USA). Alternatively, T4 RNA ligase 1 can be used for pre-adenylation (for example as described Song et al., Scientific Reports 5, Article No.: 15620 (2015)).

10 In a particularly preferred embodiment, said 5' end of said oligonucleotide is 5' pre-adenylated, said 3' end carries a 3' phosphate, and said overhang is RNA. In conjunction with this design of the oligonucleotide, it is preferred to remove the 3' phosphate in accordance with step (ac) as described further below.

15 The above disclosed combination of features is one possible way of completely or almost completely avoiding the formation of adapter-dimers. The formation of adapter-dimers is undesirable as instead of yielding the desired cloning product, adapters dimerize and are no longer available for the desired reaction. There is significant concern in the prior art about the formation of adapter-dimers. Art-established methods are cumbersome in that they typically require separate purification steps for the purpose of avoiding adapter-dimer formation.

20

The 3' end of the oligonucleotide is the starting point for reverse transcription.

In a further preferred embodiment, said one or more modified nucleotides, to the extent they are present, include LNA.

25

30 The term "LNA" designates locked nucleic acid. Locked nucleic acid as well as its building blocks, locked nucleotides, are known in the art. A methylene bridge connects 2' O and 4' C of the ribose. Locked nucleotides are a means of stabilizing the helical confirmation of double-stranded portions. Accordingly, locked nucleotides can be used to shorten double-stranded portions while maintaining a given melting temperature.

In a second aspect, the present invention provides the use of the oligonucleotide as defined in any one of the preceding aspects and embodiments for the generation of a library, preferably a sequencing library, from single-stranded nucleic acid.

35

In a preferred embodiment of said use, said generation involves reverse transcription and amplification. Preferred means and methods for implementing reverse transcription as well as amplification are disclosed further below.

- 5 In a further preferred embodiment of said use, said oligonucleotide and primers for said amplification are the only oligonucleotides to be used.

As noted above, a salient feature of the oligonucleotide according to the invention is that it contains sites for reverse transcription as well as binding sites for primers, said primers to be  
10 used for amplification. As a consequence, the oligonucleotide according to the invention is the only adapter molecule required for generating libraries from single-stranded nucleic acid.

In a third aspect, the present invention provides a method of producing a library, preferably a sequencing library, from single-stranded RNA and/or single-stranded DNA, said method  
15 comprising or consisting of: (a) ligating said single-stranded RNA and/or single-stranded DNA to the oligonucleotide according to the first aspect; (b) reverse transcribing or transcribing, respectively, said single-stranded RNA or single-stranded DNA; (c) hydrolyzing any RNA; (d) circularizing the remaining DNA; and (e) subjecting the product of (d) to amplification by PCR.

20 This method yields a library which is amplified and ready for sequencing. In principle, sequencing can be effected by any means. To the extent the above described optional compatibility feature is used to implement complementarity with primers of a certain sequencing kit, it is obviously preferred to perform the downstream sequencing step with that particular kit.

25 Ligating in accordance with (a) can be effected with any ligase. Preferred is the use of T4 RNA ligase 1 in which case preference would be given to 5' phosphorylated oligos. Alternatively, it is possible to use T4 RNA ligase 2 (truncated form, preferably the KQ mutant) in which case preference is given to a pre-adenylated oligo. Pre-adenylated oligos are also compatible with  
30 T4 RNA ligase 1. Finally, RtcB ligase may be used, preferably in conjunction with a 5' OH oligonucleotide.

During ligation, said single-stranded RNA or single-stranded DNA is preferably bound to a solid support such as beads.

35

Considering the preferred upstream pre-processing as detailed below, a solid support such as beads may e.g. be introduced in the course of step (2) of said pre-processing which step is purifying of a given protein. In the course of said purifying, use may be made of beads.

5 A solid support such as beads may also be introduced into the procedure at the ligation step. The use of a solid support provides a specific implementation which permits to entirely avoid or substantially entirely avoid the formation of adapter-dimers. As such, in addition to the solution reaction avoiding adapter-dimer formation as described above, the present invention provides a second method which allows for the avoidance of adapter-dimer formation which second  
10 method employs a solid support. Further details in relation thereto are described further below in relation to the step (aa) of removing excess oligonucleotide after step (a) ligating.

After ligating, one may conveniently proceed to reverse transcribing or transcribing, respectively, in accordance with (b). There is no need to add and anneal an oligonucleotide to  
15 prime a reverse transcription. This advantageous property is inherent to the oligonucleotide according to the first aspect of the present invention. Use can be made of any polymerase having reverse transcriptase activity. The polymerase may comprise further activities, but this is not required. Additional activities which may be present on certain polymerases in addition to reverse transcriptase activity include helicase activity and strand displacement activity. An  
20 example of a polymerase having reverse transcriptase activity and strand displacement activity is Bst DNA polymerase. If a library is to be produced from single-stranded DNA as starting material, preference is given to polymerases which have reverse transcriptase activity and strand displacement activity. A commercially available Bst polymerase product is Bst 3.0 from New England Biolabs (Ipswich, MA, USA).

25 For the purpose of hydrolyzing in accordance with (c), any means which does not destroy the non-RNA part of the nucleic acid molecule generated in (b) may be used. Preferred means are discussed further below.

30 Circularizing in accordance with (d) is preferably done with CircLigase or CircLigase II.

Finally, amplification in accordance with (e) can be effected without any further purification or linearization. This is made possible by the non-nucleic acid spacer which causes the polymerase to cease.

35 In a preferred embodiment of the method according to the third aspect, said method comprises one or more of the following further steps: (aa) removing excess oligonucleotide after step (a)

and prior to step (b); (ab) pooling of different samples containing ligation products obtained in (a), wherein ligation products originating from different samples are barcoded differently in accordance with feature (c)(cb) of the above disclosed first aspect, and wherein said step (ab) to be effected after step (a) and prior to step (b); (ac) removing, after step (a) and prior to step  
5 (b), the 3' phosphate from said ligation products to the extent it is present; and (ca) removing protein and small molecules from single-stranded DNA obtained in (c), step (ca) to be effected after step (c) and prior to step (d).

A preferred order of steps (aa), (ab) and (ac), in particular to the extent two or all three of them  
10 are to be performed, is the following temporal order: (aa), followed by (ab), and then (ac).

To the extent single stranded RNA or single stranded DNA is bound to a solid support such as beads (see above), removing excess oligonucleotide after step (a) in accordance with step (aa) is especially preferred and is preferably done by washing. As a consequence, after  
15 ligating in accordance with step (a) and washing in accordance with step (aa), every ligated target has a single reverse transcription primer which is already covalently linked thereto. By adopting this procedure, surprisingly the vast majority of adapter-dimer formation is eliminated. Importantly, no separate purification step such as purification by gel electrophoresis is necessary. This greatly increases efficiency and scalability of the protocol.

20

Step (ab) provides for the above mentioned multiplexing.

Presence of a 3' phosphate on the oligonucleotide of the invention is preferred in order to prevent self-ligation thereof. Self-ligation would lead to incomplete ligation of the single-  
25 stranded RNA or DNA which is undesirable. After ligation in step (a), the 3' phosphate is removed in accordance with (ac).

Presence of a 3' on the oligonucleotide of the invention during ligation and subsequent step (ac) are also especially useful if the avoidance of adapter-dimer formation is to be effected in  
30 solution (as opposed to the use of a solid support). The use and removal of the 3' phosphate in accordance with step (ac) is especially preferred in conjunction with the specific oligonucleotide design which is disclosed herein above. According to this specific design, the 5' end of said oligonucleotide is 5' pre-adenylated, its 3' end carries a 3' phosphate, and the overhang is RNA. In that case, step (aa) is not necessary and preferably not performed and  
35 step (ac) is preferably performed in the mixture directly obtained in step (a).

Step (ac) is generally implemented by adding phosphatase.

In principle, any phosphatase can be used for the purpose of 3' phosphate removal. T4 phosphonucleotide kinase (T4 PNK), preferably at a pH of about 6, can be used for that purpose. Any other phosphatase such as CIP, SAP, Antarctic phosphatase, FastAP, under the conditions specified by manufacturers, may be used in the alternative.

5

In a preferred embodiment, removal of the 3' phosphate in accordance with step (ac) is done by adding phosphatase to the ligation mixture formed in step (a). Preferably, a defined amount of time is allowed to lapse after start of the ligation procedure (a) until phosphatase in accordance with step (ac) is added. Preferred amounts of time are between about 10 minutes and about 2 hours, more preferably between about 30 and about 90 minutes, most preferred about 60 minutes.

10

As an alternative to adding phosphatase to the ligation mixture formed in step (a), but less preferred, a transfer step from one vessel to another vessel and/or a purification step may occur between ligation in accordance with (a) and adding of phosphatase in accordance with (ac).

15

In a further preferred embodiment of the method of the invention, said hydrolyzing any RNA is effected, (c1) if the overhang of said oligonucleotide is RNA, by adding RNaseH; or (c2) if said overhang contains DNA, by adding 0.1M NaOH.

20

Routinely, incubation with RNaseH is done at 37°C for 10 to 60 minutes.

For the removal of protein and small molecules in accordance with (ca), use can be made of the art-established means and methods including phenol/chlorophorm extraction followed by ethanol precipitation, silica or glass fibre columns suitable for DNA purification, or Solid Phase Reversible Immobilization (SPRI) beads.

25

Treatment with 0.1M NaOH in accordance with (c2) is preferably done at a temperature between 37°C and 70°C, preferably at 70°C, for about 10 minutes. This will yield a hydroxy group at the 5' end of the remaining DNA molecule. This may be converted into a phosphate by using T4 PNK at neutral or slightly basic pH (pH 7 to 8) in the presence of ATP.

30

The method of generating a library as described so far may optionally be combined with an upstream pre-processing. This is the subject of the following preferred embodiment.

35



Accordingly, in a further preferred embodiment of the method in accordance with the third aspect, said method comprises the following steps prior to step (a): (1) optionally cross-linking a sample comprising cells; (2) purifying from said sample a given protein, said protein being bound or cross-linked to single-stranded or double-stranded nucleic acid; (3) converting said nucleic acid, if it is double-stranded, into single-stranded nucleic acid; and (4) trimming of the single-stranded nucleic acid bound or cross-linked to said protein.

Converting double-stranded DNA into single-stranded DNA may be effected with lambda exonuclease.

Trimming in accordance with step (4) is preferably done with RNAses. Useful enzymes include RNase I, RNase A, RNase V1, RNase T1, MNase and benzonase.

Given that RNase enzymes may leave a 2',3'-cyclic phosphate, a 3' phosphate or a 2' phosphate at the 3' end, further modification will be needed in order to generate an end which is suitable for ligation. Any phosphatase such as CIP, SAP, antarctic phosphatase, FastAP and T4 PNK may be used. Preference is given to T4 PNK at a pH of about 6.

In a particularly preferred embodiment of the above disclosed preferred embodiment, (1) cross-linking is effected by (1.1) UV-C; (1.2) UV-A after feeding said cells 4-thiouridine; or (1.3) formaldehyde; (2) purifying is effected by using (2.1) tandem affinity purification, in which case cross-linking is compulsory; (2.2) an antibody specific for said protein; and/or (3) trimming of said single-stranded nucleic acid, to the extent it is RNA, is effected with an RNase and/or by sonication.

As regards purifying, preference is given to tandem affinity purification. In that respect it is preferred to exploit the biotin/streptavidin interaction in one of the two steps. Avidin, NeutrAvidin or any suitable streptavidin derivative may be used as well.

In a further particularly preferred embodiment, said method comprises the following further steps after step (e): (f) sequencing one, more or all members of the library obtained in step (e); and (g) determining the most 5' nucleotide and/or the most 3' nucleotide of said nucleic acid bound or cross-linked to said protein, thereby determining the binding site of said protein on said nucleic acid.

Assuming that the site where the polymerase ceases is to be interpreted as the binding site of the RNA binding protein, this embodiment allows to pinpoint protein/single-stranded nucleic

acid binding sites. Generally, it is the most 5' nucleotide on the positive strand and/or the most 3' nucleotide on the negative strand which allow determining of the binding site.

5 In a fourth aspect, the present invention provides a kit comprising (a) an oligonucleotide in accordance with the first aspect; (b) one or more enzymes selected from a ligase, a reverse transcriptase, a CircLigase, and a DNA polymerase; and (c) optionally a manual containing instructions for performing the method in accordance with the third aspect.

10 As regards the embodiments characterized in this specification, in particular in the claims, it is intended that each embodiment mentioned in a dependent claim is combined with each embodiment of each claim (independent or dependent) said dependent claim depends from. For example, in case of an independent claim 1 reciting 3 alternatives A, B and C, a dependent claim 2 reciting 3 alternatives D, E and F and a claim 3 depending from claims 1 and 2 and reciting 3 alternatives G, H and I, it is to be understood that the specification unambiguously  
15 discloses embodiments corresponding to combinations A, D, G; A, D, H; A, D, I; A, E, G; A, E, H; A, E, I; A, F, G; A, F, H; A, F, I; B, D, G; B, D, H; B, D, I; B, E, G; B, E, H; B, E, I; B, F, G; B, F, H; B, F, I; C, D, G; C, D, H; C, D, I; C, E, G; C, E, H; C, E, I; C, F, G; C, F, H; C, F, I, unless specifically mentioned otherwise.

20 Similarly, and also in those cases where independent and/or dependent claims do not recite alternatives, it is understood that if dependent claims refer back to a plurality of preceding claims, any combination of subject-matter covered thereby is considered to be explicitly disclosed. For example, in case of an independent claim 1, a dependent claim 2 referring back to claim 1, and a dependent claim 3 referring back to both claims 2 and 1, it follows that the  
25 combination of the subject-matter of claims 3 and 1 is clearly and unambiguously disclosed as is the combination of the subject-matter of claims 3, 2 and 1. In case a further dependent claim 4 is present which refers to any one of claims 1 to 3, it follows that the combination of the subject-matter of claims 4 and 1, of claims 4, 2 and 1, of claims 4, 3 and 1, as well as of claims 4, 3, 2 and 1 is clearly and unambiguously disclosed.

30

The Figures show:

**Figure 1:** Schematic representation of an exemplary oligonucleotide of the invention.

5 **Figure 2:** Implementation with Illumina-compatible sequences. (A) Exemplary oligonucleotide. (B) Illustration of workflow.

10 **Figure 3:** Example of RNA-binding profiles after Illumina sequencing. IGV (Integrative Genomics Viewer; distributed by Broad Institute) screenshots showing the distribution of cross-linking events on three example genes. KHDRBS2\_old: uvCLAP prior art method, does not use the oligonucleotide of the invention profile of KHDRBS2 protein. KHDRBS2\_New\_RNA: Method of the invention that uses an all-RNA oligonucleotide and RNaseH mediated phosphorylation of the 5'-end of the cDNA. KHDRBS2\_New\_DNA: Method of the invention that uses a  
15 single RNA moiety in the 5' overhang that is cleaved with NaOH after which the 5'-end of the cDNA is phosphorylated with T4 PNK. 5'-phosphorylation is necessary for the circularization step.

20 **Figure 4:** Avoidance of adapter-dimer formation. Workflow.  
Target: T  
pre-adenyl group: App  
ssRNA extension of the stem-loop oligonucleotide: R  
Top part of the helix that follows R: A  
Top part of ssDNA that follows the helix: B  
25 C18 spacer: S  
Bottom part of the ssDNA that follows S: C  
Bottom part of the helix that follows C: D  
5'-phosphate: 5p  
3'-phosphate: 3p  
30 3'-hydroxyl: 3-OH

**Figure 5:** Avoidance of adapter-dimer formation. Evidence. Three sequencing libraries were prepared from enzymatically fragment poly(A+) RNA using a phosphorylated adapter (lane 1) and two pre-adenylated adapters (lanes 2 and 3). In lane 2 the phosphorylated adapter was pre-adenylated with *Mth* RNA ligase, in lane 3 pre-adenylation was carried out with T4 RNA Ligase 1. The ligation reactions were carried out with T4 RNA Ligase 2 KQ in the absence of  
35

ATP. In lane 1, only adapter-dimers could be amplified, whereas in lanes 2 and 3 adapter-dimer formation was severely reduced and proper sequencing libraries could be generated.

5

The Examples illustrate the invention.

### Example 1

#### 10 Exemplary protocol

The protocol given below uses an oligonucleotide of the invention, e.g. the oligonucleotide displayed in the Figures.

- 15 (1) Protein of interest is enriched with a tandem-affinity purification.
- (2) Bound RNA is trimmed with an RNase enzyme. RNaseI is used which digests all bonds and leaves a 2',3'-cyclic phosphate group at the 3'-end of RNAs.
- 20 (3) Certain RNase enzymes may leave a 2',3'-cyclic phosphate, a 3'-phosphate or a 2'-phosphate group at the 3'-end. This type of end is not suitable for ligation, and must be repaired. T4 PNK at low pH (6.0) may be used for this purpose.
- (4) The single-oligo is then ligated to the trimmed and repaired RNA, e.g. on beads. T4  
25 RNA Ligase 1 and 5'-phosphorylated oligos are used.
- (5) Excess single-oligo is washed away. At this stage, if desired, differentially labeled samples through the internal index part (barcode) in the single-oligo can be mixed together, before reverse transcription. Depending on how the indices are designed,  
30 between two to hundreds of samples can be mixed together.
- (6) If the single-oligo contains a 3'-phosphate group which inhibits self-ligation, it is removed before reverse transcription. T4 PNK at low pH (6.0) is used.
- 35 (7) Reverse transcriptase is added together with necessary reagents such as dNTPs etc. There is no need to add and anneal an oligo to prime the reverse transcription.

(8) After reverse transcription two things are done:

- a. Remove all RNA
- b. Provide single-stranded DNA that has a 5'-phosphate and 3'-hydroxyl  
(necessary for CircLigase reaction)

There are two ways to achieve that:

(9) a) If the single-stranded extension of the single-oligo is made entirely of RNA, then RNaseH is added to the sample and incubated at 37°C for 10-60 minutes. RNaseH is a processive RNase that degrades RNA in RNA/DNA hybrids, and leaves a 5'-phosphate at the 5'-end.

b) Alternatively, the sample is treated with 0.1N NaOH (pH >13) at 70°C for 10 minutes first (or similar conditions that break down RNA but leave DNA intact). This destroys all RNA in the sample and denatures dsDNA to give rise to ssDNA. However, the 5'-end of the DNA molecule is a hydroxyl. This group is then converted to a phosphate using T4 PNK at neutral or slightly basic pH (7-8) in the presence of ATP. The second way can be used on all types of extensions e.g. all-RNA, single-RNA, or anything in between as long as the first nucleotide just after the double-stranded portion is RNA.

(10) The reaction is cleaned-up with any method that removes small molecules and protein while retaining ssDNA. SPRI (AMPure or SILANE) beads or columns, or phenol-chloroform extraction followed by ethanol precipitation may be used.

(11) The purified, 5'-phosphate, 3'-hydroxyl ssDNA is circularized with CircLigase or CircLigaseII.

(12) Circularized DNA – without further purification or linearization (the non-nucleic acid spacer makes this possible) – is then used directly as a template in a PCR reaction to generate the final sequencing library.

## Results

The determination of a binding site of an RNA-binding protein on its cognate mRNA is illustrated in Figure 3. 5' terminal nucleotides of the sequencing reads of the positive strand are

indicative of a binding site. Similarly, 3' terminal nucleotides of the sequencing reads of the negative strand are indicative of a binding site. The height of these peaks in this figure corresponds to the number of sequencing reads which terminate at the given position.

## 5 Example 2

### Avoidance of adapter-dimer formation in solution

This is a preferred process which is particularly suitable for the avoidance of adapter-dimer formation. The procedure starts with the RNA target T and oligonucleotide AppRABSCD3p where A and D base-pair, R is made of ribonucleotides, and A, B, C and D are deoxyribonucleotides. "3p" designates a 3' phosphate.

The workflow is given below and illustrated in Figure 4. This oligonucleotide is an oligonucleotide of the invention.

- (1) T is ligated to the stem-loop oligonucleotide: TRABSCD3p is generated. The 3p group prevents ligation of the AppR part of the oligonucleotide to D, which would be the preferred ligation event because of the proximity of these groups. When the 3'-end is unprotected (that is, when it is a 3'-OH group instead) this creates a circle from the stem-loop oligonucleotide and reduces ligation efficiency by depleting the available amount of stem-loop oligonucleotide. Preferably, at this stage, there will be excess stem-loop oligonucleotide in the solution in order to be able to capture as much target as possible.
- (2) Reverse transcribe the ligated target. The 3p group, however, does not permit this since it is a polymerization inhibitor as well as a ligation inhibitor. Phosphatase is added directly into the ligation mixture. Two reactions happen at this stage:
  - a) TRABSCD3p is converted to TRABSCD3-OH, which can undergo reverse transcription
  - b) The excess oligonucleotide, AppRABSCD3p goes through a very unexpected transformation: first it becomes AppRABSCD3-OH, but later, since the dephosphorylation is carried out in the ligation reaction mixture, it is circularized with very high efficiency and becomes RABSCD (imagine a line connecting R and D). This essentially chemically inactivates and neutralizes the excess stem-

loop oligonucleotide, without a requirement for gel electrophoresis or any other type of purification.

- 5 (3) Ligation and reverse transcription buffers preferably are very similar in their composition. For reverse transcription, the phosphatase can be inactivated by heat (e.g. 65°C for 5 minutes is recommended for the phosphatase rSAP), dNTP and the reverse transcriptase is added in the tube, and reverse transcription is allowed to take place, preferably at 40 - 60°C and depending on the enzyme. No primer is added, as described above.
- 10 a) TRABSCD3-OH then becomes TRABSCDR'T'3-OH where R' and T' are the reverse complements of R and T respectively, are made of deoxribonucleotides and base pair perfectly with R and T forming RNA/DNA hybrids.
- 15 b) Nothing happens to the circularized, excess stem-loop oligonucleotide, since it does not have a free 3'-OH group for primer extension. It stays as RABSCD and importantly contains no RNA/DNA hybrids.
- 20 (4) The same single-tube where ligation was carried out is still being used. At the end of reverse transcription, RNaseH is added into the same tube, and the samples are incubated, preferably at 37°C for 20-60 minutes.
- 25 a) RNaseH is not used to get rid of RNA. It is used to specifically label the 5'-end of TRABSCDR'T'3-OH with a phosphate group. Now TRABSCDR'T'3-OH becomes 5pABSCDR'T'3-OH. In the process T and R are degraded.
- b) The circularized, excess stem-loop oligonucleotide again stays as RABSCD and does not contain a 5'-phosphate group.
- 30 (5) To remove all RNA and to break the circularized, excess stem-loop oligonucleotide, sodium hydroxide is added to the reaction (to ~0.1N), and the samples are incubated at 70-80°C for 10-30 minutes.
- 35 a) 5pABSCDR'T'3-OH stays as it is, ready for circularization with CircLigase.
- b) RABSCD loses the bond between R and D and becomes 5-OH-RABSCD3p, where 5-OH is a 5'-hydroxyl group and 3p is either a 3'-phosphate or a 2',3'-

cyclic bisphosphate. These are chemically equivalent at this point. This molecule cannot circularize in the next step.

- 5 (6) Reactions are preferably cleaned up with either a column, SPRI beads or any other applicable means. This is the step where a second tube is used. All of the above was a single-tube reaction without purification or oligonucleotide addition except for the initial stem-loop oligo.
- 10 (7) 5pABSCDR'T'3-OH is circularized with CirLigase. It now becomes ABSCDRT, A and T linked with a 3'-5' phosphate bond. Another way to write this is CDRTABS where AB is used as a primer binding site for the "reverse PCR primer" and the reverse-complement of CD (C'D') serves as a binding site for the "forward PCR primer". This generates a library ready for sequencing and identifying the original target T, and is
- 15 100% directional.
- The linearized, circularized, excess stem-loop oligonucleotide 5-OH-RABSCD3p cannot generate a PCR product or participate in any kind of polymerization due to the its blocked 3'-end.
- 20 In the end, virtually no adapter-dimer is generated, even without carrying out any kind of size fractionation, be it SPRI beads or gel-purification. Reference is made to Figure 5.



## Claims

- 5 1. An oligonucleotide comprising
- (a) a double-stranded portion, which double-stranded portion is DNA and 9 to 30  
base pairs in length;
- (b) a loop connecting the 3' end of the first strand of said double-stranded portion  
with the 5' end of the second strand of said double-stranded portion, said loop  
10 comprising, in 5' to 3' direction:
- (ba) a first DNA portion which is 4 to 20 nucleotides in length;
- (bb) a non-nucleic acid spacer which
- (i) does not interfere with the formation of a stem-loop by said  
oligonucleotide; and
- 15 (ii) causes polymerases to cease;
- and
- (bc) a second DNA portion which is 4 to 20 nucleotides in length; wherein  
said first DNA portion and said second DNA portion are not complementary to  
each other;
- 20 (c) a single-stranded overhang at its 5' end, said overhang being 5 to 40  
nucleotides in length, wherein
- (ca) the bond connecting said double-stranded portion with the nucleotide of  
said overhang which is directly adjacent to said double-stranded portion  
is cleavable under alkaline conditions; and
- 25 (cb) said overhang optionally comprises a barcode sequence, said barcode  
sequence preferably being 5 to 10 nucleotides in length;
- and optionally
- (d) within one, more or all of (a), (b) and (c), one, more or all of the following:
- (da) one or more modified nucleotides;
- 30 (db) one or more sequences conferring compatibility with nucleic acid  
sequencing kits, such compatibility being preferably the presence of  
regions within said oligonucleotide which are complementary to primers  
comprised in said sequencing kits; and
- (dc) one or more random bases.
- 35

2. The oligonucleotide of claim 1, wherein
- (a) the melting temperature of said double-stranded portion is between 37°C and 70°C; and/or
- (b) the number of mismatches between said first strand and said second strand is
- 5 0, 1, 2, or 3.
3. The oligonucleotide of claim 1 or 2, wherein said non-nucleic acid spacer is an oligo-ethyleneglycol spacer, preferably  $-(\text{CH}_2\text{CH}_2\text{O})_n-$ , n being an integer number between 3 and 10, preferably 6.
- 10 4. The oligonucleotide of any one of the preceding claims, wherein said nucleotide of said overhang which is directly adjacent to said double-stranded portion comprises ribose.
- 15 5. The oligonucleotide of any one of the preceding claims, wherein said overhang is RNA.
6. The oligonucleotide of claim 5, wherein the 5' end of said oligonucleotide is pre-adenylated and the 3' end of said oligonucleotide comprises a 3' phosphate.
- 20 7. Use of the oligonucleotide as defined in any one of the preceding claims for the generation of a library, preferably a sequencing library, from single-stranded nucleic acid.
- 25 8. The use of claim 7, wherein said generation involves reverse transcription and amplification.
9. The use of claim 7 or 8, wherein said oligonucleotide and primers for said amplification are the only oligonucleotides to be used.
- 30 10. A method of producing a library, preferably a sequencing library, from single-stranded RNA and/or single-stranded DNA, said method comprising or consisting of:
- (a) ligating said single-stranded RNA and/or single-stranded DNA to the oligonucleotide as defined in any one of claims 1 to 6;
- 35 (b) reverse transcribing or transcribing, respectively, said single-stranded RNA or single-stranded DNA;
- (c) hydrolyzing any RNA;

- (d) circularizing the remaining DNA; and
  - (e) subjecting the product of (d) to amplification by PCR.
11. The method of claim 10, comprising one or more of the following further steps:
- 5 (aa) removing excess oligonucleotide after step (a) and prior to step (b);
  - (ab) pooling of different samples containing ligation products obtained in (a), wherein ligation products originating from different samples are barcoded differently in accordance with claim 1(c)(cb), and wherein said step (ab) to be effected after step (a) and prior to step (b);
  - 10 (ac) removing, after step (a) and prior to step (b), the 3' phosphate from said ligation products to the extent it is present; and
  - (ca) removing protein and small molecules from single-stranded DNA obtained in (c), step (ca) to be effected after step (c) and prior to step (d).
12. The method of claim 11, wherein step (ac) is performed while step (aa) is not performed, wherein said oligonucleotide is as defined in claim 6, and wherein preferably step (ac) is performed in the mixture directly obtained in step (a).
13. The method of any one of claims 10 to 12, wherein hydrolyzing any RNA is effected,
- 20 (c1) if the overhang of said oligonucleotide is RNA, by adding RNaseH; or
  - (c2) if said overhang contains DNA, by adding 0.1M NaOH.
14. The method of any one of claims 10 to 13, comprising the following steps prior to step (a):
- 25 (1) optionally cross-linking a sample comprising cells;
  - (2) purifying from said sample a given protein, said protein being bound or cross-linked to single-stranded or double-stranded nucleic acid; and
  - (3) converting said nucleic acid, if it is double-stranded, into single-stranded nucleic acid; and
  - 30 (4) trimming of the single-stranded nucleic acid bound or cross-linked to said protein.
15. The method of claim 14, wherein
- 35 (1) cross-linking is effected by
    - (1.1) UV-C;
    - (1.2) UV-A after feeding said cells 4-thiouridine; or
    - (1.3) formaldehyde;

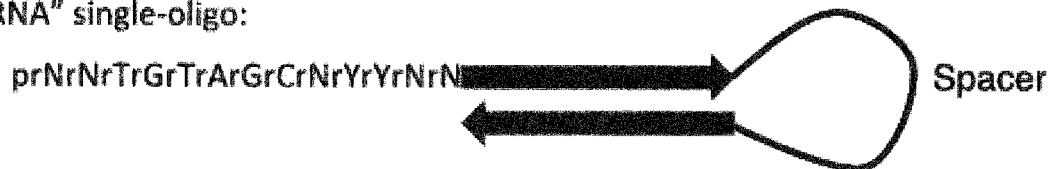
- (2) purifying is effected by using
- (2.1) tandem affinity purification, in which case cross-linking is compulsory;
  - (2.2) an antibody specific for said protein;
- 5 and/or
- (3) trimming of said single-stranded nucleic acid, to the extent it is RNA, is effected with an RNase and/or by sonication.
16. The method of claim 14 or 15, comprising one or more the following further steps
- 10 after step (e):
- (f) sequencing one, more or all members of the library obtained in step (e); and
  - (g) determining the most 5' nucleotide and/or the most 3' nucleotide of said nucleic acid bound or cross-linked to said protein, thereby determining the binding site of said protein on said nucleic acid.
- 15
17. A kit comprising
- (a) an oligonucleotide as defined in any one of claims 1 to 6;
  - (b) one or more enzymes selected from a ligase, a reverse transcriptase, a CircLigase, and a DNA polymerase; and
- 20 (c) optionally a manual containing instructions for performing the method as defined in any one of claims 10 to 16.

Figure 1

"DNA" single-oligo:



"RNA" single-oligo:



rN: ribonucleotide

p: 5'-phosphate (can be adenylated, not depicted)

All other nucleotides are deoxyribonucleotides.


 : self-complementary sequence.  
Might be phosphorylated at the 3'-end (not depicted)

Figure 2  
(A)

N: RNA  
TGTAGC: index

pNNTGTAGCNYYNNAGATCGGAAGAGCGTGT~18C~ACGTGTGCTCTCCGATCT

Figure 2

(B)

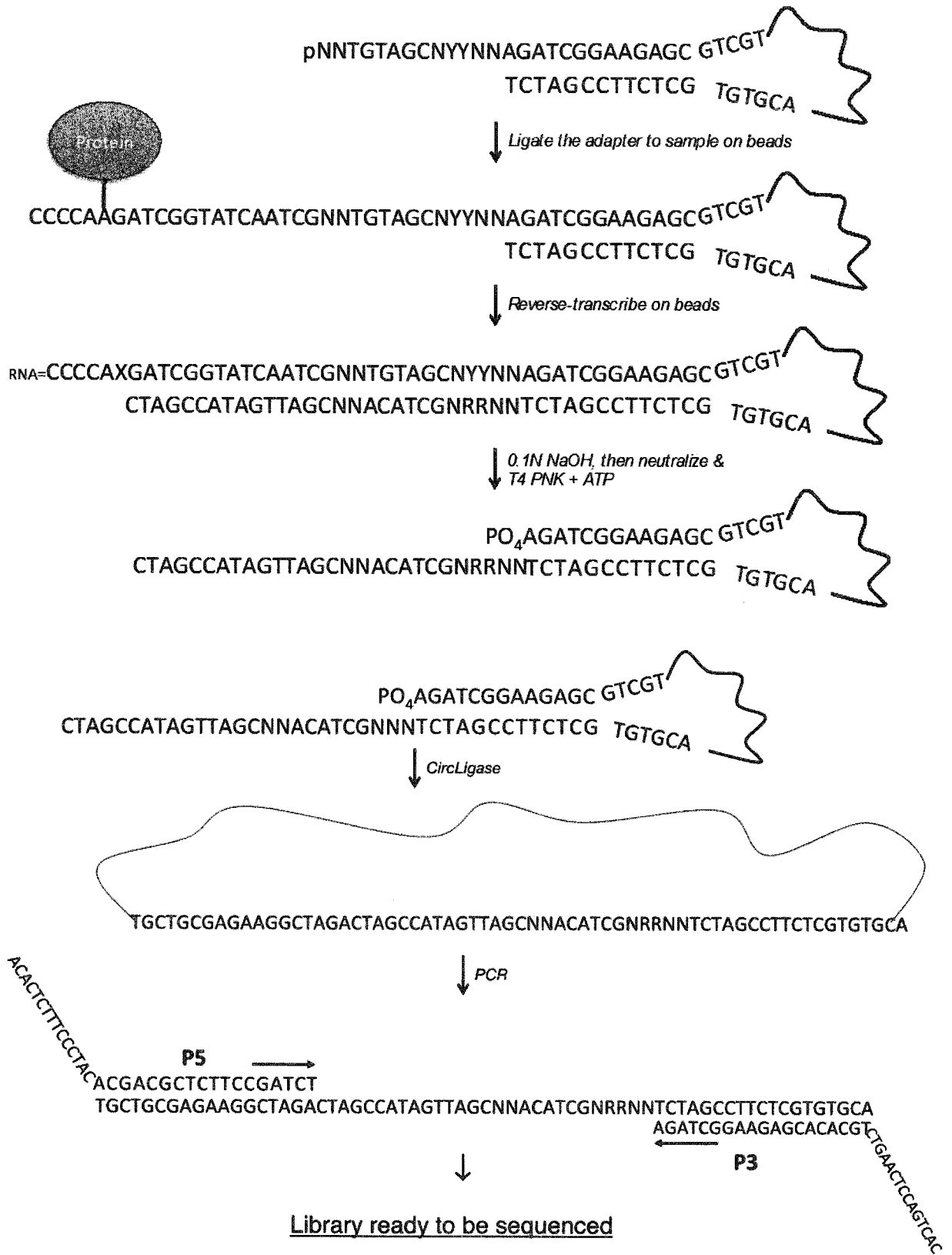


Figure 3

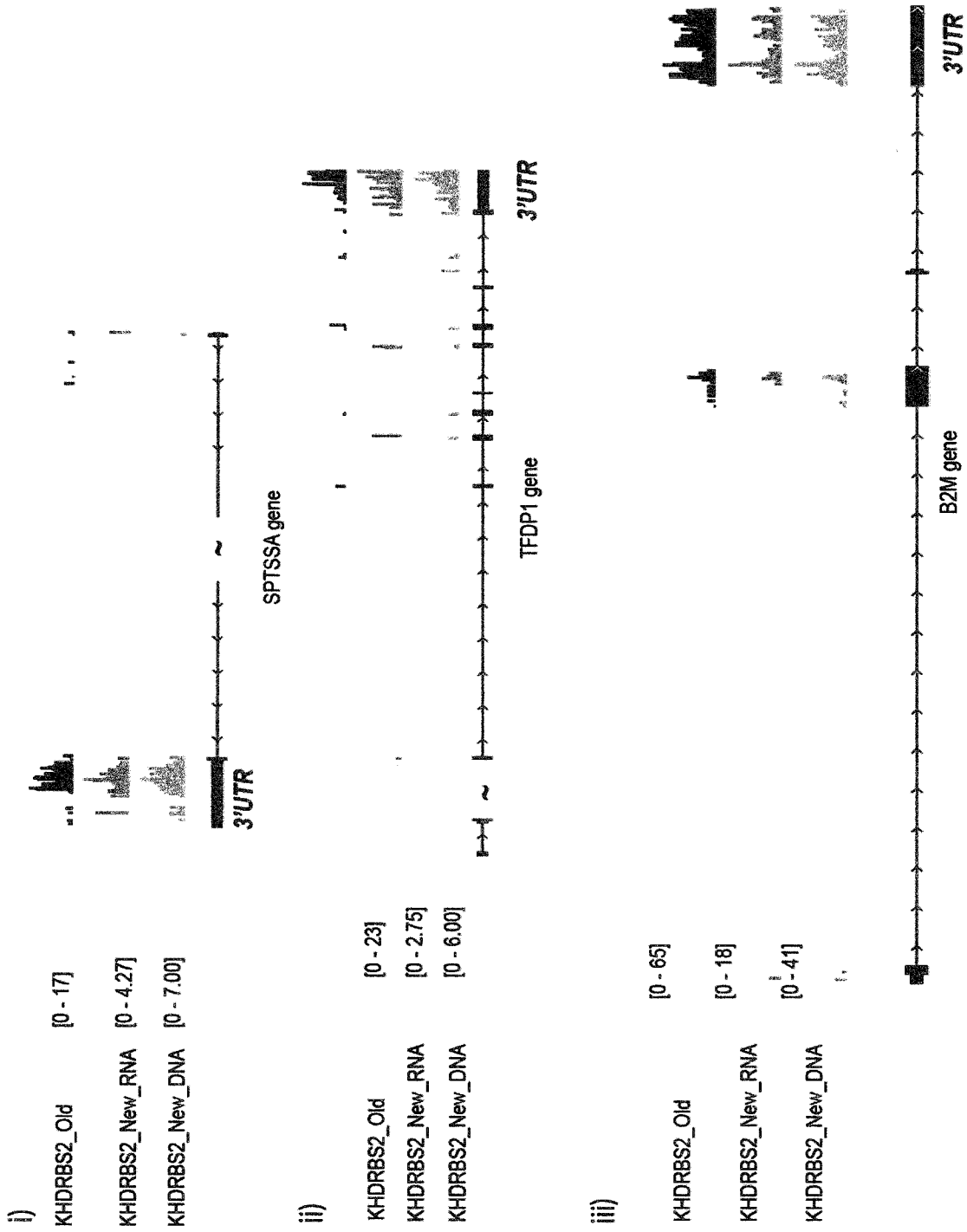




Figure 4

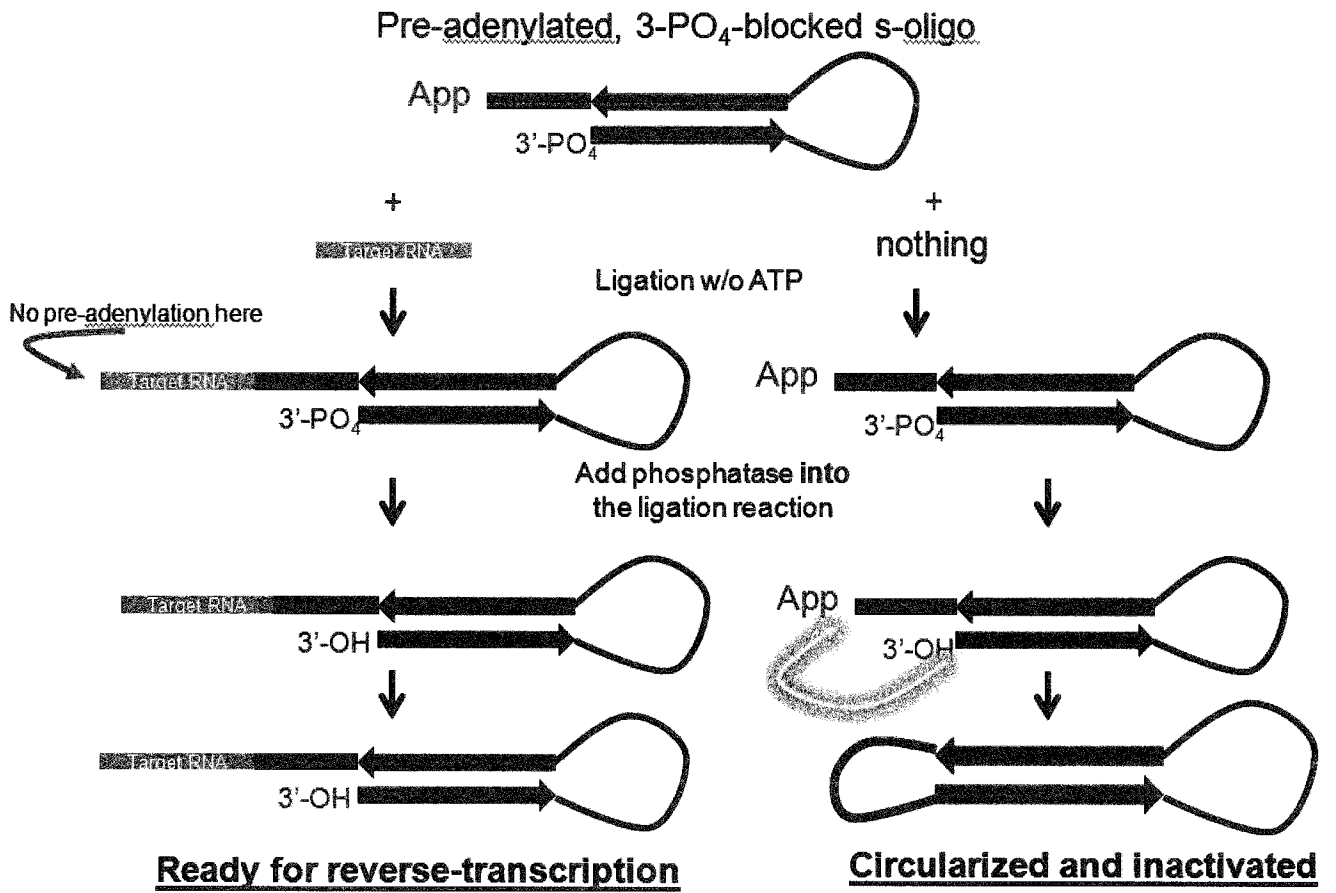
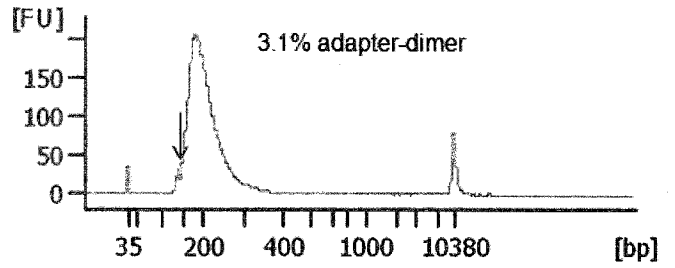
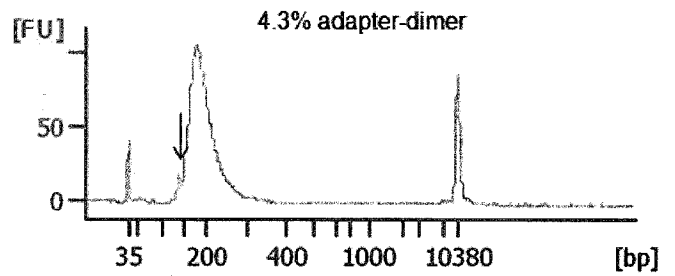
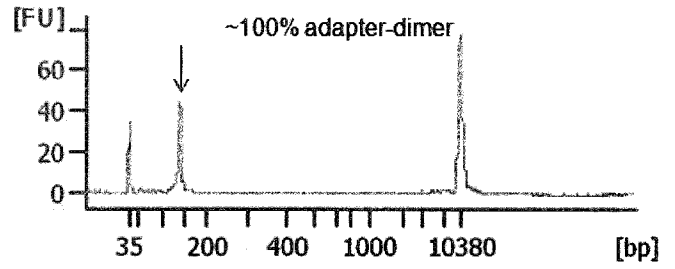
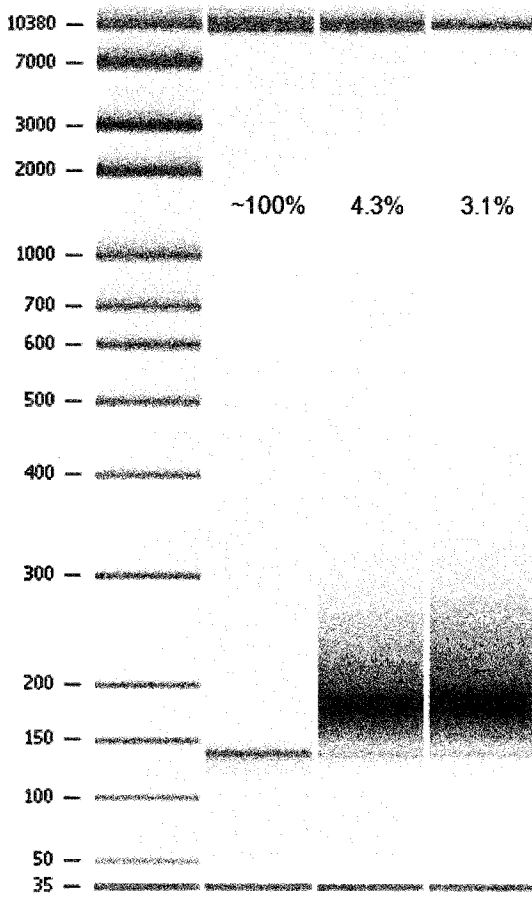


Figure 5



INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2016/066876

A. CLASSIFICATION OF SUBJECT MATTER  
INV. C12N15/10 C12Q1/68  
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)  
C12N 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	WO 2014/071070 A1 (PACIFIC BIOSCIENCES CALIFORNIA [US]) 8 May 2014 (2014-05-08) the whole document -----	1-17
Y	WO 2012/103154 A1 (NUGEN TECHNOLOGIES INC [US]; KURN NURITH [US]; STAPLETON MARK [US]; WA) 2 August 2012 (2012-08-02) cited in the application in particular see figure 1-2 and pages 15, 18-22; the whole document ----- -/--	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  23 September 2016	Date of mailing of the international search report  07/10/2016
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  Vix, Olivier

## INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2016/066876

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2013/112923 A1 (NUGEN TECHNOLOGIES INC [US]; AMORESE DOUG [US]; ARMOUR CHRIS [US]; KUR) 1 August 2013 (2013-08-01) generation of single stranded nucleic acid, see claims 1 and 15-16, pages 5-6 and 8-11, 16-17, 20-21; the whole document	1-17
Y	WO 2011/159256 A1 (UNIV SINGAPORE [SG]; TOO HENG PHON [SG]; ANWAR AZLINDA B [SG]) 22 December 2011 (2011-12-22) in particular see pages 3-5 and claims 1-7; the whole document	1-17
A	METZKER MICHAEL L: "APPLICATIONS OF NEXT-GENERATION SEQUENCING Sequencing technologies - the next generation", NATURE REVIEWS GENETICS, NATURE PUBLISHING GROUP, GB, vol. 11, no. 1, 1 January 2010 (2010-01-01), pages 31-46, XP002659622, ISSN: 1471-0056, DOI: 10.1038/NRG2626 the whole document	1-17
A	US 2012/283145 A1 (WANG YAN [US]) 8 November 2012 (2012-11-08) the whole document	1-17
A	YOICHIRO SUGIMOTO ET AL: "Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions", GENOME BIOLOGY, BIOMED CENTRAL LTD., LONDON, GB, vol. 13, no. 8, 3 August 2012 (2012-08-03), page R67, XP021138398, ISSN: 1465-6906, DOI: 10.1186/GB-2012-13-8-R67 the whole document	1-17
A	WO 2008/043543 A2 (LOOP DIAGNOSTICS GMBH [DE]; THRIPPLETON IAN [DE]) 17 April 2008 (2008-04-17) the whole document	1-17

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2016/066876

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014071070 A1	08-05-2014	US 2014134610 A1 WO 2014071070 A1	15-05-2014 08-05-2014
WO 2012103154 A1	02-08-2012	NONE	
WO 2013112923 A1	01-08-2013	CA 2862552 A1 CN 104093890 A CN 105861487 A EP 2807292 A1 GB 2513793 A GB 2533882 A SG 10201504490Q A SG 11201404243W A US 2013231253 A1 WO 2013112923 A1	01-08-2013 08-10-2014 17-08-2016 03-12-2014 05-11-2014 06-07-2016 30-07-2015 28-08-2014 05-09-2013 01-08-2013
WO 2011159256 A1	22-12-2011	CN 103210092 A EP 2580355 A1 HK 1187378 A1 JP 5851496 B2 JP 2013530698 A SG 185776 A1 US 2013177915 A1 WO 2011159256 A1	17-07-2013 17-04-2013 26-08-2016 03-02-2016 01-08-2013 30-01-2013 11-07-2013 22-12-2011
US 2012283145 A1	08-11-2012	US 2012283145 A1 US 2014206581 A1	08-11-2012 24-07-2014
WO 2008043543 A2	17-04-2008	AU 2007306616 A1 CA 2667631 A1 CN 101573454 A DK 2097541 T3 EP 2097541 A2 ES 2401183 T3 HK 1134324 A1 JP 5645407 B2 JP 2010505429 A PT 2097541 E US 2012258876 A1 US 2014255926 A1 WO 2008043543 A2	17-04-2008 17-04-2008 04-11-2009 18-03-2013 09-09-2009 17-04-2013 03-05-2013 24-12-2014 25-02-2010 18-03-2013 11-10-2012 11-09-2014 17-04-2008