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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 February 2014 (27.02.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/030066 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/IB2013/002256

(22) International Filing Date:
22 August 2013 (22.08.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/692,208 22 August 2012 (22.08.2012) US
61/705,469 25 September 2012 (25.09.2012) US

(72) Inventors; and
(71) Applicants : **BERNITZ, Mats, Nilsson** [SE/SE]; Appelviksvagen 1, S-16753 Bromma (SE). **LARSSON, Chatarina** [SE/SE]; Cellovagen 124a, S-75654 Uppsala (SE). **GRUNDBERG, Ida** [SE/SE]; Marielundsgatan 3a, S-75325 Uppsala (SE).

(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, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/030066 A2

(54) **Title**: METHODS FOR IDENTIFYING NUCLEIC ACID SEQUENCES

(57) **Abstract**: Embodiments relate to the detection of RNA in a sample of cells. More particularly, methods concern the localized detection of RNA in situ. The method relies on the conversion of RNA to complementary DNA prior to the targeting of the cDNA with a padlock probe(s). The hybridization of the padlock probe(s) relies on the nucleotide sequence of the cDNA which is derived from the corresponding nucleotide sequence of the target RNA. Rolling circle amplification of the subsequently circularized padlock probe produces a rolling circle product which may be detected. Advantageously, this allows the RNA to be detected in situ. In additional methods, rolling circle amplification products are sequenced.

DESCRIPTION

METHODS FOR IDENTIFYING NUCLEIC ACID SEQUENCES

CROSS –REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. Provisional Patent Application Serial Nos. 61/692,208, filed August 22, 2012, and 61/705,469, filed September 25, 2012, both hereby incorporated by reference in their entirety.

BACKGROUND

[0002] It is generally desirable to be able sensitively, specifically, qualitatively and/or quantitatively to detect RNA, and in particular mRNA, in a sample, including for example in fixed or fresh cells or tissues. It may be particularly desirable to detect an mRNA in a single cell. For example, in population-based assays that analyze the content of many cells, molecules in rare cells may escape detection. Furthermore, such assays provide no information concerning which of the molecules detected originate from which cells. Expression in single cells can vary substantially from the mean expression detected in a heterogeneous cell population. It is also desirable that single-cell studies may be performed with single-molecule sensitivity which allows the fluctuation and sequence variation in expressed transcripts to be studied. Fluorescence in situ hybridization (FISH) has previously been used to detect single mRNA molecules in situ. Although permitting determination of transcript copy numbers in individual cells, this technique cannot resolve highly similar sequences, so it cannot be used to study, for example, allelic inactivation or splice variation and cannot distinguish among gene family members.

[0003] The only option available for assigning transcript variants to a single cell in a given tissue involves polymerase chain reaction (PCR) of laser-capture microdissected material, which is time consuming and error prone, and thus not suitable for diagnostics.

[0004] As an alternative to PCR- and hybridization-based methods, padlock probes (Nilsson et al., 1994) have for many years been used to analyze nucleic acids. These highly

selective probes are converted into circular molecules by target-dependent ligation upon hybridization to the target sequence. Circularized padlock probes can be amplified by RCA in situ (Lizardi et al., 1998), and thus can be used to provide information about the localization of target molecules, including, where DNA targets are concerned, at the single-cell level. Such a protocol is described in Larrson et al., 2004), in which the target DNA molecule is used to prime the RCA reaction, causing the RCP to be anchored to the target molecule, thereby preserving its localization and improving the in situ detection.

[0005] While RNA molecules can also serve as templates for the ligation of padlock probes (Nilsson et al., 2000), RNA detection with padlock probes in situ has so far proven more difficult than DNA detection and is subject to limitations (Lagunavicius et al., 2009). For example, the high selectivity reported for padlock probes with in situ DNA detection and genotyping has not been reproduced with detection of RNA targets in situ. This is possibly due to problems with ligation of DNA molecules on an RNA template, since it is known that both the efficiency and the specificity of the ligation reaction are lower compared to ligation on a DNA template (Nilsson et al.., 2000; Nilsson et al. 2001). It has recently been demonstrated that RNA molecules may be detected in situ with padlock probes and target-primed RCA (Lagunavicius et al., 2009; Stougaard et al., 2007). However, thus far, detection through target-primed RCA has for the most part been restricted to sequences in the 3′-end of non-polyadenylated RNA or sequences adjacent to the poly(A)-tail of mRNA. Since target-priming of the RCA reaction is dependent on a nearby free 3′-end that can be converted into an RCA primer, it is thought that this limitation results from the formation of RNA secondary structures which impede the polymerase action (3′ exonucleolysis) required to convert the RNA into a reaction primer. The detection efficiency of direct mRNA detection with padlock probes has been estimated to be as low as 1% (Nilsson et al., 2001). For the detection of non-polyadenylated RNA molecules, it has been noted that ligation of the probes using an internal hairpin structure as template resulted in higher detection efficiency than using the RNA molecule itself as ligation template (Stougaard et al., 2007). This indicates that better ligation conditions are required to be able to efficiently detect and genotype RNA directly with padlock probes in situ.

[0006] None of the methods for in situ detection of RNA presented thus far provide the possibility to detect sequence variation at the single nucleotide level and in particular to genotype transcripts. In the present embodiment, by converting an RNA target molecule into

cDNA, the reduction in padlock probe ligation efficiency and specificity is avoided and the excellent genotyping properties provided by padlock probes are preserved. In addition, it has been found that unlike many previously described methods, embodiments are not restricted to detection of sequences positioned at specific sites in the RNA molecules.

SUMMARY OF THE INVENTION

[0007] Embodiments generally concern the characterization, detection, and/or identification of nucleic acid sequences using sensitive and specific padlock probes that are capable of distinguishing between sequences with as few as one nucleotide difference.

[0008] Some embodiments concern the detection of RNA, especially mRNA, in a sample of cells. More particularly, in particular embodiments methods concern the localized detection of RNA, particularly mRNA, in situ. In certain aspects, the method relies on the conversion of RNA to complementary DNA (cDNA) prior to the targeting of the cDNA with a padlock probe(s). The hybridization of the padlock probe(s) relies on the nucleotide sequence of the cDNA which is derived from the corresponding nucleotide sequence of the target RNA. Rolling circle amplification (RCA) of the subsequently circularized padlock probe produces a rolling circle product (RCP) which allows detection of the RNA. Advantageously, the RCP may be localized to the RNA allowing the RNA to be detected in situ. Also, provided are kits for performing such methods.

[0009] Methods and compositions advantageously allow for detection of RNA, and particularly, the detection of single nucleotide variations in RNA. For example, a detection resolution may be achieved that allows the study of differences in the relative expression of two allelic transcripts directly in tissue. Such studies have recently been recognized as important in the context of large-scale analyses of allele-specific expression, since it has been shown that many genes undergo this type of transcriptional regulation and that the allelic expression can differ among tissues. Furthermore, it has been shown that most human genes undergo alternative splicing, which could now be studied at the single-cell level using the methods described herein. No other in situ method exists today that can perform multiplex detection of expressed single nucleotide sequence variants in RNA. It is believed that the present method can meet this need, and that the ability it provides to visualize transcriptional

variation directly in cells and tissues will be of value in both research and diagnostics, providing new insights about the human transcriptome.

[0010] Other methods and compositions can be used for one or more of the following: characterizing a target RNA, determining the sequence of a target RNA, determining the sequence of a target RNA by sequencing a complement of the target RNA, detecting in situ a target RNA, determining the sequence of a target RNA in situ, identifying sequence information for a target RNA, detecting in situ a nucleic acid and determining the sequence of the nucleic acid, identifying a cancer mutation in a target RNA, localizing a cancer mutation in a sample, identifying a cell expressing a mutant RNA, identifying a cell expressing an RNA associated with cancer, or analyzing a target RNA.

[0011] According to some methods, transcript detection in situ is accomplished by first converting the at least one mRNA into localized cDNA molecules that are detected with padlock probes and target-primed RCA (FIG. 1). Whilst of particular applicability to mRNA, the method may be used for the detection of any RNA molecule present in a cell, including but not limited to viral RNA, tRNA, rRNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), microRNA (miRNA), small interfering RNA (siRNA), piwi-interacting RNA (piRNA), antisense RNA and non-coding RNA. The RNA is converted into cDNA, typically in a reverse transcriptase reaction comprising a reverse transcriptase enzyme and one or more reverse transcriptase primers. A ribonuclease is employed to digest the RNA in the resultant RNA:DNA duplex thus making the cDNA strand available for hybridization to a padlock probe(s). Hybridization of the padlock probe(s) to the cDNA allows circularization of the probe by direct or indirect ligation of the ends of the probe(s). The circularized padlock probe is then subjected to RCA and a RCP is detected by any appropriate means available in the art. The method may, in specific embodiments, also be used for localizing more than one target RNA, e.g., 2, 3, 4, 5, 6 or more target RNAs. These target RNAs may be derived from the same gene, or from different genes, or be derived from the same genomic sequence, or from different genomic sequences.

[0012] In one embodiment, there are methods for in situ detection of at least one target RNA in a sample of one or more cells, comprising: generating a cDNA complementary to an RNA in the sample; adding a ribonuclease to the sample to digest the RNA hybridized to the cDNA; contacting the sample with one or more padlock probes wherein the padlock probe(s)

comprise terminal regions complementary to immediately adjacent regions on the cDNA and hybridizing the padlock probe to the cDNA at the complementary terminal regions; ligating the ends of the padlock probe(s); subjecting the circularized padlock probe(s) to rolling circle amplification (RCA); and detecting the rolling circle amplification product(s).

[0013] Some methods are provided for detecting a genetic mutation in a gene of a cell comprising: a) incubating the cell with a reverse transcriptase and a primer to hybridize the primer to RNA from the gene to generate a cDNA of all or part of the gene; b) incubating the cell with a ribonuclease under conditions to digest the RNA; c) incubating the cDNA with at least one padlock probe under conditions to hybridize the padlock probe to the cDNA, wherein the padlock probe comprises two terminal ends that are complementary to different but adjacent regions of the cDNA; d) incubating the cDNA and at least one padlock probe with a ligase under conditions to join terminal ends of the padlock probe; e) incubating the cDNA and the at least one ligated padlock probe with a polymerase and labeled nucleotides under conditions to amplify the at least one padlock probe and generate labeled, amplified padlock probes; and, f) assaying for one or more labeled nucleotides in the amplified probes to detect the genetic mutation.

[0014] In additional embodiments, there are methods for determining a nucleic acid sequence from a tissue section comprising: (a) generating a cDNA complementary to an RNA containing the nucleic acid sequence in the tissue section; (b) incubating the cDNA with a ribonuclease to the sample to digest the RNA; (c) hybridizing one or more padlock probes to the cDNA, wherein the padlock probe(s) comprise one or two terminal regions having the nucleic acid sequence; (d) incubating the hybridized padlock probes and cDNA with ligase under conditions to ligate the ends of the padlock probe(s); (e) incubating the replicating the padlock probe(s) using a polymerase to create an amplified product; and, (f) sequencing the complement of the nucleic acid sequence in the amplified product to determine the nucleic acid sequence and/or its complement.

[0015] In a further embodiment, there are methods for determining the presence and/or location of a genetic sequence in a cell in a biological sample comprising: (a) hybridizing a DNA complement having the genetic sequence to RNA; (b) digesting RNA hybridized to the DNA complement; (c) hybridizing a first padlock probe to at least a portion of the DNA complement, wherein the padlock probe comprises the genetic sequence on one of two

terminal ends that are complementary to different but immediately adjacent regions of the DNA complement; (d) joining the two terminal ends of the padlock probe; and (e) replicating the circularized probe to yield a nucleic acid molecule comprising multiple copies of the replicated probe or subjecting the circularized probe to rolling circle amplification (RCA). In certain embodiments, methods also comprise (f) detecting the rolling circle amplification products. In some embodiments, there is a step of detecting the presence or absence of the genetic sequence in the cell using a probe that hybridizes to the nucleic acid molecule. In certain aspects, the method further comprises generating the DNA complement that is hybridized to the RNA. In further embodiments, the two terminal ends of the padlock probe are joined using a ligase.

[0016] In another embodiment, methods for identifying a cell in a tissue sample that has a specific nucleic acid sequence are provided comprising: (a) incubating the cell with a DNA complement that includes the specific nucleic acid sequence to generate an RNA-DNA hybrid; (b) incubating the RNA target molecule with a ribonuclease under conditions to digest at least part of the RNA-DNA hybrid; (c) incubating the DNA complement with a padlock probe under conditions to hybridize the padlock probe to the DNA complement comprising the specific nucleic acid sequence, wherein the padlock probe comprises two terminal ends that are complementary to different but immediately adjacent regions of the DNA complement; (d) incubating the DNA complement and padlock probe with a ligase under conditions to join terminal ends of the padlock probe; (e) incubating the ligated padlock probe with a polymerase and nucleotides under conditions to prime replication of the padlock probe with the DNA complement and generate a nucleic acid with multiple copies of the replicated padlock probe; and (f) incubating the nucleic acid with multiple copies of the replicated padlock probe with one or more complementary oligonucleotides to detect the presence or absence of the specific sequence.

[0017] In one embodiment, there are methods for identifying a cell in a cell sample that has a specific nucleic acid sequence comprising: (a) incubating the cell sample with a ribonuclease-resistant primer that is immobilized to the sample and reverse transcriptase under conditions to generate a DNA complement of an RNA, wherein the DNA complement comprises the specific nucleic acid sequence; (b) incubating the cell sample with a ribonuclease under conditions to digest at least part of the RNA; (c) incubating the DNA complement with a padlock probe under conditions to hybridize the padlock probe to the

DNA complement comprising the specific nucleic acid sequence, wherein the padlock probe comprises two terminal ends that are complementary to different but immediately adjacent regions of the DNA complement; (d) incubating the DNA complement and padlock probe with a ligase under conditions to join terminal ends of the padlock probe; (e) incubating the ligated padlock probe with a polymerase and nucleotides under conditions to prime replication of the padlock probe with the DNA complement and generate a nucleic acid with multiple copies of the replicated padlock probe; and (f) incubating the nucleic acid with multiple copies of the replicated padlock probe with one or more nucleic acid probes to detect the presence or absence of the specific sequence.

[0018] In another embodiment, methods are provided for in situ localization of a nucleic acid sequence in a cell in a biological sample on a slide comprising: (a) incubating an immobilized biological sample on solid support with reverse transcriptase and a ribonuclease-resistant primer under conditions to generate a nucleic acid molecule that contains the nucleic acid sequence and that hybridizes to a complementary RNA molecule in the cell to form an RNA-DNA hybrid; (b) adding a ribonuclease and incubating the ribonuclease under conditions to digest RNA in the RNA-DNA hybrid; (c) incubating the digested RNA-DNA hybrid under conditions to hybridize a complementing padlock probe to the DNA portion of the digested RNA-DNA hybrid, wherein the padlock probe comprises the nucleic acid sequence and has two terminal ends that are complementary to different but immediately adjacent regions of the DNA; (d) incubating the padlock probe hybridized to the DNA portion of the RNA-DNA hybrid with a ligase under conditions to ligate the terminal ends of the padlock probe; (e) incubating the ligated padlock probe with a polymerase and nucleotides under conditions to create a primer from the DNA that is used to replicate the padlock probe and generate a nucleic acid with multiple copies of the replicated padlock probe; and (f) incubating the nucleic acid with one or more complementing nucleic acid probes to detect the presence or absence of the specific sequence.

[0019] In specific embodiments of the methods for identifying a cell in a tissue sample, the methods for identifying a cell in a cell sample, or the methods for in situ localization of a nucleic acid sequence in a cell in a biological sample, e.g. as mentioned above, the sample is a formalin-fixed paraffin-embedded tissue section.

[0020] In another embodiment, there are methods for localized in situ detection of at least one RNA in a sample of cells, the methods comprising: (a) contacting the sample with a reverse transcriptase and a reverse transcriptase primer to generate cDNA from RNA in the sample; (b) adding a ribonuclease to the sample to digest the RNA hybridized to the cDNA; (c) contacting the sample with one or more padlock probes wherein the padlock probe(s) comprise terminal regions complementary to the cDNA and hybridizing the padlock probe(s) to the cDNA at the complementary terminal regions; (d) circularizing the padlock probe(s) by ligating, directly or indirectly, the ends of the padlock probe(s); (e) subjecting the circularized padlock probe(s) to rolling circle amplification (RCA) using a DNA polymerase having 3′-5′ exonuclease activity wherein, if necessary, the exonuclease activity digests the cDNA to generate a free 3′ end which acts as a primer for the RCA; and (f) detecting the rolling circle amplification product(s).

[0021] Methods of certain embodiments concern localizing or detecting in situ at least one target RNA in a sample of one or more cells, comprising: a) hybridizing the target RNA with a complementary nucleic acid that comprises a region complementary to the target RNA; b) digesting the RNA hybridized to the complementary nucleic acid; c) contacting the sample with one or more padlock probes, wherein the padlock probe(s) comprise terminal regions complementary to the cDNA and hybridizing the padlock probe to the cDNA at the complementary terminal regions; d) joining the ends of the padlock probe(s); and, e) subjecting the circularized padlock probe(s) to rolling circle amplification (RCA). In some embodiments, methods also involve generating the complementary nucleic acid while in others methods involve obtaining or providing the complementary nucleic acid. In some embodiments, methods further comprise sequencing all or part of one or more rolling circle amplification product(s).

[0022] The methods thus involve detecting the rolling circle amplification product (RCP) as a means of detecting the target RNA. The RCP is generated as a consequence of padlock probe recognition of a cDNA complementary to the target RNA (i.e. padlock probe binding to the cDNA complement of the target RNA by hybridization to complementary sequences in the cDNA) and ligation of the padlock probe to generate a circular template for the RCA reaction. The RCP may thus be viewed as a surrogate marker for the cDNA, which is detected to detect the RNA.

[0023] As discussed above, the method may be used for the detection of any RNA molecule type or RNA sequence present in a cell. In some embodiments, the method is used for the detection of mRNA. The cDNA complementary to the RNA in the sample may be generated by contacting the sample with an RNA-dependent DNA polymerase and a primer. The RNA dependent DNA polymerase may be, for example, a reverse transcriptase, such as an MMLV reverse transcriptase or an AMV reverse transcriptase.

[0024] In certain aspects, the primer used for first strand cDNA synthesis is ribonuclease resistant. A primer which is "ribonuclease resistant" means that it exhibits some (i.e., a measurable or detectable) degree of increased resistance to ribonuclease action (in particular to the action of an RNase H) over a naked, unmodified primer of the same sequence. Thus the primer is at least partially protected from digestion by the ribonuclease, or more particularly when the primer is hybridized to its RNA template, the primer/template hybrid is at least partially protected from ribonuclease digestion. In some embodiments at least 50% survives the ribonuclease treatment, while in further embodiments at least 60, 70, 80 or 90%, or even 100% survives the ribonuclease treatment. A primer may, for example, comprise 2'O-Me RNA, methylphosphonates or 2' Fluor RNA bases, locked nucleic acid residues, or peptide nucleic acid residues, which make the primer resistant to digestion by ribonucleases.

[0025] In one embodiment, the primer comprises 2, 3, 4, 5, 6, 7, 8, 9 or more locked nucleic acids separated by 1 or more natural or synthetic nucleotides in the primer sequence. In certain embodiments, the primer comprises between 4 to 9 locked nucleic acids, with each locked nucleic acid being separated for the other locked nucleic acids by 1 or more natural or synthetic nucleotides in the primer sequence.

[0026] The term "reverse transcriptase primer" or "RT primer" as used herein (also known as a cDNA primer) refers to an oligonucleotide capable of acting as a point of initiation of cDNA synthesis by an RT under suitable conditions. Thus, a reverse transcription reaction is primed by an RT primer. The appropriate length of an RT primer typically ranges from 6 to 50 nucleotides or from 15 to 35 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the mRNA template, but may still be used. Shortening the primer from 30 to 25 nucleotides did not affect its function. A primer need not reflect the exact sequence of the template nucleic acid, but must be

sufficiently complementary to hybridize with the template. The design of suitable primers for cDNA synthesis is well known in the art.

[0027] Typically, an RT primer is designed to bind to the region of interest in the RNA, for example a region within a particular RNA it is desired to detect, or a region within which sequence variations may occur (for example, allelic or splice variants, polymorphisms or mutations, etc., e.g. SNPs, etc.).Thus, in seeking to detect the presence or absence of particular mutations, etc. (e.g. in a genotyping context), the RT primer may be designed to bind in or around the region within which such mutations occur (e.g. near to such a region, for example within 100, 70, 50, 30, 20, 15, 10 or 5 nucleotides of such a region). Such mutations or sequence variations may be associated with disease (e.g. cancer) or disease risk or predisposition, or may with response to a therapeutic treatment, etc.

[0028] RT primers can incorporate additional features which allow for the immobilization of the primer to or within a cell in the sample but do not alter the basic property of the primer, that of acting as a point of initiation of cDNA synthesis. Thus it is contemplated that the primer may be provided with a functional moiety or means for immobilization of the primer to a cell or cellular component. This may for example be a moiety capable of binding to or reacting with a cell or cellular component and, as described above, such a cellular component may include RNA. Thus, the functional moiety may include a moiety(ies) which allow the primer to remain hybridized to the primer binding site within the template RNA, namely a moiety(ies) which render the primer resistant to ribonuclease digestion.

[0029] The primer may be modified to incorporate one or more reactive groups, e.g. chemical coupling agents, capable of covalent attachment to cells or cellular components. This may be achieved by providing the primer with chemical groups or modified nucleotide residues which carry chemical groups such as a thiol, hydroxy or amino group, a phosphate group via EDC (1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride), NHS (N-hydroxysuccinimide)-esters, etc. which are reactive with cellular components such as proteins, etc. Such chemical coupling groups and means of introducing them into nucleic acid molecules are well known in the art. Potential reactive functionalities thus include nucleophilic functional groups (amines, alcohols, thiols, hydrazides), electrophilic functional groups (aldehydes, esters, vinyl ketones, epoxides, isocyanates, maleimides), functional groups capable of cycloaddition reactions, forming disulfide bonds, or binding to metals.

Specific examples include primary and secondary amines, hydroxamic acids, N-hydroxysuccinimidyl esters, N-hydroxysuccinimidyl carbonates, oxycarbonylimidazoles, nitrophenylesters, trifluoroethyl esters, glycidyl ethers, vinylsulfones, and maleimides.

[0030] Alternatively or in addition, the primer may be provided with an affinity binding group capable of binding to a cell or cellular component or other sample component. Such an affinity binding group may be any such binding group known in the art which has specific binding activity for a corresponding binding partner in or on a cell, tissue, sample component, etc. Thus, representative binding groups include antibodies and their fragments and derivatives (.e.g. single chain antibodies, etc.), other binding proteins, which may be natural or synthetic, and their fragments and derivatives, e.g. lectins, receptors, etc., binding partners obtained or identified by screening technology such as peptide or phage display, etc., aptamers and such like, or indeed small molecule binding partners for proteins e.g. for receptors and other proteins on or within cells. Such immobilization systems may work best in relation to cellular components which are abundant e.g. actin filaments.

[0031] The target RNA or the synthesized cDNA may be attached to a synthetic component in the sample, e.g. a synthetic gel matrix, instead of the native cellular matrix to preserve the localization of the detection signals. The cells or tissue may be immersed in a gel solution that upon polymerization will give rise to a gel matrix to which the cDNA or target can be attached. For example, if an Acrydite modification is included at the 5′ end of the cDNA primer, the cDNA can be covalently attached to a polyacrylamide matrix (Mitra and Church, 1999).

[0032] Alternatively or in addition to the aforementioned modifications to the RT primer, the modification described above may be used in which the 5′ phosphate of the primer may be linked to amines present on proteins in the cellular matrix via EDC-mediated conjugation, thus helping to maintain the localization of the RNA relative to other cellular components. Such a technique has previously been described in relation to microRNAs and their detection via in situ hybridization (Pena et al., 2009).

[0033] To ensure good ribonuclease resistance it may in certain instances be advantageous to use several modified residues in the RT primer, such as 2, 3, 4, 5 or 6 modified residues in a row for example. In some embodiments, modified residues may be incorporated into the RT primer every second, or every third, residue. In additional embodiments, the RT primer may

comprise, comprise at least, or comprise at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more modified residues (or any range derivable therein). In the literature various modifications of nucleic acids that impart ribonuclease resistance have been described and any modification that prevents, or partially prevents, digestion of the RT primer or the RNA to which it is hybridized is encompassed in this method.

[0034] In one embodiment the modifications (e.g. modified residues) are placed at the 5′ end of the primer (in the 5′ region of the primer) and the 3′ end is left unmodified. For example, in some embodiments, at least or at most 1, 2, 3, 4, 5 or 6 residues from the 3′ end (or any range of derivable therein) are unmodified.

[0035] A preferred modification to confer ribonuclease resistance is the incorporation of LNA residues into the RT primer. Thus the RT primer may include at least 1 LNA residue and in certain embodiments include at least or at most 2, 3, 4, 5, 6, 7, 8 or 9 LNA residues (or any range derivable therein). As well conferring ribonuclease resistance, LNA monomers have enhanced hybridization affinity for complementary RNA, and thus may be used to enhance hybridization efficiency.

[0036] In a representative embodiment, the RT primer comprises LNA residues every second, or every third, residue. LNA is a bicyclic nucleotide analogue wherein a ribonucleoside is linked between the 2'-oxygen and the 4'-carbon atoms by a methylene unit. Primers comprising LNA exhibit good thermal stabilities towards complementary RNA, which permits good mismatch discrimination. Furthermore, LNA offers the possibility to adjust Tm values of primers and probes in multiplex assays.

[0037] The cDNA that is generated may be from 10 nucleotides to 1000 nucleotides in length, and in certain embodiments may range from 10 to 500 nucleotides in length including from 50 to 500 nucleotides in length, e.g., from 90 to 400 nucleotides in length, such as from 90 to 200 nucleotides in length, from 90 to 100 nucleotides in length, and so on. In certain representative embodiments, the cDNA may range in length from 10 to 100 nucleotides in length, from 30 to 90 nucleotides in length, from 14 to 70 nucleotides in length, from 50 to 80 nucleotides in length, and any length of integers between the stated ranges.

[0038] The cDNA may be made up of deoxyribonucleotides and/or synthetic nucleotide residues that are capable of participating in Watson-Crick-type or analogous base pair

interactions. Thus the nucleotides used for incorporation in the reverse transcriptase step for synthesis of the cDNA may include any nucleotide analogue or derivative that is capable of participating in the reverse transcriptase reaction (i.e., capable of being incorporated by the reverse transcriptase).

[0039] Ribonucleases, also known as RNases, are a class of enzymes that catalyze the hydrolysis of RNA. A ribonuclease for use according to various embodiments will be able to degrade RNA in an RNA:DNA duplex. The RNases H are a family of ribonucleases that cleave the 3′-O-P-bond of RNA in a DNA:RNA duplex to produce 3′-hydroxyl and 5′-phosphate terminated products. Since RNase H specifically degrades the RNA in RNA:DNA hybrids and will not degrade DNA or unhybridized RNA it is commonly used to destroy the RNA template after first-strand cDNA synthesis by reverse transcription. RNase H thus represents a preferred class of enzymes for use. Members of the RNase H family can be found in nearly all organisms, from archaea and prokaryota to eukaryota. Again, suitable ribonuclease, particularly RNase H, enzymes are well-known and widely available.

[0040] Upon the hybridization of the terminal regions of a padlock probe to a complementary cDNA sequence, the padlock probe is "circularized" by ligation. The circularization of the padlock probe(s) may be carried out by ligating, directly or indirectly, the ends of the padlock probe(s). Procedures, reagents and conditions for this are well known and described in the art and may be selected according to choice. Suitable ligases include e.g., Tth DNA ligase, Taq DNA ligase, Thermococcus sp. (strain 9oN) DNA ligase (9oNTM DNA ligase, New England Biolabs), AmpligaseTM (Epicentre Biotechnologies) and T4 DNA ligase. In specific embodiments, the in the circularization of the padlock probe(s) step, the terminal regions of the padlock probe may hybridize to non-contiguous regions of the cDNA such that there is a gap between the terminal regions. In further specific embodiments of this method, the gap may be a gap of 1 to 60 nucleotides, such as a gap of 1 to 40 nucleotides or a gap of 3 to 40 nucleotides. In specific embodiments, the gap may be a gap of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 30, 32, 35, 37, 40, 42, 45, 47, 50, 52, 55, 57 or 60 nucleotides, of any integer of nucleotides in between the indicated values. In further embodiments, the gap may be larger than 60 nucleotides. In further embodiments, the gap may have a size of more than 60 nucleotides. In further embodiments, the gap between the terminal regions may be filled by a gap oligonucleotide or by extending the 3' end of the padlock probe. The gap oligonucleotide may

accordingly have a size of 1 to 60 nucleotides, e.g. a size of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 30, 32, 35, 37, 40, 42, 45, 47, 50, 52, 55, 57 or 60 nucleotides, or any integer of nucleotides in between the indicated values. In further embodiments, the size of the gap oligonucleotide may be more than 60 nucleotides.

[0041] Rolling circle amplification or "RCA" of the circularized padlock probe results in the synthesis of a concatemeric amplification product containing numerous tandem repeats of the probe nucleotide sequence. RCA reactions and the conditions therefore are widely described in the literature and any such conditions, etc. may be used, as appropriate. The ligation reaction may be carried out at the same time (i.e. simultaneously) as the RCA reaction of step, i.e. in the same step. In some embodiments, the RCA reaction is primed by the 3′ end of the cDNA strand to which the padlock probe has hybridized. In other embodiments, instead of priming the RCA reaction with the 3′ end of the cDNA, a primer is hybridized to the padlock probe and primes the RCA reaction. In certain aspects, this primer hybridizes to a region of the padlock probe other than the 5′ and 3′ terminal regions of the padlock probe.

[0042] Where the RCA reaction is primed by the 3′ end of the cDNA strand to which the padlock probe has hybridized, any unpaired 3′ nucleotides in the cDNA are removed in order to generate the primer for RCA. This may be achieved by using a polymerase having 3′-5′ exonuclease activity. Such target-primed RCA procedures are known and described in the art as are appropriate polymerase enzymes for such use. Thus, for example, a DNA polymerase such phi29 (ϕ29) polymerase, Klenow fragment, Bacillus stearothermophilus DNA polymerase (BST), T4 DNA polymerase, T7 DNA polymerase, or DNA polymerase I may be used. The skilled person may readily determine other suitable polymerases which might be used, including, for example, DNA polymerases that have been engineered or mutated to have desirable characteristics. In the RCA reaction, the polymerase thus extends the 3′ end of the cDNA using the circularized padlock probe as template. As a result of RCA, concatemeric amplification products containing numerous tandem repeats of the probe nucleotide sequence are produced and may be detected as indicative of the presence and/or nature of a RNA in the sample. Alternatively, a separate enzyme having 3′-5′ exonuclease activity may be added to the reaction to generate the free 3′ end, in which case a DNA polymerase lacking 3′-5′ exonuclease activity could then be used for RCA. In some cases, depending on the proximity of the hybridized padlock probe to the 3′ end of the target cDNA,

it may not be necessary to digest the cDNA to generate a free 3′ end at the appropriate position for it to act as a primer for RCA.

[0043] In various methods described herein, there may be one or more wash steps. Multiple washes may be employed at one or more points during a process. On certain embodiments there are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more wash steps (and any range derivable therein) in a method. Each wash may involve the same or different washing reagents depending on the purpose for the wash.

[0044] The terms "padlock probe" and "probe" and their plural forms are synonymous and are used interchangeably throughout this specification. The use of a single padlock probe occurs in the case of a "simplex" (as opposed to "multiplex") embodiment of the method, i.e. when a single RNA or a single variant in a RNA are to be detected. It will be understood that the term "single" as used in relation to a padlock probe, or the RNA, means single in the sense of a "single species," i.e. a plurality of RNA molecules of the same type may be present in the sample for detection, and a plurality of identical padlock probes specific for that RNA may be used, but such pluralities relate only to a unique sequence of RNA or padlock probe. In multiplex embodiments, two or more different target RNAs are to be detected in a sample of cells. In such embodiments, the sample of cells is contacted with a plurality of padlock probes for each target RNA, such that the number of probes contacted with the sample may be two or more, e.g., three or more, four or more, etc. Optionally, up to 10, 15 or 20 probes may be used. Such methods find particular use in high-throughput applications. For example, the method may employ or may employ at least 1, 2, 3, 4, 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, or any range derivable therein, padlock probes in a single reaction.

[0045] For example, in one embodiment, the method comprises contacting the sample with at least a first and a second padlock probe, wherein the first padlock probe comprises terminal regions complementary to immediately adjacent regions on the cDNA, and wherein the second padlock probe comprises terminal regions that differ from the terminal regions of the first padlock probe only by a single nucleotide at the 5′ or 3′ terminus of the second padlock probe. In this manner, the two padlock probes can be used to detect a single nucleotide differences in an RNA sequence. For example, the first padlock probe may be configured to hybridize to a cDNA complementary to a wild-type mRNA sequence, and the second padlock

probe is configured to hybridize to a cDNA complementary to a single nucleotide variant of the mRNA sequence. In addition to detecting nucleic acid substitutions, the padlock probes may be configured to detect insertions or deletions in a nucleic acid sequence.

[0046] The padlock probe may be of any suitable length to act as an RCA template. For example, the padlock probe may have an overall length (including two arms and a backpiece) of between 50 and 150 nucleotides, of between 60 to 120 nucleotides, or of between 70 to 100 nucleotides. Thus, the padlock probe may have, for instance, a length of, of at least, or of at most 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 nucleotides, or any range derivable therein. The arms of the padlock probes may have any suitable length, e.g. each may have a length of 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, e.g.13, 24, 25, 26, 27, 28, 29, 30, 32, 35, 36, 37, 38, 39 or 40 nucleotides, or any range derivable therein. The length of the two arms of the padlock probes may, in certain embodiments, be identical or essentially identical, e.g. showing a length difference of 1-2 nucleotides. In further embodiments, the length of the two arms may differ one from the other by more than 2 nucleotides, e.g. one arm having a length of 15 nucleotides, whereas the other having a length of 20 nucleotides. The length difference in some embodiments may not surpass 5 to 7 nucleotides. In addition to the end regions, which are complementary to the cDNA, the probe may contain features or sequences or portions useful in RCA or in the detection or further amplification of the RCA product. Such sequences may include binding sites for an RCA primer, hybridization probes, and/or amplification or sequencing primers. Thus, a padlock probe may be viewed as having a "back piece" which links the 3' and 5' target-complementary regions. By including within this back piece or linking region a particular sequence, to which when amplified by RCA of the circularized probe, a detection probe or primer may bind in the RCP, the padlock probe may be seen as having, or more particularly as providing, a detection site for detection of the RCP. Accordingly, the padlock probe may contain an arbitrary "tag" or "barcode" sequence which may be used diagnostically to identify the cDNA, and by extension the corresponding mRNA, to which a given RCA product relates, in the context of a multiplex assay. Such a sequence is simply a stretch of nucleotides comprising a sequence designed to be present only in the padlock probe which is "specific for" (i.e. capable of hybridizing only to) a particular cDNA. Thus, for example in the context of padlock probes for genotyping, the tag sequence

(or detection site) may be different for the padlock probes designed to detect the wild-type sequence and the mutant(s)/sequence variant(s) thereof.

[0047] In certain embodiments, a detection probe that is complementary to the backbone sequence of a padlock probe may be labeled or tagged. In some embodiments, the detection probe has a substance attached to it that can be detected directly or indirectly. In further embodiments, the substance attached to the detection probe is or contains an epitope for an antibody or antibody fragment. In certain embodiments, the substance is or comprises a hapten. Therefore, in some methods provided herein, a detection probe is recognized by an antibody or antibody fragment. Detection of the rolling circle amplification product may involve detection of the antibody or antibody fragment directly or indirectly. A direct method might involve the use of an antibody is itself is labeled with a detectable moiety. Alternatively, detection of the antibody of antibody fragment may involve a secondary antibody that is specific for the antibody or antibody fragment that binds the substance on the detection probe.

[0048] A padlock probe may have ends that can be joined when juxtaposed with one another upon hybridization of a complementing nucleic acid. The ends may be chemically or enzymatically joined. In some embodiments, ends are joined by ligating the ends using a ligase.

[0049] In certain aspects, the padlock probes comprise a "tag" or "detection probe binding region." The detection probe binding region may be used to incorporate detection probe binding regions into the rolling circle amplification products for subsequent hybridization to labeled detection probes. Different padlock probes may have different detection probe binding regions such that differentially labeled detection probes may be used in the detection of the rolling circle amplification products. For example, a first padlock probe may comprise a first detection probe binding region, and a second padlock probe may comprise a second detection probe binding region. The sample may then be contacted with a first labeled detection probe comprising a sequence identical to the first detection probe binding region of the first padlock probe, and a second labeled detection probe comprising a sequence identical to the second detection probe binding region of the first padlock probe, such that the first and second labeled detection probes hybridize to the rolling circle amplification products, if any, generated by the first and second padlock probes.

[0050] The term "detection" is used broadly herein to include any means of determining, or measuring (e.g. quantitatively determining), the presence of at least one RNA (i.e. if, or to what extent, it is present, or not) in the sample. "Localized" detection means that the signal giving rise to the detection of the RNA is localized to the RNA. The RNA may therefore be detected in or at its location in the sample. In other words the spatial position (or localization) of the RNA within the sample may be determined (or "detected"). This means that the RNA may be localized to, or within, the cell in which it is expressed or to a position within the cell or tissue sample. Thus "localized detection" may include determining, measuring, assessing or assaying the presence or amount and location, or absence, of RNA in any way. Quantitative and qualitative determinations, measurements or assessments are included, including semi-quantitative. Such determinations, measurements or assessments may be relative, for example when two or more different RNAs in a sample are being detected. In certain embodiments, a detection probe is labelled. In some embodiments, a labeled detection or padlock probe comprises one or more fluorescent labels, enzymatic labels, chromogenic labels, radioactive labels, luminescent labels, magnetic labels, or electron-density labels. In particular embodiments, one or more probes that are differentially labeled with respect to one another may be employed or included in methods or kits discussed herein.

[0051] In some embodiments, a rolling circle amplification product or replicated circularized padlock probe is labeled directly. In some embodiments, methods involve subjecting the circularized padlock probe(s) to rolling circle amplification by adding labeled nucleotides to generate labeled, amplified padlock(s). In some embodiments, at least two different padlock probes are used. In further aspects, the at least two different padlock probes have different backbone sequences. It is contemplated that in some cases a difference in backbone sequence comprises a difference in nucleotide content. In particular embodiments, at least two differentially labelled nucleotides are employed. In some cases, at least two different padlock probes are used and the amplified padlock probes are differentially labeled. The term “backbone sequence” refers to the contiguous sequence in the padlock probe that is not complementary to the target sequence. The backbone sequence lies between the two arms of a padlock probe that are complementary to a target sequence.

[0052] In some embodiments, there may be at least two padlock probes that have different backbone sequences. In some instances, a difference between backbone sequences comprises a difference in nucleotide content by at least 2x. For example, one backbone may have at

least twice as many of a specific nucleotide as the other backbone. The nucleotide content of G, A, T, or C may vary between padlock probes such that they can be distinguished from one another. In one embodiment, a difference in the nucleotide content comprises a difference in the number of guanines or G nucleotides in the backbone sequence. In another embodiment a difference in the nucleotide content comprises a difference in the number of adenines or A nucleotides in the backbone sequence. In another embodiment a difference in the nucleotide content comprises a difference in the number of thymidine or T nucleotides in the backbone sequence. In another embodiment a difference in the nucleotide content comprises a difference in the number of cytosine or C nucleotides in the backbone sequence. In certain cases, a backbone sequence does not have at least one of the four nucleotides (G, A, T, or C). In further examples, one of the padlock probes is missing at least one of G, A, T, or Cs, and another padlock probe is also missing at least one of G, A, T, or Cs; the two can be distinguished if they differ by which nucleotide each is missing. In further embodiments, four different padlock probes that each lack a different nucleotide may be employed.

[0053] In other embodiments, rolling circle amplification product(s) are detected by sequentially adding at least two probes. In some cases, each probe is detected prior to the addition of a next probe. Methods may involve eliminating what is detected prior to the addition of the next probe, where "eliminating" means a level of detection that is at or below background levels of detection. In some cases, this may involve photobleaching.

[0054] In some embodiments, rolling circle amplification product(s) are detected with one or more probes that comprises one or more branches having one or more labeling moieties on each branch. These branched probes increase signal that can be detected.

[0055] As used herein, the term "in situ" refers to the detection of at least one RNA in its native context, i.e. in the cell, bodily fluid, or tissue in which it normally occurs. Thus, this may refer to the natural or native localization of an RNA. In other words, the RNA may be detected where, or as, it occurs in its native environment or situation. Thus, the RNA is not moved from its normal location, i.e. it is not isolated or purified in any way, or transferred to another location or medium, etc. Typically, this term refers to the RNA as it occurs within a cell or within a cell, organ, bodily fluid, or tissue sample, e.g. its native localization within the cell or tissue and/or within its normal or native cellular environment. In certain embodiments, a sample is on a solid support. The solid support may be a slide, a bead, an

array, or chip. This is in contrast to a sample in solution, such as in a tube. In certain embodiments, the solid support is a slide, which may or may not have a cover, such as a coverslip or covertile.

[0056] In certain embodiments, methods may also involve a sample that is stained or a sample that is stained during or after contact with one or more padlock probes. In specific embodiments, a sample, such as cells or tissue, is stained prior to contact with one or more padlock probes. In particular embodiments, a sample is stained with hematoxylin and eosin.

[0057] A variety of labels are known for labeling nucleic acids and may be used in the detection of rolling circle amplification products. Non-limiting examples of such labels include fluorescent labels, chromogenic labels, radioactive labels, luminescent labels, magnetic labels, and electron-density labels. Labels may be incorporated directly into the amplification product, such as with modified or labeled dNTPs during amplification. Alternatively, the amplification products may be labeled indirectly, such as by hybridization to labeled probes. In multiplex reactions, it is contemplated that a different label may be used for each different amplification product that may be present in the reaction.

[0058] The method of detection will depend on the type of label used. In certain embodiments, the detection is by imaging or direct visualization of fluorescent or chromogenic labels. Accordingly, the present method allows for the detection of the amplification products in situ at the location of the target RNA. This sensitivity permits, for example, genotyping at the single-cell level.

[0059] In certain embodiments, methods will also include incubating the amplification product with a detection probe under conditions to allow hybridization between the product and the probe. In some embodiments, the detection probe has one or more fluorescent labels, enzymatic labels, epitope labels, chromogenic labels, radioactive labels, luminescent labels, magnetic labels, or electron-density labels. Furthermore, methods may also include incubating the detection probe with one or more polypeptides that binds the one or more labels. In some situations, at least one or more polypeptides is an antibody. It is further contemplated that methods may also involve incubating the one or more polypeptides that binds the one or more labels with a secondary polypeptide that binds the label-binding polypeptide. In some instances, the secondary polypeptide is an antibody. Furthermore, some methods involve a label-binding polypeptide or a secondary polypeptide that comprises a

detectable label. Additional embodiments, involve detecting the rolling circle amplification product(s), which may be achieved by steps that include assaying for the label(s) on the detection probe. In certain embodiments, methods also involve incubating the rolling circle amplification product(s) and label(s) on the probe with an enzyme substrate to detect the rolling circle amplification product(s).

[0060] In certain embodiments, following rolling circle amplification or replication of a circularized padlock probe, resulting nucleic acid molecules such as rolling circle amplification products or replicated circularized padlock probes may be sequenced. All or part of the products or probes may be sequenced. In certain embodiments, the identity of a single nucleotide in the RCA product or replicated circularized probe may be determined. In certain embodiments, the identity is determined by sequencing that position or nucleotide.

[0061] The "sample" may be any sample of cells in which an RNA molecule may occur, to the extent that such a sample is amenable to in situ detection. Typically, the sample may be any biological, clinical or environmental sample in which the RNA may occur, and particularly a sample in which the RNA is present at a fixed, detectable or visualizable position in the sample. The sample will thus be any sample which reflects the normal or native (in situ) localization of the RNA, i.e. any sample in which it normally or natively occurs. The sample may, for example, be derived from a tissue or organ of the body, or from a bodily fluid. Such a sample will advantageously be or comprise a cell or group of cells such as a tissue. The sample may, for example, be a colon, lung, pancreas, prostate, skin, thyroid, liver, ovary, endometrium, kidney, brain, testis, lymphatic fluid, blood, plasma, urinary bladder, or breast sample, or comprise colon, lung, pancreas, prostate, skin, thyroid, liver, ovary, endometrium, kidney, brain, testis, lymphatic fluid, blood, urinary bladder, or breast cells, groups of cells or tissue portions.

[0062] Particularly preferred are samples such as cultured or harvested or biopsied cell or tissue samples, e.g., as mentioned above, in which the RNA may be detected to reveal the qualitative nature of the RNA, i.e. that it is present, or the nucleotide sequence of the mRNA or the presence and/or identity of one or more nucleotides in the mRNA, and localization relative to other features of the cell. The sample of cells may be freshly prepared or may be prior-treated in any convenient way such as by fixation or freezing. Accordingly, fresh,

frozen or fixed cells or tissues may be used, e.g. FFPE tissue (Formalin Fixed Paraffin Embedded).

[0063] In specific embodiments, the sample of cells or tissues may be prepared, e.g. freshly prepared, or may be prior-treated in any convenient way, with the proviso that the preparation is not a preparation of fresh frozen tissues. In further specific embodiments, the sample of cells or tissues may be prepared, e.g. freshly prepared, or may be prior-treated in any convenient way, with the proviso that the preparation is not a preparation including seeding on Superfrost Plus slides. In yet additional specific embodiments, the sample of cells or tissues may be prepared, e.g. freshly prepared, or may be prior-treated in any convenient way, with the proviso that the preparation is not a preparation as disclosed in Larsson et al., Nature Methods, 2010, Vol 7 (5), pages 395-397. In very specific embodiments, the sample of cells or tissues may be prepared, e.g. freshly prepared, or may be prior-treated in any convenient way, with the proviso that the preparation is not a preparation as disclosed in section "Preparation of tissue sections" and/or "Sample pretreatment for in situ experiments" of Online methods of Larsson et al., Nature Methods, 2010, Vol 7 (5), pages 395-397.

[0064] Thus, tissue sections, treated or untreated, may be used. Alternatively a touch imprint sample of a tissue may be used. In this procedure a single layer of cells is printed onto a surface (e.g. a slide) and the morphology is similar to normal tissue sections. The touch imprint are obtained using fresh tissue sample. Other cytological preparations may be used, e.g. cells immobilized or grown on slides, or cell prepared for flow cytometry. In specific embodiments, the sample of cells or tissues may be prepared, e.g. freshly prepared, or may be prior-treated in any convenient way. In certain embodiments, the sample comprises fixed tissue. In some aspects, the sample is fixed with an alcohol, ketone, aldehyde, glycol or mixture thereof.

[0065] The sample may comprise any cell type that contains RNA including all types of mammalian and non-mammalian animal cells, plant cells, algae including blue-green algae, fungi, bacteria, protozoa, etc. Representative samples thus include clinical samples, e.g. whole blood and blood-derived products, blood cells, tissues, biopsies, as well as other samples such as cell cultures and cell suspensions, etc. In certain aspects, the sample contains, or is suspected of containing, cancer cells, such as colorectal cancer or lung cancer cells, pancreas cancer, prostate cancer, skin cancer, thyroid cancer, liver cancer, ovary cancer,

endometrium cancer, kidney cancer, cancer of the brain, testis cancer, acute non lymphocytic leukemia, myelodysplasia, urinary bladder cancer, head and neck cancer or breast cancer cells. For example, the sample may be a colon, lung, pancreas, prostate, skin, thyroid, liver, ovary, endometrium, kidney, brain, testis, lymphatic fluid, blood, plasma, urinary bladder, or breast sample suspected to be cancerous, or suspected to comprise an mRNA found in a cancer or cancerous cell, or cancerous cell group or tissue.

[0066] In some embodiments, a sample is obtained from a patient who previously was known to have cancer, which was treated or went into remission. In some cases, the patient may have a recurrent cancer. In other embodiments, the patient may have a metastasis or be suspected of having a metastasis or be at risk for metastasis. A patient at risk for cancer or metastasis may be at risk because of familial history or at determination of other genetic predispositions. In other embodiments, the patient may have been determined or may be determined to have cells exhibiting the pathology of cancer or precancer cells.

[0067] Cancer "recurrence," in pathology nomenclature, refers to cancer re-growth at the site of the primary tumor. For many cancers, such recurrence results from incomplete surgical removal or from micrometastatic lesions in neighboring blood or lymphatic vessels outside of the surgical field. Conversely, "metastasis" refers to a cancer growth distant from the site of the primary tumor. Metastasis of a cancer is believed to result from vascular and/or lymphatic permeation and spread of tumor cells from the site of the primary tumor prior to surgical removal. The prevailing clinical nomenclature used for cancer statistics is somewhat confusing in that patients who experience a second episode of a treated cancer are referred to as having undergone a "recurrence", whereas these lesions are usually temporally remote metastases at sites distant from the primary cancer. This clinical terminology will be used herein, i.e., the term "recurrence" denotes these late-arising metastatic lesions, unless specific pathologic nomenclature is needed to separate the two forms of clinical recurrence.

[0068] In certain embodiments, the sample contains pre-cancerous or premalignant cells, including but not limited to metaplasias, dysplasias, and/or hyperplasias. It may also be used to identify undesirable but benign cells, such as squamous metaplasia, dysplasia, benign prostate hyperplasia cells, and/or hyperplastic lesions.

[0069] In additional embodiments, methods and compositions are implemented with respect to a specific type of lung cancer. They may be implemented with patients diagnosed,

at risk for, or exhibiting symptoms of a specific type of lung cancer. In some embodiments, the specific type of lung cancer is non-small cell lung cancer (NSCLC) as distinguished from small cell lung cancer (SCLC). In other embodiments, the NSCLC is squamous cell carcinoma (or epidermoid carcinoma), adenocarcinoma, bronchioalveolar carcinoma, or large-cell undifferentiated carcinoma.

[0070] In certain embodiments, methods and compositions are implemented with respect to a specific type of colon cancer. They may be implemented with patients diagnosed, at risk for, or exhibiting symptoms of a specific type of colon cancer. In some embodiments, the specific type of colon cancer is an adenocarcinoma, leiomyosarcoma, colorectal lymphoma, melanoma, neuroendocrine tumors (aggressive or indolent). In the case of adenocarcinomas, the cancer may be further subtyped into mucinous or signet ring cell.

[0071] The terms "target, "target sequence", "target region", and "target nucleic acid," etc. are used synonymously herein and refer to the nucleic acid, or to a region or sequence thereof, which is to be detected or to which a reagent used in the method binds, for example the RNA to be detected, or the cDNA, or more particularly the regions thereof, to which the padlock probe is hybridized. Thus a target sequence may be within a cDNA, in which case it is to be understood that the cDNA nucleotide sequence is derived from and is complementary to the target RNA nucleotide sequence. The target may, in certain embodiments, be a single RNA molecule. In other embodiments, the target may be at least one RNA molecule, e.g. a group of 2, 3, 4, 5, 6 or more RNA molecules. These RNA molecules may differ in molecule type, and/or may differ in sequence.

[0072] The term "hybridization," as used herein, refers to the formation of a duplex structure by two single-stranded nucleic acids due to complementary base pairing. Hybridization can occur between fully complementary nucleic acid strands or between "substantially complementary" nucleic acid strands that contain minor regions of mismatch. Conditions under which hybridization of fully complementary nucleic acid strands is strongly preferred are referred to as "stringent hybridization conditions" or "sequence-specific hybridization conditions". Stable duplexes of substantially complementary sequences can be achieved under less stringent hybridization conditions; the degree of mismatch tolerated can be controlled by suitable adjustment of the hybridization conditions. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number

of variables including, for example, the length and base pair composition of the oligonucleotides, ionic strength, and incidence of mismatched base pairs, following the guidance provided by the art (see, e.g., Sambrook et al., 1989; Wetmur, 1991; Owczarzy et al., 2008, which are incorporated herein by reference). Thus the design of appropriate primers and probes, and the conditions under which they hybridize to their respective targets is well within the routine skill of the person skilled in the art.

[0073] Mutations in KRAS are common in several types of cancer. In certain embodiments, methods are provided for detecting the presence or absence of KRAS mutations in situ. In particular embodiments, the method uses padlock probe(s) configured to hybridize to cDNA(s) corresponding to one or more mutant KRAS mRNA sequences selected from the group consisting of 12AGT, 12CGT, 12TGT, 12GAT, 12GCT, 12GTT, and 13GAC (wherein the wild-type sequence is 12GGT and 13GGC) and mutants of KRAS codon 61, mutants of KRAS codon 146, and mutants of the 3' untranslated region of KRAS. In certain embodiments, the method uses padlock probe(s) configured to hybridize to cDNA(s) corresponding to the wild-type KRAS sequence. In further embodiments, the method uses padlock probe(s) configured to hybridize to cDNA(s) corresponding to one or more mutant KRAS mRNA sequences selected from the group consisting of 12AGT, 12CGT, 12TGT, 12GAT, 12GCT, 12GTT, and 13GAC (wherein the wild-type sequence is 12GGT and 13GGC) and mutants of KRAS codon 61, mutants of KRAS codon 146, and mutants of the 3' untranslated region of KRAS; and to one or more wild-type KRAS mRNA sequences selected from the group consisting of 12GGT and 13GGC, wild-type sequences of KRAS codon 61, KRAS codon 146, and of the 3' untranslated region of KRAS.

[0074] In another embodiment, methods are provided for detecting the presence or absence of mutations in mRNA that codes for HER2, cMyc, TERT, APC, Braf, PTEN, PI3K, and/or EGFR. In particular embodiments, the method uses padlock probe(s) configured to hybridize to cDNA(s) corresponding to one or more mutant HER2, cMyc, TERT, Braf, APC, PTEN and/or PI3K mRNA sequences. In further embodiments, the method uses padlock probe(s) configured to hybridize to cDNA(s) corresponding to one or more wild-type HER2, cMyc, TERT, Braf, APC, PTEN and/or PI3K mRNA sequences. In further embodiments, padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant Braf, PTEN and/or PI3K mRNA sequences, and to one or more wild-type Braf, APC, PTEN and/or PI3K mRNA sequences. Accordingly methods are provided for detecting the presence or

absence of a rolling circle amplification product corresponding to one or more of mutant and wild-type Braf, APC, PTEN and/or PI3K mRNA sequences.

[0075] In yet another group of embodiments, the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant KRAS mRNA sequences and to one or more mutant Braf mRNA sequences; or to one or more mutant KRAS mRNA sequences and to one or more mutant APC mRNA sequences; or to one or more mutant KRAS mRNA sequences and to one or more mutant PTEN mRNA sequences; or to one or more mutant KRAS mRNA sequences and to one or more mutant PI3K mRNA sequences. Embodiments accordingly provide methods for detecting the presence or absence of a rolling circle amplification product corresponding to mutant KRAS and mutant Braf mRNA sequences; or corresponding to mutant KRAS and mutant APC mRNA sequences; or corresponding to mutant KRAS and mutant PTEN mRNA sequences; or corresponding to mutant KRAS and mutant PI3K mRNA sequences.

[0076] In further embodiments, the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to wild-type KRAS and wild-type Braf mRNA sequences; or corresponding to wild-type KRAS and wild-type APC mRNA sequences; or corresponding to wild-type KRAS and wild-type PTEN mRNA sequences; or corresponding to wild-type KRAS and wild-type PI3K mRNA sequences. Accordingly methods are provided for detecting the presence or absence of a rolling circle amplification product corresponding to wild-type KRAS and Braf mRNA sequences; or corresponding to wild-type KRAS and APC mRNA sequences; or corresponding to wild-type KRAS and PTEN mRNA sequences; or corresponding to wild-type KRAS and PI3K mRNA sequences.

[0077] In a further group of embodiments, the padlock probe(s) are configured to hybridize to cDNA(s) (i) corresponding to one or more mutant KRAS mRNA sequences and to one or more mutant Braf mRNA sequences; or corresponding to one or more mutant KRAS mRNA sequences and to one or more mutant APC mRNA sequences; or corresponding to one or more mutant KRAS mRNA sequences and to one or more mutant PTEN mRNA sequences; or corresponding to one or more mutant KRAS mRNA sequences and to one or more mutant PI3K mRNA sequences; and (ii) corresponding to wild-type KRAS and Braf mRNA sequences; or corresponding to wild-type KRAS and APC mRNA sequences; or corresponding to wild-type KRAS and PTEN mRNA sequences; or corresponding to wild-

type KRAS and PI3K mRNA sequences. Methods are provided for detecting the presence or absence of a rolling circle amplification product corresponding to one or more mutant and wild-type K*RAS* and *Braf* mRNA sequences; or corresponding to one or more mutant and wild-type *KRAS* and *APC* mRNA sequences; or corresponding to one or more mutant and wild-type *KRAS* and *PTEN* mRNA sequences; or corresponding to one or more mutant and wild-type *KRAS* and *PI3K* mRNA sequences.

[0078] One embodiment provides a collection of padlock probes specific for mutations to the *KRAS* gene, comprising:

(a) Y1-X1-Z1-A

(b) Y1-X1-Z1-T

(c) Y1-X1-Z1-C

(d) Y2-X1-Z2-A

(e) Y2-X1-Z2-T

(f) Y2-X1-Z2-C, and

(g) Y3-X1-Z3-A;

where:

X1 is from 5-50 nucleotides;

Y1+Z1= 20 to 40 nucleotides;

Y2+Z2= 20 to 40 nucleotides;

Y3+Z3= 20 to 40 nucleotides;

Y1 is GTGGCGTAGGCAAGA (SEQ ID NO:1), GTGGCGTAGGCAAG (SEQ ID NO:2), GTGGCGTAGGCAA (SEQ ID NO:3), GTGGCGTAGGCA (SEQ ID NO:4), GTGGCGTAGGC (SEQ ID NO:5), GTGGCGTAGG (SEQ ID NO:6), GTGGCGTAG, GTGGCGTA, GTGGCGT, GTGGCG, GTGGC, GTGG, GTG, GT, G;

Y2 is TGGCGTAGGCAAGAG (SEQ ID NO:7), TGGCGTAGGCAAGA (SEQ ID NO:8), TGGCGTAGGCAAG (SEQ ID NO:9), TGGCGTAGGCAA (SEQ ID NO:10), TGGCGTAGGCA (SEQ ID NO:11), TGGCGTAGGC (SEQ ID NO:12), TGGCGTAGG, TGGCGTAG, TGGCGTA, TGGCGT, TGGCG, TGGC, TGG, TG, T;

Y3 is TGGCGTAGGCAAGAGTGC (SEQ ID NO:13), TGGCGTAGGCAAGAGTG (SEQ ID NO:14), TGGCGTAGGCAAGAGT (SEQ ID NO:15), TGGCGTAGGCAAGAG (SEQ ID NO:7), TGGCGTAGGCAAGA (SEQ ID NO:8), TGGCGTAGGCAAG (SEQ ID NO:9), TGGCGTAGGCAA (SEQ ID NO:10), TGGCGTAGGCA (SEQ ID NO:11),

TGGCGTAGGC (SEQ ID NO:12), TGGCGTAGG, TGGCGTAG, TGGCGTA, TGGCGT, TGGCG, TGGC, TGG, TG, T;

Z1 is TGGTAGTTGGAGCT (SEQ ID NO:27), GGTAGTTGGAGCT (SEQ ID NO:28), GTAGTTGGAGCT (SEQ ID NO:29), TAGTTGGAGCT (SEQ ID NO:30), AGTTGGAGCT (SEQ ID NO:31), GTTGGAGCT, TTGGAGCT, TGGAGCT, GGAGCT, GAGCT, AGCT, GCT, CT, T, or a bond;

Z2 is GGTAGTTGGAGCTG (SEQ ID NO:16), GTAGTTGGAGCTG (SEQ ID NO:17), TAGTTGGAGCTG (SEQ ID NO:18), AGTTGGAGCTG (SEQ ID NO:19), GTTGGAGCTG (SEQ ID NO:20), TTGGAGCTG, TGGAGCTG, GGAGCTG, GAGCTG, AGCTG, GCTG, CTG, TG, G or a bond; and

Z3 is AGTTGGAGCTGGTG (SEQ ID NO:21), GTTGGAGCTGGTG (SEQ ID NO:22), TTGGAGCTGGTG (SEQ ID NO:23), TGGAGCTGGTG (SEQ ID NO:24), GGAGCTGGTG (SEQ ID NO:25), GAGCTGGTG (SEQ ID NO:26), AGCTGGTG, GCTGGTG, CTGGTG, TGGTG, GGTG, GGTG, GTG, TG, G or a bond.

[0079] In some embodiments, the collection of *KRAS* probes, further comprises:

(h) Y1-X2-Z1-G

(i) Y2-X2-Z2-G

(j) Y3-X2-Z3-G

where X2 is from 10-50 nucleotides and differs from X1.

[0080] In a specific embodiment, the collection of *KRAS* probes, further comprises:

(h) Y1-X2-Z1-G

(i) Y2-X2-Z2-G

(j) Y3-X2-Z3-G

where X2 is from 10-50 nucleotides and differs from X1.

[0081] Further embodiments provide a collection of padlock probes specific for mutations to the *Braf* gene comprising:

(k) Y1-X1-Z1-A

where:

X1 is from 5-50 nucleotides;

Y1+Z1= 20 to 40 nucleotides;

Y1 is GAAATCTCGATGGAG (SEQ ID NO:102), AAATCTCGATGGAG (SEQ ID NO:103), AATCTCGATGGAG (SEQ ID NO:104), ATCTCGATGGAG (SEQ ID NO:105), TCTCGATGGAG (SEQ ID NO:106), CTCGATGGAG (SEQ ID NO:107), TCGATGGAG, CGATGGAG, GATGGAG, ATGGAG, TGGAG, GGAG, GAG, AG, G; and

Z1 is TGGTCTAGCTACAG (SEQ ID NO:108), GGTCTAGCTACAG (SEQ ID NO:109), GTCTAGCTACAG (SEQ ID NO:110), TCTAGCTACAG (SEQ ID NO:111), CTAGCTACAG (SEQ ID NO:112), TAGCTACAG, AGCTACAG , GCTACAG, CTACAG, TACAG, ACAG, CAG, AG, G, or a bond.

[0082] In some embodiments, the collection of *Braf* probes further comprises:

(l) Y1-X2-Z1-T

where X2 is from 10-50 nucleotides.

[0083] In a specific embodiment, the collection of *Braf* probes further comprises:

(l) Y1-X2-Z1-T

where X2 is from 10-50 nucleotides and differs from X1.

[0084] Further embodiments provide a collection of padlock probes specific for mutations to the *APC* gene, *PTEN* gene, *PI3K* gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR, comprising:

(m) Y1-X1-Z1-W

where:

X1 is from 5-50 nucleotides;

Y1+Z1= 20 to 40 nucleotides;

wherein Y1 comprises 5-20 nucleotides 3' to a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR;

wherein Z1 comprises 5-20 nucleotides in the 5' to a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR; and

wherein W is a nucleotide complementary to a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR.

[0085] In some embodiments, the collection of probes specific for mutations to the *APC* gene, *PTEN* gene, *PI3K* gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR, further comprises:

(n) Y1-X2-Z1-V

where X2 is from 10-50 nucleotides; and

wherein V is a nucleotide complementary to a wildtype sequence at the site of a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR.

[0086] In specific embodiments, the collection of probes specific for mutations to the *APC* gene, *PTEN* gene, *PI3K* gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR, further comprises:

(n) Y1-X2-Z1-V

where X2 is from 10-50 nucleotides and differs from X1; and

wherein V is a nucleotide complementary to a wildtype sequence at the site of a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR.

[0087] In some embodiments, X1 is from 25-50 nucleotides. In certain embodiments, X1 comprises at least one labeled nucleotide. In some embodiments, each probe (a)-(g) has the same X1. In some embodiments, each probe selected from (a)-(g), (k) and (m) has the same X2.

[0088] In certain aspects, each of Y1+Z1, Y2+Z2 and Y3+Z3 is at least 25 nucleotides.

[0089] In certain aspects, each probe in the collection of probes has a GC content of at least 40%.

[0090] Some embodiments provide a collection of padlock probes specific for mutations to the KRAS gene, specific for mutations to the Braf gene, specific for mutations to the APC gene, PTEN gene, PI3K gene, KRAS gene codon 61 or codon 146, or KRAS gene 3'UTR, and optionally collections of padlock probes specific for corresponding wild-type sequences, e.g. as defined above, the collection being capable of detecting a plurality of mutations in (i) the KRAS gene, (ii) the KRAS gene and the Braf gene, (iii) the KRAS gene and the APC

gene, (iv) the KRAS gene and the PTEN gene, or (v) the KRAS gene and the PI3K gene, wherein the plurality of mutations constitute at least 40% of KRAS mutations associated with cancer.

[0091] Additional embodiments provide a collection of padlock probes specific for mutations to the KRAS gene, specific for mutations to the Braf gene, specific for mutations to the APC gene, PTEN gene, PI3K gene, KRAS gene codon 61 or codon 146, or KRAS gene 3'UTR, and optionally collections of padlock probes specific for corresponding wild-type sequences, e.g. as defined above, wherein the detection of mutations to (i) the KRAS gene, (ii) the KRAS gene and the Braf gene, (iii) the KRAS gene and the APC gene, (iv) the KRAS gene and the PTEN gene, or (v) the KRAS gene and the PI3K gene allows to determine the presence of cancer or a predisposition for cancer.

[0092] In a specific embodiment the cancer or predisposition for cancer is determined in at least or in at most 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95% (or any range derivable therein) of patients bearing a KRAS-mutant associated with tumor development.

[0093] Further embodiments provide the use of a collection of padlock probes specific for mutations to the KRAS gene, specific for mutations to the Braf gene, specific for mutations to the APC gene, PTEN gene, PI3K gene, KRAS gene codon 61 or codon 146, or KRAS gene 3'UTR, and optionally collections of padlock probes specific for corresponding wild-type sequences, e.g. as defined above, for the determination of the presence or absence of a KRAS-mutant tumor or for the determination of a predisposition for a KRAS-mutant tumor in a patient or group of patients.

[0094] In specific embodiments, the determination of the presence or absence of a KRAS-mutant tumor or for the determination of a predisposition for a KRAS-mutant tumor in a patient or group of patients allows to determine the presence of cancer in at least or in at most 5%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95% (or any range derivable therein) of a patient group bearing a KRAS-mutant associated with tumor development.

[0095] A "patient bearing a KRAS-mutant associated with tumor development" or a "patient group bearing a KRAS-mutant associated with tumor development" refers to an

individual or group of individuals, wherein each patient or group member comprises at least one mutation in the KRAS gene (or a corresponding mutant), that has been described in the scientific literature or is known to the skilled person as being associated with tumor development, e.g., associated with preforms of tumors or predispositions for tumors, associated with different tumor development stages, or associated with full grown tumors or cancer. In specific embodiments, these mutations or mutants comprise mutations as can be derived from the Sanger database as of August 22, 2012 being associated with cancer or precancer (on the world wide web at sanger.ac.uk).

[0096] In specific embodiments, the patient group, i.e. each member of the patient group, may bear a KRAS-mutant associated with tumor development and an additional mutation in the Braf gene, and/or the APC gene, and/or PTEN gene, and/or the PI3K gene. These combinations of mutations may contribute to tumor development associated with KRAS mutations; or they may constitute mutational combinations associated with cancer or precancer forms, or predispositions for cancer. In further specific embodiments, the patient group, i.e., each member of the patient group, may bear a mutation in the Braf gene, and/or the APC gene, and/or PTEN gene, and/or the PI3K gene. These mutations are associated with cancer or precancer, or predispositions for cancer as can be derived from the Sanger database (on the world wide web at sanger.ac.uk). In further specific embodiments, the patient group, i.e. each member of the patient group, may bear a mutation in the EGFR gene, and/or the KRAS gene, and/or the Braf gene, and/or the APC gene, and/or PTEN gene, and/or the PI3K gene. These mutations are associated with cancer or precancer, or predisposition for cancer, as can be derived from the Sanger database (on the world wide web at sanger.ac.uk). Furthermore, examples of EGFR mutations that may be detected according to various embodiments, or that may be employed in the context of compositions described herein are shown in Table 7.

[0097] Methods also concern detecting or localizing a RNA transcript encoded by a gene listed in Table B. In certain embodiments, a padlock probe has a sequence that is identical or complementary to a mutation in the gene associated with cancer (which means the mutation has been correlated in a statistically significant way with the presence of cancer, pre-cancer, and/or risk of cancer). The mutation may or may not cause cancer. This padlock probe can be used, in some embodiments, in conjunction with a padlock probe that is identical or complementary to the wild-type version of the gene in order to detect a cancer mutation in the

gene using methods described herein. In some embodiments, the mutation is a point mutation, frame shift, substitution, deletion, insertion, translocation, inversion, amplification, indel, or a combination thereof. The mutation may constitute a mutation of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides.

[0098] In certain embodiments, the cancer is colorectal cancer, lung cancer, pancreas cancer, prostate cancer, skin cancer, thyroid cancer, liver cancer, ovary cancer, endometrium cancer, kidney cancer, cancer of the brain, testis cancer, acute non lymphocytic leukemia, myelodysplasia, urinary bladder cancer, head and neck cancer or breast cancer. In further embodiments, the predispositions to cancer are predispositions to colorectal cancer, lung cancer, pancreas cancer, prostate cancer, skin cancer, thyroid cancer, liver cancer, ovary cancer, endometrium cancer, kidney cancer, cancer of the brain, testis cancer, acute non lymphocytic leukemia, myelodysplasia, urinary bladder cancer, head and neck cancer or breast cancer.

[0099] In further embodiments the colorectal cancer is metastatic colorectal cancer, adenocarcinoma, leiomyosarcoma, colorectal lymphoma, melanoma or neuroendocrine tumor. In other embodiments, the lung cancer is a non-small cell lung cancer (NSCLC), or small cell lung cancer (SCLC).

[00100] Also provided are collections of padlock probes specific for mutations to the KRAS gene, specific for mutations to the Braf gene, specific for mutations to the APC gene, PTEN gene, PI3K gene, KRAS gene codon 61 or codon 146, or KRAS gene 3'UTR, and optionally collections of padlock probes specific for corresponding wild-type sequences, e.g. as defined above, or uses thereof, e.g. as defined above, allowing to determine

(i) the presence of colorectal cancer in at least 25 to 60 % of a patient group bearing a KRAS-mutant associated with tumor development;

(ii) the presence of lung cancer in at least 25 to 60 % of a patient group bearing a KRAS-mutant associated with tumor development;

(iii) the presence of pancreas cancer in at least 80 to 90 % of a patient group bearing a KRAS-mutant associated with tumor development;

(iv) the presence of prostate cancer in at least 5 to 25 % of a patient group bearing a KRAS-mutant associated with tumor development;

(v) the presence of skin cancer in at least 5 to 25 % of a patient group bearing a KRAS-mutant associated with tumor development;

(vi) the presence of thyroid cancer in at least 5 to 60 % of a patient group bearing a KRAS-mutant associated with tumor development;

(vii) the presence of liver cancer in at least 10 to 25 % of a patient group bearing a KRAS-mutant associated with tumor development;

(viii) the presence of ovary cancer in at least 5 to 50 % of a patient group bearing a KRAS-mutant associated with tumor development;

(ix) the presence of endometrium cancer in at least 10 to 40 % of a patient group bearing a KRAS-mutant associated with tumor development;

(x) the presence of kidney cancer in at least 5 to 50 % of a patient group bearing a KRAS-mutant associated with tumor development;

(xi) the presence of cancer of the brain in at least 5 to 15 % of a patient group bearing a KRAS-mutant associated with tumor development;

(xii) the presence of testis cancer in at least 10 to 45 % of a patient group bearing a KRAS-mutant associated with tumor development;

(xiii) the presence of acute non lymphocytic leukemia in at least 5 to 15 % of a patient group bearing a KRAS-mutant associated with tumor development;

(xiv) the presence of urinary bladder cancer in at least 5 % of a patient group bearing a KRAS-mutant associated with tumor development;

(xv) the presence of head and neck cancer in at least 5 to 10 % of a patient group bearing a KRAS-mutant associated with tumor development; or

(xvi) the presence of breast cancer in at least 5 to 10 % of a patient group bearing a KRAS-mutant associated with tumor development.

[00101] In some embodiments, the above-mentioned collections of probes are provided in a kit along with one or more of the following:

(ii) an reverse transcriptase primer comprising one or more locked nucleic acid and capable of hybridizing to the target RNA;

(iii) a reverse transcriptase;

(iv) a ribonuclease;

(v) a ligase;

(vi) a polymerase having 3′ exonuclease activity;

(vii) a detection probe capable of hybridizing to a complement of the padlock probe; or

(ix) nucleotides.

[00102] In further embodiments, methods are provided for localized in situ detection of mRNA which codes for one or more mutations of the KRAS gene in a sample of cells on a slide surface, comprising:

(a) generating cDNA from mRNA in the sample, wherein the primer is provided with a functional moiety capable of binding to or reacting with a cell or cellular component or an affinity binding group capable of binding to a cell or cellular component;

(b) adding a ribonuclease to the sample to digest the mRNA hybridized to the cDNA;

(c) contacting the sample with one or more padlock probes specific for mutations to the KRAS gene, wherein each padlock probe comprises a sequence selected from the collection of padlock probes specific for mutations to the KRAS gene, specific for mutations to the Braf gene, specific for mutations to the APC gene, PTEN gene, PI3K gene, KRAS gene codon 61 or codon 146, or KRAS gene 3'UTR, and optionally collections of padlock probes specific for corresponding wild-type sequences, e.g. as defined above.

[00103] In one embodiment, there are methods for localized in situ detection of mRNA which codes for one or more mutations of the KRAS gene in a sample of cells on a slide surface, comprising: (a) generating cDNA from mRNA in the sample, wherein the primer is provided with a functional moiety capable of binding to or reacting with a cell or cellular component or an affinity binding group capable of binding to a cell or cellular component; (b) adding a ribonuclease to the sample to digest the mRNA hybridized to the cDNA; (c) contacting the sample with one or more padlock probes specific for mutations to the *KRAS*

gene, wherein each padlock probe comprises a sequence selected from the group consisting of:

(a) Y1-X1-Z1-A

(b) Y1-X1-Z1-T

(c) Y1-X1-Z1-C

(d) Y2-X1-Z2-A

(e) Y2-X1-Z2-T

(f) Y2-X1-Z2-C, and

(g) Y3-X1-Z3-A;

where:

X1 is from 5-50 nucleotides;

Y1+Z1= 20 to 40 nucleotides;

Y2+Z2= 20 to 40 nucleotides;

Y3+Z3= 20 to 40 nucleotides;

Y1 is GTGGCGTAGGCAAGA (SEQ ID NO:1), GTGGCGTAGGCAAG (SEQ ID NO:2), GTGGCGTAGGCAA (SEQ ID NO:3), GTGGCGTAGGCA (SEQ ID NO:4), GTGGCGTAGGC (SEQ ID NO:5), GTGGCGTAGG (SEQ ID NO:6), GTGGCGTAG, GTGGCGTA, GTGGCGT, GTGGCG, GTGGC, GTGG, GTG, GT, G;

Y2 is TGGCGTAGGCAAGAG (SEQ ID NO:7), TGGCGTAGGCAAGA (SEQ ID NO:8), TGGCGTAGGCAAG (SEQ ID NO:9), TGGCGTAGGCAA (SEQ ID NO:10), TGGCGTAGGCA (SEQ ID NO:11), TGGCGTAGGC (SEQ ID NO:12), TGGCGTAGG, TGGCGTAG, TGGCGTA, TGGCGT, TGGCG, TGGC, TGG, TG, T;

Y3 is TGGCGTAGGCAAGAGTGC (SEQ ID NO:13), TGGCGTAGGCAAGAGTG (SEQ ID NO:14), TGGCGTAGGCAAGAGT (SEQ ID NO:15), TGGCGTAGGCAAGAG (SEQ ID NO:7), TGGCGTAGGCAAGA (SEQ ID NO:8), TGGCGTAGGCAAG (SEQ ID NO:9), TGGCGTAGGCAA (SEQ ID NO:10), TGGCGTAGGCA (SEQ ID NO:11), TGGCGTAGGC (SEQ ID NO:12), TGGCGTAGG, TGGCGTAG, TGGCGTA, TGGCGT, TGGCG, TGGC, TGG, TG, T;

Z1 is TGGTAGTTGGAGCT (SEQ ID NO:27), GGTAGTTGGAGCT (SEQ ID NO:28), GTAGTTGGAGCT (SEQ ID NO:29), TAGTTGGAGCT (SEQ ID NO:30), AGTTGGAGCT (SEQ ID NO:31), GTTGGAGCT, TTGGAGCT, TGGAGCT, GGAGCT, GAGCT, AGCT, GCT, CT, T, or a bond;

Z2 is GGTAGTTGGAGCTG (SEQ ID NO:16), GTAGTTGGAGCTG (SEQ ID NO:17), TAGTTGGAGCTG (SEQ ID NO:18), AGTTGGAGCTG (SEQ ID NO:19), GTTGGAGCTG (SEQ ID NO:20), TTGGAGCTG, TGGAGCTG, GGAGCTG, GAGCTG, AGCTG, GCTG, CTG, TG, G or a bond; and

Z3 is AGTTGGAGCTGGTG (SEQ ID NO:21), GTTGGAGCTGGTG (SEQ ID NO:22), TTGGAGCTGGTG (SEQ ID NO:23), TGGAGCTGGTG (SEQ ID NO:24), GGAGCTGGTG (SEQ ID NO:25), GAGCTGGTG (SEQ ID NO:26), AGCTGGTG, GCTGGTG, CTGGTG, TGGTG, GGTG, GGTG, GTG, TG, G or a bond;

(d) ligating, directly or indirectly, the ends of the padlock probe(s);

(e) subjecting the circularized padlock probe(s) to rolling circle amplification (RCA) using a DNA polymerase having 3′-5′ exonuclease activity wherein if necessary the exonuclease activity digests the cDNA to generate a free 3′ end which acts as a primer for the RCA; and

(f) detecting the rolling circle amplification product(s).

[00104] In some embodiments of the method, step (c) further comprises contacting the sample with padlock probes (h), (i) and (j), wherein each is specific for wild-type *KRAS* gene and have sequences:

(h) Y1-X2-Z1-G

(i) Y2-X2-Z2-G, and

(j) Y3-X2-Z3-G

where X2 is from 10-50 nucleotides.

[00105] In specific embodiments of the method, step (c) further comprises contacting the sample with padlock probes (h), (i) and (j), wherein each is specific for wild-type *KRAS* gene and have sequences:

(h) Y1-X2-Z1-G

(i) Y2-X2-Z2-G, and

(j) Y3-X2-Z3-G

where X2 is from 10-50 nucleotides and differs from X1.

[00106] In a further embodiment, there are methods for localized *in situ* detection of mRNA which codes for one or more mutations of the *Braf* gene in a sample of cells on a slide surface, comprising:

(a) generating cDNA from mRNA in the sample, wherein the primer is provided with a functional moiety capable of binding to or reacting with a cell or cellular component or an affinity binding group capable of binding to a cell or cellular component;

(b) adding a ribonuclease to the sample to digest the mRNA hybridized to the cDNA;

(c) contacting the sample with one or more padlock probes specific for mutations to the *Braf* gene, wherein each padlock probe comprises a sequence selected from the group consisting of:

(*k*) Y1-X1-Z1-A

where:

X1 is from 5-50 nucleotides;

Y1+Z1= 20 to 40 nucleotides;

Y1 is GAAATCTCGATGGAG (SEQ ID NO:102), AAATCTCGATGGAG (SEQ ID NO:103), AATCTCGATGGAG (SEQ ID NO:104), ATCTCGATGGAG (SEQ ID NO:105), TCTCGATGGAG (SEQ ID NO:106), CTCGATGGAG (SEQ ID NO:107), TCGATGGAG, CGATGGAG, GATGGAG, ATGGAG, TGGAG, GGAG, GAG, AG, G; and

Z1 is TGGTCTAGCTACAG (SEQ ID NO:108), GGTCTAGCTACAG (SEQ ID NO:109), GTCTAGCTACAG (SEQ ID NO:110), TCTAGCTACAG (SEQ ID NO:111), CTAGCTACAG (SEQ ID NO:112), TAGCTACAG, AGCTACAG , GCTACAG, CTACAG, TACAG, ACAG, CAG, AG, G, or a bond;

(d) ligating, directly or indirectly, the ends of the padlock probe(s);

(e) subjecting the circularized padlock probe(s) to rolling circle amplification (RCA) using a DNA polymerase having 3'-5' exonuclease activity wherein if necessary the exonuclease activity digests the cDNA to generate a free 3' end which acts as a primer for the RCA; and

(f) detecting the rolling circle amplification product(s).

[00107] In some embodiments of the method, step (c) further comprises contacting the sample with padlock probes (*l*), wherein each is specific for wild-type *Braf* gene and have sequences:

(*l*) Y1-X2-Z1-T

where X2 is from 10-50 nucleotides.

[00108] In specific embodiments of the method, step (c) further comprises contacting the sample with padlock probes (l), wherein each is specific for wild-type *Braf* gene and have sequences:

(*l*) Y1-X2-Z1-T

where X2 is from 10-50 nucleotides and differs from X1.

[00109] In a further embodiment, methods are provided for localized *in situ* detection of mRNA which codes for one or more mutations of the *APC* gene, *PTEN* gene, *PI3K* gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR in a sample of cells on a slide surface, comprising:

(a) generating cDNA from mRNA in the sample, wherein the primer is provided with a functional moiety capable of binding to or reacting with a cell or cellular component or an affinity binding group capable of binding to a cell or cellular component;

(b) adding a ribonuclease to the sample to digest the mRNA hybridized to the cDNA;

(c) contacting the sample with one or more padlock probes specific for mutations to the *APC* gene, *PTEN* gene, *PI3K* gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR, wherein each padlock probe comprises a sequence selected from the group consisting of:

(*m*) Y1-X1-Z1-W

where:

X1 is from 5-50 nucleotides;

Y1+Z1= 20 to 40 nucleotides;

wherein Y1 comprises 5-20 nucleotides 3' to a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR;

wherein Z1 comprises 5-20 nucleotides in the 5' to a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR; and

wherein W is a nucleotide complementary to a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR.

[00110] In some embodiments of the method step (c) further comprises contacting the sample with padlock probes (*n*), wherein each is specific for wild-type *APC* gene, *PTEN* gene, *PI3K* gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR and have sequences:

(*n*) Y1-X2-Z1-V

where X2 is from 10-50 nucleotides; and

wherein V is a nucleotide complementary to a wildtype sequence at the site of a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR.

[00111] In specific embodiments of the method step (c) further comprises contacting the sample with padlock probes (*n*), wherein each is specific for wild-type *APC* gene, *PTEN* gene, *PI3K* gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR and have sequences:

(*n*) Y1-X2-Z1-V

where X2 is from 10-50 nucleotides and differs from X1; and

wherein V is a nucleotide complementary to a wildtype sequence at the site of a point mutation in the APC gene, PTEN gene, PI3K gene, *KRAS* gene codon 61 or codon 146, or *KRAS* gene 3'UTR.

[00112] In some embodiments, X1 and X2 each comprise at least one labeled nucleotide. In certain aspects, the label is fluorophore or a chromophore. In certain embodiments, each probe selected from (a)-(g), (k) and (m) has the same X1. In certain embodiments, each probe selected from (h)-(j), (l) and (n) has the same X2.

[00113] In certain embodiments of the method, the primer comprises 2'O-Me RNA, methylphosphonates or 2' Fluor RNA bases, peptidyl nucleic acid residues, or locked nucleic acid residues. In additional embodiments, a primer is modified with biotin, an amine group, a lower alkylamine group, an acetyl group, DMTO, fluoroscein, a thiol group, or acridine. In some embodiments, the sample comprises a fixed tissue section, a fresh frozen tissue, touch imprint samples or a cytological preparation comprising one or more cells.

[00114] It is contemplated that any method or composition described herein can be implemented with respect to any other method or composition described herein.

[00115] The terms "comprise" (and any form of comprise, such as "comprises" and "comprising"), "have" (and any form of have, such as "has" and "having"), "contain" (and any form of contain, such as "contains" and "containing"), and "include" (and any form of include, such as "includes" and "including") are open-ended linking verbs. As a result, a method, composition, kit, or system that "comprises," "has," "contains," or "includes" one or more recited steps or elements possesses those recited steps or elements, but is not limited to possessing only those steps or elements; it may possess (*i.e.*, cover) elements or steps that are not recited. Likewise, an element of a method, composition, kit, or system that "comprises," "has," "contains," or "includes" one or more recited features possesses those features, but is not limited to possessing only those features; it may possess features that are not recited.

[00116] Any embodiment of any of the present methods, composition, kit, and systems may consist of or consist essentially of—rather than comprise/include/contain/have—the described steps and/or features. Thus, in any of the claims, the term "consisting of" or "consisting essentially of" may be substituted for any of the open-ended linking verbs recited above, in order to change the scope of a given claim from what it would otherwise be using the open-ended linking verb.

[00117] The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and "and/or."

[00118] Throughout this application, the term "about" is used to indicate that a value includes the standard deviation of error for the device or method being employed to determine the value.

[00119] Following long-standing patent law, the words "a" and "an," when used in conjunction with the word "comprising" in the claims or specification, denotes one or more, unless specifically noted.

[00120] Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[00121] The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

[00122] **FIG. 1**: Schematic representation of the detection of individual transcripts *in situ* with padlock probes and target-primed RCA. cDNA is created using locked nucleic acid

(LNA)-modified primers and is probed after degradation of mRNA by RNase H. RCPs are identified through hybridization of fluorescent detection probes.

[00123] FIG. 2a-d: Multiplex *in situ* detection of cancer-related transcripts in cancer and primary human cell lines. Quantification of RCPs in the different cell lines is shown in the bar graph: **(a)** human ovarian carcinoma cells (SKOV3); **(b)** human breast carcinoma cells (SKBR3); **(c)** *TERT* immortalized human fibroblast cells (BJhTERT); and **(d)** primary human fibroblast culture GM08402.

[00124] FIG. 3: Effect of LNA base incorporation in the primer for cDNA synthesis *in situ*. cDNA primers with different LNA substitutions were compared against an unmodified primer consisting of only DNA bases (No mod) for cDNA synthesis *in situ*. Synthesized cDNA was detected with padlock probes and target-primed RCA and quantified by counting RCPs/cell. The investigated primers had five, seven or nine LNA bases positioned either at every second or every third position in the 5′-end of the primers. Primers had a total length of 25 nt or 30 nt (indicated in parentheses).

[00125] FIG. 4: Investigation of cDNA synthesis length. Primers positioned at different distances from the mRNA 5′-end, where the target site of the PLP-βe1 padlock probe is located, were compared for *in situ* detection of β-actin transcripts with padlock probes and target-primed RCA to investigate the efficiency of the cDNA synthesis. When reverse transcription was carried out without addition of any primer, an average of seven RCPs were detected per cell (not shown in diagram).

[00126] FIG. 5: Detection of individual β-actin transcripts in cultured human fibroblasts. Target sites in exons 1 and 6 on the β-actin transcript were probed in GM08402 cells. A negative control was performed without addition of reverse transcriptase.

[00127] FIG. 6a-b: Quantification of RCPs in single cultured cells. Histogram showing quantification of **(a)** β-actin RCPs in 134 cells of a GM08402 culture and **(b)** *KRAS* RCPs in 77 cells of an A-427 culture.

[00128] FIG. 7: *In situ* genotyping of *KRAS* codon 12 mutations in cell lines with padlock probes and RCA. Quantification of the number of RCPs/cell detected *in situ* in the heterozygous cell line A-427, showing the allelic expression of wild type (light grey) and

mutated (dark grey) *KRAS*-RCPs in single cells. Inset represents the overall allelic ratio from 77 counted cells.

[00129] FIG. 8: Schematic overview for *in situ* genotyping with padlock probes and target-primed RCA. *KRAS* cDNA (black) is created by reverse transcription with an LNA-primer. Target mRNA (grey) is degraded by RNase H, except for the region that is hybridized to the LNA-part of the primer that is protected from degradation, anchoring the created cDNA to the target. *KRAS* genotype specific padlock probes, with similar target sites except for the single point mutated base (GGT→AGT), are hybridized to the cDNA and circularized by target-dependent ligation. The targeted *KRAS* transcripts act as primer for RCA and the resulting RCPs are labeled with fluorescence-labeled detection probes and visualized as bright spots in the cells or tissue.

[00130] FIG. 9: Example of padlock probes for a *Braf* mutant and wild-type sequence.

[00131] FIG. 10a-g: Schematic illustration of in situ sequencing. **(a)** cDNA is synthesized by using locked nucleic acid (LNA)-modified primers. **(b)** The mRNA is degraded by RNase H, **(c)** followed by hybridization of a padlock probe, which is designed such that a gap between the two ends is formed after the hybridization. **(d)** The gap, which is the target for sequencing, is then filled by DNA polymerization. The 3' end and 5' end are then joined by DNA ligase to form a completed DNA circle. **(e)** Target primer RCA is performed to amplify the DNA circle and generate RCA products, which are then subjected for sequencing by ligation. **(f)** The anchor primer is hybridized right next to the target on the 3' end, followed by ligation of interrogation probes. The mismatched interrogation probes will not be ligated and therefore washed away. **(g)** Finally, the RCA product will show the color that represents the right base.

[00132] FIG. 11a-c: *In situ* sequencing of *ACTB* in co-culture of human and mouse cells. **(a)** Raw data from four cycles, color channels merged. **(b)** Two different sequences are detected by automated image analysis, and locations marked directly in the image. **(c)** Each sequence is related to a specific cell, and the number of occurrences is quantified and represented by a pie-chart, where the total area is proportional to the number of RCPs.

[00133] FIG. 12a-d: *In situ* sequencing of *ACTB* and *HER2* mRNA in breast cancer tissue. **(a)** Raw image showing the location of sequences called from a fresh frozen breast cancer tissue section (blue: DAPI, red: general stain of sequence common to all probes). **(b)** Each diamond represents a decoded sequence, color coded as shown in **(c)**. The white line was manually drawn to separate cancer cells from adjacent non-malignant stroma. **(d)** The relative frequency of each sequence is quantified in normal and cancer tissue, and represented by a pie-chart (the number in parentheses is the number of occurrences, and the total area is proportional to the total number of RCPs). The two most abundant sequences are that from *ACTB* (light blue) and *HER2* (red) transcripts. Note that other sequences differ with as little as a single nucleotide and occur only once.

[00134] FIG. 13: Raw data images from each sequencing cycles of co-cultured mouse and human cells shown in Figure 2. From top to bottom is cycle 1 to cycle 4. Signal from each cycle is displayed in four channels, FITC for base T, Cy3 for base G, Cy3. 5 for base C and Cy5 for base A.

[00135] FIG. 14a-d: Analyzing data and sequencing four bases of co-cultured mouse and human cells shown in Figure 2 and Figure 13. **(a)** cell nuclei are detected based on size and shape. **(b)** Cells are defined by watershed segmentation seeded by the nuclei detected in (a). **(c)** Individual RCPs are detected using the general stain in the first imaging cycle and used as a template for base calling at each subsequent image. **(d)** The output from the CellProfiler pipeline is processed using a Matlab script decoding the sequence of individual RCPs and calculating their distribution per single cells.

[00136] FIG. 15a-d: Sequencing four bases of ACTB in ten human fibroblast cells. **(a)** Raw data showing cell nuclei (blue) and RCPs (general stain, red); **(b)** Detected RCPs and their sequence overlaid the original image. **(c)** Distribution of sequences for the combined cell population. **(d)** Raw data from each sequencing cycles. From top to bottom is cycle 1 to cycle 4. Signal from each cycle is displayed in four channels, FITC for base T, Cy3 for base G, Cy3.5 for base C and Cy5 for base A.

[00137] FIG. 16a-f: We compared the localization and intensity of RCPs before sequencing **(a)** and after four cycles of sequencing **(b)**. The degree of co-localization of RCPs is obvious when comparing before alignment **(c)**, and after image filtering and alignment **(d)** (red=before sequencing, green=after four cycles, yellow=colocalized). There is a significant

horizontal shift in **(c)** that has been corrected in **(d)**. Defining RCPs using the combined intensity from the first and the fourth cycle and a fixed intensity threshold (0.2) **(e)** we compare the maximum intensity of the general stain within each RCP before cycle 1 and after cycle 4 **(f)**. Signal intensities do not decrease substantially after four cycles of sequencing, and even increase for ~30% of the RCPs. However, the background fluorescence increases with the number of cycles, introducing false RCPs (marked with a red dot). These false RCPs are excluded from the analysis as we require true RCPs to be present at cycle 1. Debris fluorescing in all color channels will be excluded by a quality threshold during sequence decoding. The increased background fluorescence may cause the weakest signals to be lost in the background.

[00138] FIG. 17a-b: Sequencing of two stretches of four bases in ACTB and HER2 transcripts in a fresh frozen breast cancer tissue section. **(a)** Acquired raw data images from each sequencing cycles are displayed from top to bottom. From left to right, the images from each sequencing cycle are displayed in the following order: DAPI staining for nuclei, FITC channel for base T, Cy3 channel for base G, Cy3.5 for base C and Cy5 for base A. **(b)** Images showing individual channels from each sequencing cycles after performing enhancement of signal and suppression of background in CellProfiler. The displayed order is the same as in **(a)**. The obtained images were aligned with the image of signal from the general detection oligonucleotide binding to all RCPs. Image analysis was performed as described for cultured cells, only omitting detection of individual cells.

[00139] FIG. 18a-b: (a) Called sequences from in situ sequencing of KRAS transcripts in six co-cultured cell lines: the wild type cell line ONCO-DG-1, the heterozygous mutant cell line A-427 (12GAT), the heterozygous mutant cell line SW-480 (12GTT), the heterozygous mutant cell line HCT-15 (13GAC), the homozygous mutant cell line A-549 (12AGT) and the heterozygous mutant cell line HUP-T3 (12CGT). **(b)** Raw data from each channel in each sequencing cycles. Cycle 1 to cycle 5 are listed from top to bottom. Signal from each cycle is displayed in four channels, FITC for base T, Cy3 for base G, Cy3.5 for base C and Cy5 for base A.

DETAILED DESCRIPTION OF THE INVENTION

A. Localized synthesis of cDNA from RNA targets *in situ*

[00140] As discussed above, some embodiments concern the detection of RNA, especially mRNA, in cells. The method involves the conversion of RNA to complementary DNA (cDNA) prior to the targeting of the cDNA with a padlock probe(s). The cDNA is synthesized *in situ* at the location of the template RNA. The reverse transcriptase (RT) primer may be modified so as to be capable of immobilization, and in particular immobilization to the cell. Thus it is contemplated that the primer may be provided with a functional moiety, or functional means (*i.e.* a "functionality"), which allows or enables the primer to be immobilized to a component in the sample, *e.g.* a cell or cellular component . This may be for example a functional moiety capable of binding to or reacting with a cell or a sample or cellular component. The use of such a primer, which becomes immobilized to the sample (*e.g.* to or in a cell), has the result that the cDNA product (which is generated by extension of the RT primer and is therefore contiguous with it) also becomes immobilized to the sample (*e.g.* to or in a cell). Methods and compositions are discussed in Application PCT/IB2012/000995, U.S. Patent Application 13/397,503, U.S. Provisional Application Serial No. 61/473,662, and U.S. Provisional Application Serial No. 61/442,921, which are hereby incorporated by reference.

[00141] Since the RCA, which is performed to generate the RCP that is ultimately detected, is carried out using the cDNA as primer (*i.e.* is a target-primed RCA) the RCP is contiguous with the cDNA and thus the RCP is also anchored or attached to the sample (*e.g.* cell). Thus, the use of such a primer ensures or allows that the RCP remains localized to the site of the RNA in the sample (*e.g.* in the cell). In other words localization of the RCP to the original site of the target RNA is preserved. In this way, localization of the signal reporting the target RNA is preserved and thus it can be seen that this favors and facilitates localized *in situ* detection.

[00142] Various such modifications of the RT primer are described herein and include, for example, the provision of reactive groups or moieties in the RT primer, *e.g.* chemical coupling agents such as a thiol group, NHS-esters, *etc.*, which are capable of covalent attachment to the cells or cellular or other sample components, *e.g.* to proteins or other biomolecules in the cell, or to components in the sample *e.g.* matrix components in the

sample. Alternatively or in addition, the primer may be provided with an affinity binding group capable of binding to a cell or cellular or sample component.

[00143] In particular embodiments, a nucleic acid molecule, such as a primer or probe, has been modified to alter its characteristics, such as functionality or activity. In some embodiments, the nucleic acid is subject to depurination and ketone functionalization. Depurination of DNA introduces an aldehyde group which undergoes the full range of reactions expected of aldehydes including the Cannizaro reaction, Cyanohydrin formation, hydration, hydrazine derivatisation, hydrolysis, reductive amination reaction, Schiff base formation, Wolff-Kishner reduction, and reactions with Grignard reagents. The example is further exemplified by the reactions of a hydrazone followed by sodium cyanoborohydride reduction. *See* Mirzabekov, A.; Proudnikov, D. *Nucleic Acid Research* 1996, *24,* 4535; McMurry, J. *Organic Chemistry* 4th Ed, Brooks / Cole 1995; Hermann, G. T. *Bioconjugate Techniques 2nd Ed,* Elsevier, 2008; US2009011836; Dey, S.; Sheppard, T. L. *Organic Letters.* 2001, *3,* 3983, all of which are hereby incorporated by reference in their entirety.

[00144] In other embodiments, a nucleic acid molecule such as a probe or primer has substituted purines, which can then be crosslinked. Heterocyclic bases, nucleotides, nucleotide analogues and alkylating agents which when incorporated into the backbone of the sequence can be covalently cross-linked, such as described in U.S. Patent 6,232,463, which is hereby incorporated by reference. Other examples involve labeled nucleic acids. Included are a variety of methods to form linkers with azide functionalised nucleic acids, these include the Wittig-Horner reaction, imine formation, ether formation, for example by the Williamson method or by the palladium-catalysed Buchwald method, Claisen ester condensation, Ziegler nitrile condensation, acyloin condensation, Ruzicka condensation of carboxylic acid salts of cerium or of thorium, ester formation, amide formation, 4+2 cycloaddition, for example Diels-Alder reaction, Buchwald amination, Suzuki coupling or olefin metathesis. *See*, for instance, U.S. Patent 8,114,636, which is hereby incorporated by reference.

[00145] Other alterations may be implemented using click chemistry, which has been utilized to covalently link DNA and fluorophores through Cu^{I} [3+2] azide-alkyne cycloaddition (CuAAC) reaction and 1,3-dipolar cycloaddition chemistry (Husigens AAC reaction). Brown, T.; El-Sagheer, A. H. Chem. Soc. Rev., 2010, 39, 1388 and U.S. Patent Publication 2008/0311412, both of which are hereby incorporated by reference. Another modification that may be employed with nucleic acids used in methods described herein

include thiol modifications and amine modifications. DNA can be commercially bought with a thiol modification that allows the DNA crosslinking through a range of standard thiol chemistry including the formation of dithiols, reactions with maleimides, Haloacetyls, pyridyl disulphides, acrydites, and acryloyls (Hermann, G. T. Bioconjugate Techniques 2nd Ed, Elsevier, 2008, which is hereby incorporated by reference). Amine modification DNA can be crosslinked through a range of standard amine chemistry including the formation of isothiocyanates, acyl azides, NHS esters, sulfonyl chlorides, aldehydes, glyoxals, epoxides, oxiranes, carbonates, arylating agents, carbodiimides, and anhydrides (Hermann, G. T. Bioconjugate Techniques 2nd Ed, Elsevier, 2008, which is hereby incorporated by reference). As discussed in detail elsewhere, there are several sites on a nucleic acid at which covalent attachment is possible; these include the sugar, the phosphate, the purine and pyrimidine bases (Kricka, L, J. Clinical Chemistry, 2009, 55, 670, which is hereby incorporated by reference).

[00146] In some embodiments, enzymatic modification of a nucleic acid strand may be employed. Oligonucleotide sequences can be modified through the enzymatic action to increases reactivity. For example in the presence of adenosine triphosphate the 5' end of a single strand sequence is adenylated allowing for further crosslinking with hydroxyls (U.S. Patent 4,464,359, which is hereby incorporated by reference). Another technique that can be used is functional biopolymer modification and reagents. Bifunctional phosphorus containing monomers can be incorporated into the oligonucleotide sequence during synthesis, this allows for the introduction of phosphorus based coupling groups including phosphoramidites, phosphonamidites, H-phosphonates, phosphodiesters, phosphotriesters, thiophosphoramidates, thionoalkylphosphonates, thionoalkyl-phosphotriesters and boranophosphates. These reagents also possess a protected hydrazino or oxyamino group including heteroaromatic hydrazine, semicarbazide, carbazide, thiosemicarbazide, thiocarbazide, carbonic acid dihydrazine or hydrazine carboxylate which undergo coupling reactions (U.S. Patent 7,732,628 and U.S. Patent Publication 20110319606, which are both hereby incorporated by reference). The use of psoralens is another modification that may be employed. Psoralens (furocumarins) intercalating between bases and forming permanently bonded adducts between mRNA and primer sequences upon exposure to UV (Lipson, E. S.; Hearst, J. E. Methods in Enzomology, 1998, 164, 330 and U.S. Patent 4,124,598, which are hereby incorporated by reference).

[00147] For sequencing applications, the terminal phosphate group is used as a linker to EDC activated surfaces (U.S. Patent Publication 20100167299, which is hereby incorporated by reference). Accordingly, nucleic acid molecules have a phospholink nucleotide(s) in some embodiments.

[00148] Although cells or cellular components provide a convenient point of attachment, or site of immobilization of the RT primer, this aspect is not restricted to immobilization on or within cells, and the RT primer may be immobilized to other components present in the sample, for example extracellular components. Indeed the components may be natural or synthetic and synthetic components may be added to the sample to supplement or to replace native cellular components. For example, a synthetic matrix may be provided to a cell or tissue sample to preserve signal localization in the method (namely to preserve localization of the RCP product which is detected). Indeed, rather than immobilizing the RT primer (as a means of immobilizing the cDNA), the synthesized cDNA itself or the target RNA may be immobilized in a synthetic matrix which is provided to the sample.

[00149] Thus for example, the target RNA or the synthesized cDNA may be attached to a synthetic gel matrix instead of the native cellular matrix to preserve the localization of the detection signals. This may be achieved by immersing the sample (e.g. the cells or tissue of the sample) in a gel solution which upon polymerization will give rise to a gel matrix to which the cDNA or target RNA can be attached. To achieve such attachment the RT primer may be provided with a reactive group or moiety which can react with the matrix material, for example at the 5′ end thereof. This is described further below.

[00150] In one embodiment, modification, however, the primer is rendered resistant to the ribonuclease. Thus the primer may be modified to be ribonuclease-resistant. A ribonuclease is utilized to digest the RNA hybridized to the cDNA in an RNA:DNA duplex. As discussed below, in some embodiments a ribonuclease made be added or a sample may be incubated under conditions that allow a ribnuclease to digest RNA. In some cases, an endogenous ribonuclease may be employed. In certain embodiments, the ribonuclease may be RNase H or a ribonuclease capable of digesting RNA in an RNA:DNA duplex. In some embodiments, immobilization of the reverse transcriptase primer is achieved by virtue of it being ribonuclease resistant. In such a situation the ribonuclease cannot degrade the RNA which is hybridized to the RT primer. Thus the RT primer protects the primer binding site in the RNA from degradation. The RT primer

accordingly remains bound to the RNA in the cell and in this way is immobilized in the cell. Modifications which may be made to the primer to render it ribonuclease-resistant are described below and include in particular the use of modified nucleotides, or nucleotide analogues for example nucleotides comprising 2'O-Me RNA, methylphosphonates, 2' fluor RNA bases, etc. which when incorporated into the primer, render the primer at least partially resistant to ribonuclease digestion. Alternatively or in addition, the primer may comprise locked nucleic acids (LNAs) or peptide nucleic acids (PNAs). Thus, in some embodiments, it is envisaged that the 5′ end of the cDNA remains bound to the target RNA molecule via a ribonuclease resistant reverse transcriptase primer.

[00151] Methods may involve digestion of mRNA, but in some embodiments, complete digestion of mRNA is not desirable because this allows the primer and hybridized cDNA to diffuse within the cell, which may reduce specificity and resolution. Therefore, in certain embodiments, methods are employed to reduce this by chemically modifying the mRNA and/or the primers to enhance RNase resistance. In other embodiments, a primer is modified to promote its surface conjugation to native proteins in order to prevent its diffusion.

[00152] Embodiments for chemical modification of mRNA and primers to inhibit RNase digestion include a variety of techniques. SHAPE is Selective 2'-Hydroxyl Acylation Analyse by Primer Extension. It involves a 2'-OH group present in a nucleotide backbone that is an essential component in the mechanism of mRNA hydrolysis by RNase. It has been demonstrated that nucleotide backbone exposure to electrophilic reagents results in 2-O' adducts that inhibit RNase digestion, providing chemical resistance. Reagents include 1-methyl-7-nitroisatoic (1M7) and N-methylisatoic anhydride (NMIA). Therefore sight selective synthesis at the appropriate position on the backbone should provide RNAse resistance to both a primer and mRNA. *See* Steen, *et al.*, *Nature Protocols* **2011**, *6*, 1683, Merino, *et al.*, *J. Am. Chem. Soc.* 2005, *127,* 4223, which are both hereby incorporated by reference. Another protocol involves modification of the phosphate backbone. Part of the RNase digestion mechanism involves cyclization of the phosphodiester nucleotide backbone bond with 2'-OH. Modification of this group through the synthesis of phosphorothioates, N3'-P5' phosphoamidates and all of their derivatives prevents RNase hydrolysis by preventing the traditional mechanistic route, and therefore providing site selective RNase resistance. See Gao, et al., Mol Pharmacol, 1992, 41, 223, U.S. Patent 6143881, Gryaznoz; J. Am. Chem. Soc. 1994, 116, 3143, and U.S. Patent 4,415,732, all of which are

incorporated by reference. The use of metal chelators may also be implemented. It has been demonstrated that transition metals such as vanadium (v), oxocanadium (IV) and oxorhenium (V) form metal chelates with the Uracil backbone of RNA. This complex prevented RNase digesting beyond this point therefore site selective chelation may therefore provide a further method for RNase resistance. See Janda, et al. Am. Chem. Soc. 1996, 118, 12521, and U.S. Patent 4,837,312, which are hereby incorporated by reference. Another technique involves primer-mRNA crosslinking. Chemically crosslinking the primer to mRNA will ensure that the primer will neither dissociate from the point of conjugation and also ensure that RNase cannot hydrolyse the mRNA through adding steric hindrance. There are various methods of crosslinking including those involving: psoralens (furocumarins) intercalating between bases and forming permanently bonded adducts between mRNA and primer sequences upon exposure to UV; thiolation of bases to from dithiols; metal complexation; 1,4-phenyldiglyoxal crosslinking between sequences; quinine crosslinking between sequences; and, azinomycin crosslinking between sequences. See Lipson et al., Methods in Enzymology, 1998, 164, 330, Sigurdsson, S. J.; Methods, A Companion to Methods in Enzymology, 1999, 18, 71, Mohammed et al., Bio-Organic and Medicinal Chem. Let. 1999, 9, 1703, Wagner et al., Nucleic Acid Research, 1978, 5, 4065, Pang et al, J. Am. Chem. Soc. 2003, 125, 1116, Alcaro et al., J. Chem. Inf. Model. 2005, 45, 602, U.S. Patent 5,681,941, and U.S. Patent 4,196,281, all of which are hereby incorporated by reference.

[00153] In other embodiments, chemical modification of the primer may be implemented to allow for its surface conjunction with native proteins. A variety of techniques may be employed. For example, methods may involve chemical coupling of modified primers to proteins native to the sample surface. If complete digestion of mRNA occurs the primer-cDNA complex will dissociate from its point of origin, reducing resolution and sensitivity. To prevent this one may modify the primer backbone / 5' end so that it contains a chemical group that can bind with any native proteins on the sample surface. As there is little opportunity to modify the sample prior to use, the conjugation possibilities include four native groups (primary amine, carboxyls, thiols and carbonyls) on the protein surface. There are currently multiple cross-linkers that can be added to the primer to ensure conjugation including: amine reactive groups such as NHS esters, imidoesters and hydroxymethyl phosphene; carboxyl reactive groups such as carbodiimides' thiol reactive groups such as malemides, thiosulphonates and vinylsulphones; and, carbonyl reactive

groups such as hydrazide. Functionalization of the primer with any of these groups followed by the appropriate conditions should cross link the primer to the surface following hybridisation to mRNA ensuring even if complete digestion occurs the primer/cDNA will not dissociate. See Hermann, G. T. Bioconjugate Techniques 2nd Ed, Elsevier, 2008, which is hereby incorporated by reference.

[00154] In some embodiments, a nucleic acid such as a primer may be modified to provide one or more additional properties. In some embodiments, the modification enables the primer to be resistant to degradation, as discussed above. In other embodiments, the modification enables the primer to be attached or localized. In certain embodiments, the modification allows the primer to be crosslinked to one or more chemical moieties. In some cases, the crosslinking occurs via a linker that may or may not be cleavable. In some embodiments, there may be a primary amine reactive group, while in others, there may be a thiol reactive group. In further embodiments, both functional groups may be employed in a linker.

[00155] In another example, heterobifunctional cross-linking reagents and methods of using the cross-linking reagents are described (U.S. Patent 5,889,155, specifically incorporated herein by reference in its entirety). The cross-linking reagents combine a nucleophilic hydrazide residue with an electrophilic maleimide residue, allowing coupling in one example, of aldehydes to free thiols. The cross-linking reagent can be modified to cross-link various functional groups and is thus useful for cross-linking polypeptides and sugars. Table A details certain hetero-bifunctional cross-linkers considered useful in the some embodiments. Moreover, the following references disclose information about such linkers, and are hereby incorporated by reference: Hermann, G. T. *Bioconjugate Techniques 2nd Ed*, Elsevier, 2008; May, J.M, Biochemistry, 28, 1718; Fujiwara, K., *J. Immunol. Methods.* 1998, *112,* 77 ; Tiberi, M., *J. Biol. Chem.* 1996, *271,* 3771 ; Kitagawa, T. *Chem. Pharm. Bull.* 1981, *29,* 1130 ; Trail, P, A. *Science,* 1993, *261,* 212.

TABLE A

HETERO-BIFUNCTIONAL CROSS-LINKERS

Linker	Reactive Toward	Advantages and Applications	Spacer Arm Length\after cross-linking
SMPT	Primary amines Sulfhydryls	· Greater stability	11.2 A
Sulfo-LC-SMPT	Primary amines to Sulfhydryls	· Water Soluble	20.0 A
SPDP	Primary amines Sulfhydryls	· Thiolation · Cleavable cross-linking	6.8 A
LC-SPDP	Primary amines Sulfhydryls	· Extended spacer arm	15.6 A
Sulfo-LC-SPDP	Primary amines Sulfhydryls	· Extended spacer arm · Water-soluble	15.6 A
SMCC	Primary amines Sulfhydryls	· Stable maleimide reactive group · Enzyme-antibody conjugation ·Hapten-carrier protein conjugation	11.6 A
Sulfo-SMCC	Primary amines Sulfhydryls	· Stable maleimide reactive group · Water-soluble · Enzyme-antibody conjugation	11.6 A
AMAS	Primary amines Sulfhydryls	· Low potential for immune response	4.4 A
BMPS	Primary amines Sulfhydryls	· Low potential for immune response	
MBS	Primary amines Sulfhydryls	· Enzyme-antibody conjugation · Hapten-carrier protein conjugation	9.9 A
Sulfo-MBS	Primary amines Sulfhydryls	· Water-soluble	9.9 A
GMBS	Primary amines Sulfhydryls	· Increased stability over MBS	10.2 A
Sulfo-GMBS	Primary amines Sulfhydryls	· Water-soluble	10.2 A
EMCS	Primary amines Sulfhydryls	· Low potential for immune response	9.4 A
Sulfo-EMCS	Primary amines Sulfhydryls	· Water-soluble	9.4 A

TABLE A

HETERO-BIFUNCTIONAL CROSS-LINKERS

Linker	Reactive Toward	Advantages and Applications	Spacer Arm Length\after cross-linking
SIAB	Primary amines Sulfhydryls	· Enzyme-antibody conjugation	10.6 A
Sulfo-SIAB	Primary amines Sulfhydryls	· Water-soluble	10.6 A
SMPB	Primary amines Sulfhydryls	· Extended spacer arm · Enzyme-antibody conjugation	14.5 A
Sulfo-SMPB	Primary amines Sulfhydryls	· Extended spacer arm · Water-soluble	14.5 A
Sulfo-KMUS	Primary amines Sulfhydryls	· Water-soluble	16.3 A
SMPH	Primary amines Sulfhydryls	· Non-cleavable	14.2 A
SIAX / SIAXX	Primary amines Sulfhydryls	· Highly specific to sulfhydryls · Good membrane penetration	10.5 A / 24 A
SIAC / SIACX	Primary amine Sulfhydryls	· Highly specific to sulfhydryls	12.0 A / 24 A
NPIA	Primary amine Sulfhydryls	· Short linker allows for the study of biological interactions	3.0 A
SATA	Primary amine Sulfhydryls	· Hapten-carrier protein conjugation	6.5 A
EDC/Sulfo-NHS	Primary amines Carboxyl groups	· Hapten-Carrier conjugation	0
ABH	Carbohydrates Nonselective	· Reacts with sugar groups	11.9 A
NHS-ASA	Primary amines Photoreactives	· Reacts with sugar groups	8.0 A
Sulfo-NHS-ASA	Primary amines Photoreactives	· Provides an iodination site for tracking	8.0 A
Sulfo-NHS-LC-ASA	Primary amines Photoreactives	· Extended spacer arm · Water-soluble	18.0 A
SASD	Primary amines Photoreactives	· Water-soluble · Potentially Cleavable	18.9 A

TABLE A

HETERO-BIFUNCTIONAL CROSS-LINKERS

Linker	Reactive Toward	Advantages and Applications	Spacer Arm Length\after cross-linking
HSAB	Primary amines Photoreactives	· Short linker allows for stable derivatives	9.0 A
NHS-ASA	Primary amines Photoreactives	· Reacts with sugar groups	8.0 A
Sulfo-NHS-ASA	Primary amines Photoreactives	· Provides an iodination site for tracking	8.0 A
Sulfo-NHS-LC-ASA	Primary amines Photoreactives	· Extended spacer arm · Water-soluble	18.0 A
SASD	Primary amines Photoreactives	· Water-soluble · Potentially Cleavable	18.9 A
HSAB	Primary amines Photoreactives	· Short linker allows for stable derivatives	9.0 A
Sulfo-HSAB	Primary amines Photoreactives	· Water-soluble	9.0 A
SANPAH	Primary amines Photoreactives	· Selectivly activates reactive nitrenes	18.2 A
Sulfo-SANPAH	Primary amines Photoreactives	· Water-soluble	18.2 A
SAND	Primary amines Photoreactives	· Water-soluble · Photoactive at higher wavelengths	18.5 A
ANB-NOS	Primary amines Photoreactives	· Photoactive at higher wavelengths	7.7 A
SADP	Primary amines Photoreactives	· Cleavable	13.9 A
Sulfo-SADP	Primary amines Photoreactives	· Water Soluble	13.9 A
Sulfo-SAPB	Primary amines Photoreactives	· Water Soluble	12.8 A
SAED	Primary amines Photoreactives	· Water Soluble · Fluoresces after activation	23.6 A
Sulfo-SAMCA	Primary amines Photoreactives	· Fluoresces after activation	12.8 A

TABLE A

HETERO-BIFUNCTIONAL CROSS-LINKERS

Linker	Reactive Toward	Advantages and Applications	Spacer Arm Length\after cross-linking
PND	Primary amines Photoreactives	· Potential bonding agent for paramagnetic beads	12.0 A
PNP-DTP	Primary amines Photoreactives	· Can probe active centres of receptor molecules	12.0 A
Sulfo-SANPAH	Primary amines Photoreactives	· Water Soluble · Non cleavable	18.2 A
NHS-Diazirine	Primary amines Photoreactives	· Greater photostability than aryl azides	3.9 A
Sulfo-NHS-Diazirine	Primary amines Photoreactives	· Greater photostability than aryl azides	3.9 A
SDA	Primary amines Photoreactives	· Greater photostability than aryl azides · Cleavable	3.9 A
Sulfo-SDA	Primary amines Photoreactives	· Greater photostability than aryl azides · Cleavable · Water Soluble	3.9 A
Sulfo-LC-SDA	Primary amines Photoreactives	· Greater photostability than aryl azides · Cleavable · Water Soluble · Extended spacer arm	12.5 A
BMPH	Sulfhydryls Carbohydrates	· Water Soluble · Non Cleavable	8.1A
EMCH	Sulfhydryls Carbohydrates	· Non Cleavable	11.8 A
MPBH	Sulfhydryls Carbohydrates	· Extended spacer arm	17.9 A
KMUH	Sulfhydryls Carbohydrates	· Extended spacer arm	19.0 A

TABLE A HETERO-BIFUNCTIONAL CROSS-LINKERS			
Linker	**Reactive Toward**	**Advantages and Applications**	**Spacer Arm Length\after cross-linking**
PDPH	Sulfhydryls Carbohydrates	· Cleavable	9.2 A
ASIB	Sulfhydryls Photoreactives	· Excellent specificity	18.8 A
APDP	Sulfhydryls Photoreactives	· Extended spacer arm	19.5 A
B4I	Sulfhydryls Photoreactives	· High Yield	12.0 A
ASBA	Carboxylate Photoreactives	· Extended spacer arm	16.3 A

[00156] Examples of types of chemical reactions that might be employed include, but are not limited to the following: Diels-Alder chemistry, supramolecular chemistry, click chemistry, or thiol crosslinking (such as SMCC and other described below). Examples of modifications include biotin, amine molecules, thiol molecules, a combination of modifications discussed herein, dendrimers, and random primers.

[00157] Click chemistry that has been employed with nucleic acids is described in El-Sagheer et al., Chem. Soc. Rev. 2010, 39, 1388-1405, which is hereby incorporated by reference. This review describes the use of the copper catalyzed akyne-azide cycloaddition (CuAAC) reaction for use with nucleic acids. In particular embodiments, a primer may employ the CuAAC reaction to add a label to a nucleic acid, such as a fluorescent label, or to add a sugar, peptide, or other reporter groups.

[00158] A "reverse transcription reaction" is a reaction in which RNA is converted to cDNA using the enzyme "reverse transcriptase" ("RT"), which results in the production of a single-stranded cDNA molecule whose nucleotide sequence is complementary to that of the RNA template. However, reverse transcription results in a cDNA that includes thymine in all instances where uracil would have occurred in an RNA complement. The reverse transcription reaction is typically referred to as the "first strand reaction" as the single-

stranded cDNA may subsequently be converted into a double-stranded DNA copy of the original RNA by the action of a DNA polymerase (i.e. the second strand reaction). However, in the present method, a single cDNA strand is formed to act as a target for a sequence-specific padlock probe. The reverse transcription reaction is catalyzed by an enzyme that functions as an RNA-dependent DNA polymerase. Such enzymes are commonly referred to as reverse transcriptases. Reverse transcriptase enzymes are well known in the art and widely available. Any appropriate reverse transcriptase may be used and the choice of an appropriate enzyme is well within the skill of a person skilled in the art.

B. Padlock Probes and RCA

[00159] As mentioned above, the cDNA serves as a target for a padlock probe. Embodiments relating to the use of padlock probes for detection of specific sequences can be found in U.S. Nonprovisional Patent Application 13/397,503, PCT Application PCT/US12/25279, U.S. Provisional Application Serial No. 61/473,662, and U.S. Provisional Application Serial No. 61/442,921, all of which are incorporated by reference in their entirety. Padlock probes are well known and widely used and are well-reported and described in the prior art. Thus the principles of padlock probing are well understood and the design and use of padlock probes is known and described in the art. Reference may be made for example to WO 99/49079. A padlock probe is essentially a linear circularizable oligonucleotide which has free 5′ and 3′ ends which are available for ligation, to result in the adoption of a circular conformation. It is understood that for circularization (ligation) to occur, the padlock probe has a free 5′ phosphate group. To allow the juxtaposition of the ends of the padlock probe for ligation, the padlock probe is designed to have at its 5′ and 3′ ends regions of complementarity to its target sequence (in this case the synthesized cDNA molecule in the cell sample to be analyzed). These regions of complementarity thus allow specific binding of the padlock probe to its target sequence by virtue of hybridization to specific sequences in the target. Padlock probes may thus be designed to bind specifically to desired or particular targets. In the case of some methods, the sequence of the cDNA target is defined by the sequence of the target RNA, i.e. the RNA molecule it is desired to detect. By hybridization to the cDNA target the ends of the padlock probe are brought into juxtaposition for ligation. As described in more detail below, the ligation may be direct or indirect. In other words, the ends of the padlock probe may be ligated directly to each other or they may be ligated to an intervening nucleic acid molecule/sequence of nucleotides. Thus the end regions of the padlock probe may be complementary to adjacent,

or contiguous, regions in the cDNA product of step (a), or they may be complementary to non-adjacent (non-contiguous) regions of the cDNA (in which case, for ligation to occur, the "gap" between the two ends of the hybridized padlock probe is filled by an intervening molecule/sequence).

[00160] Upon addition to a sample, the ends of the padlock probe(s) hybridize to complementary regions in a cDNA molecule(s). Following hybridization, the padlock probe(s) may be circularized by direct or indirect ligation of the ends of the padlock probe(s) by a ligase enzyme. The circularized padlock probe is then subjected to RCA primed by the 3′ end of the cDNA (i.e. the RCA is target-primed). A DNA polymerase with 3′-5′ exonuclease activity is used. This permits the digestion of the cDNA strand in a 3′-5′ direction to a point adjacent to the bound padlock probe. Alternatively, the cDNA may be of appropriate length and may act as the primer for the DNA polymerase-mediated amplification reaction without such digestion. In this way the 5′ end of the RCP is advantageously continuous with the cDNA molecule. As a further alternative, instead of priming the RCA with the cDNA molecule, a separate primer that hybridizes to the padlock probe may be used in the reaction.

[00161] In some embodiments, the DNA polymerase is phi29. In additional embodiments, the phi29 enzyme or any other enzyme used herein has been modified or mutated to alter one or more properties such as stability (in the reaction or shelf-life during storage), activity, fidelity, processivity, speed, or specificity. In certain embodiments, methods involve an enzyme that is added to a sample or reaction. In some embodiments, there may be about, at least about, or at most about 0.1, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7. 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5.0, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9, 8.0, 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.7, 8.8, 8.9, 9.0, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 10.0, 10.5, 11.0, 11.5, 12.0, 12.5, 13.0, 13.5, 14.0, 14.5, 15.0, 15.5, 16.0, 16.5, 17.0, 17.5, 18.0, 18.5, 19.0. 19.5, 20.0, 1, 2, 3, 4, 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, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170,

175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290, 295, 300, 305, 310, 315, 320, 325, 330, 335, 340, 345, 350, 355, 360, 365, 370, 375, 380, 385, 390, 395, 400, 410, 420, 425, 430, 440, 441, 450, 460, 470, 475, 480, 490, 500 U or U/μl (or any range derivable therein) of a specific enzyme such as phi29 (or a combination of enzymes) may added or used. Other enzymes that may be used in these amounts include, but are not limited to, ligase, reverse transcriptase, RNAse, and other polymerases.

[00162] It will be understood by the skilled person that ribonuclease digestion of RNA, hybridization of padlock probes to the cDNA, ligation of the padlock probes, and RCA may be carried out sequentially or simultaneously. Thus, for example, the ribonuclease, the padlock probe(s), the ligase, and the DNA polymerase for RCA may be added to the sample sequentially or substantially at the same time. Furthermore, any combination of steps of the method can be carried out simultaneously and are contemplated within the scope of the methods and composition described herein such that the RCP produced by the method is capable of detection and is indicative of the presence, absence and/or nature of an RNA in a sample. For example, ribonuclease digestion of RNA and hybridization of the padlock probe may be carried out simultaneously, or in the same step, or ligation of the padlock probe and RCA may be carried out simultaneously, or in the same step.

[00163] The "complementary regions" of the padlock probe correspond to the 5′ and 3′ end regions of the probe which hybridize to the cDNA. The padlock probe is thus designed to bind to the cDNA in a target-specific manner. The padlock probe may be designed to detect the presence of a particular RNA, for example to determine if a particular gene is expressed. It may also be designed for genotyping applications, for example to detect the presence of particular sequence variants or mutants in a cell or tissue sample - padlock probes may be designed which are specific for particular known mutants of genes (e.g. known mutations in the KRAS gene, as described further below) or for the wild-type sequence and accordingly may be used to detect or determine the presence, or the distribution (within the context of a tissue sample) of particular mutations or sequence variants, etc.

[00164] Accordingly, based on principles which are known in the art, a padlock probe may be designed to bind to the cDNA at a site selected to detect the presence of a particular sequence or sequence variant in the corresponding RNA. The probes may be

designed and used to verify or confirm the presence of particular mutations or sequence variations (e.g. targeted genotyping) or they may be used on a sample with unknown mutation/variant status, to detect whether or not a mutation/variant is present, and/or the specific nature of the mutation/variant (blinded genotyping). For example a mixture of padlock probes may be used, one designed to detect the wild-type, and one more others designed to detect specific mutations/variants. For such genotyping applications, padlock probes may be designed to have identical complementary regions, except for the last nucleotide at the 3′ and/or 5′ end, which differs according to the genotype the probe is designed to detect; the DNA ligase which is used for circularization of the padlock probe does not accept mismatches when joining the ends of the padlock probe and hence ligation will only occur when the probe hybridizes to a sequence which it "matches" at the terminal nucleotide. In this way, single nucleotide differences may be discriminated.

[00165] In the hybridization reaction both ends of the padlock probe bind to the corresponding portion of, or region in, the cDNA such that they may become ligated, directly or indirectly, to each other resulting in circularization of the probe. Hybridization in this step does not require, but does include, 100% complementarity between the regions in the cDNA and the padlock probe. Thus "complementary", as used herein, means "functionally complementary", i.e. a level of complementarity sufficient to mediate a productive hybridization, which encompasses degrees of complementarity less than 100%. Thus, the region of complementarity between the cDNA and the region of the padlock probe may be at least 5 nucleotides in length, and is in some embodiments 10 or more nucleotides in length, e.g., 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50 or more nucleotides (and any range derivable therein). It may be up to 30, 40, 50, 60, 70, 80, 90 or 100 nucleotides in length (or any range derivable therein) in certain embodiments.

[00166] As noted above the ends of the padlock probe may be ligated directly or indirectly. "Direct ligation" of the ends of the padlock probe means that the ends of the probe hybridize immediately adjacently on the cDNA strand to form a substrate for a ligase enzyme resulting in their ligation to each other (intramolecular ligation). Alternatively, "indirect" means that the ends of the probe hybridize non-adjacently to the cDNA, i.e. separated by one or more intervening nucleotides. In such an embodiment the ends are not ligated directly to each other, but circularization of the probe instead occurs

either via the intermediacy of one or more intervening (so-called "gap" or "gap-filling" (oligo)nucleotides) or by the extension of the 3′ end of the probe to "fill" the "gap" corresponding to the intervening nucleotides (intermolecular ligation). Thus, in the former case, the gap of one or more nucleotides between the hybridized ends of the padlock probe may be "filled" by one or more "gap" (oligo)nucleotide(s) which are complementary to the intervening part of the cDNA. The gap may be a gap of 1 to 60 nucleotides or a gap of 1 to 40 nucleotides or a gap of 3 to 40 nucleotides. In specific embodiments, the gap may be a gap of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 30, 32, 35, 37, 40, 42, 45, 47, 50, 52, 55, 57 or 60 nucleotides, of any integer (or range of integers) of nucleotides in between the indicated values. In further embodiments, the gap may have a size of more than 60 nucleotides. In further embodiments, the gap between the terminal regions may be filled by a gap oligonucleotide or by extending the 3' end of the padlock probe, .e.g. a gap oligonucleotide as defined herein above. Circularization of the padlock probe thereby involves ligation of the ends of the probe to at least one gap (oligo)nucleotide, such that the gap (oligo)nucleotide becomes incorporated into the resulting circularized probe. Hence, in such an embodiment the template for the RCA contains the padlock probe and the gap (oligo)nucleotide. In such an embodiment, the intervening part of the cDNA may be of any length sufficient to allow a productive hybridization with the gap oligonucleotide, wherein by "productive hybridization", it is meant a hybridization capable of templating the indirect ligation (*i.e.* via the gap oligonucleotide) of the ends of the padlock probe. The padlock probe should be designed so that is does not contain any sequence which is complementary to the intervening part of the cDNA (*i.e.* the gap between the hybridized probe ends). The gap oligonucleotide may contain sequences useful for amplification or detection or sequencing, *etc.*, of the eventual RCA product. Additionally or alternatively, the gap oligonucleotide may contain one or more tag or barcode sequences (discussed below). It will be seen that in a related embodiment more than one gap oligonucleotide might be used, which gap oligonucleotides hybridize to the intervening part of the cDNA in such a way that they, and the ends of the padlock probe, are ligated together end-to-end during the ligation step. In the latter case, the gap between the ends of the padlock probe hybridized to the cDNA may be filled by polymerase-mediated extension of the 3′ end of the padlock probe. Suitable polymerases are known in the art. Once the 3′ end has been extended as far as the 5′ end of the padlock probe, the ends may be joined in a ligation reaction. Hybridization of the probe and/or the (oligo)nucleotide to the cDNA is advantageously dependent on the nucleotide sequence of the cDNA thus allowing for the sensitive, specific, qualitative and/or

quantitative detection of one or more cDNA, and by extension the corresponding RNA nucleotide sequences.

C. Samples

[00167] The methods and compositions disclosed herein may be used to evaluate RNA in any sample of cells in which an RNA molecule may occur, so long as the sample is amenable to *in situ* detection. A representative sample may comprise a fixed tissue section, a fresh frozen tissue or a cytological preparation comprising one or more cells. In certain embodiments, the formalin fixed paraffin embedded (FFPE) cells or tissue may be used. The sample may be permeabilized to render the RNA accessible. Appropriate means to permeabilize cells are well known in the art and include for example the use of detergents, *e.g.* appropriately diluted Triton X-100 solution, *e.g.* 0.1% Triton X-100, or Tween, 0.1% Tween, or acid treatment *e.g.* with 0.1M HCl. Permeabilization of tissue samples may also comprise treatment of the sample with one or more enzymes, *e.g.* pepsin, proteinase K, trypsinogen, or pronase, *etc.* Also, microwave treatment of the sample may be carried out as described in the art.

[00168] The sample may also be treated to fix RNA contained in the cells to the sample, for example to fix it to the cell matrix. Such procedures are known and described in the art. For example, in the field of *in situ* hybridization, reagents are known for fixing mRNA to cells. In particular, 5′ phosphate groups in the RNA may be linked to amines present on proteins in the cellular matrix via EDC-mediated conjugation (EDC: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide), thus helping to maintain the localization of the RNA relative to other cellular components. Such a technique has previously been described in relation to microRNAs and their detection via *in situ* hybridization (Pena *et al.,* 2009).

[00169] In certain embodiments, a sample is fixed with formalin. In addition to or instead of formalin, a sample may be fixed with formaldehyde, ethanol, methanol, and/or picric acid. In other embodiments, a sample may be fixed in a non-formalin-based solution, such as Carnoys, Modified Carnoys / Clarkes solution, Ethanol, FineFX, Methacarn, Methanol, Molecular Fixative (UMFIX), BoonFix, Polyethylene glycol based fixatives, RCL2, Uni-Fix, Glyco-Fix, Gluteraldehyde, HistoCHOICE, HistoFix, HOPE Fixation, Ionic liquid, Mirsky's fixative, NOTOXhisto, Prefer, Preserve, or Zenker. *See* NHS "Evidence Review: Non-formalin fixatives" August 2009, which is hereby incorporated by reference. In additional embodiments, fixation techniques include fixation in acetone, methanol, ethanol, methanol

acetone (e.g., fix in methanol, remove excess methanol, permeabilize with acetone), methanol-acetone mix (e.g., 1:1 methanol and acetone mixture), methanol-ethanol mix (e.g., 1:1 methanol and ethanol mixture), formalin, paraformaldehyde, gluteraldehyde, Histochoice, Streck cell preservative (Streck Labs., Nebraska), Bouin's solution (a fixation system containing picric acid), and/or Sed-Fix (a polyethylene glycol based fixation system available from Leica Biosystems, Buffalo Grove VA), FineFix (Leica Biosystems, Buffalo Grove VA).

[00170] Pieces of tissue may be embedded in paraffin wax to increase their mechanical strength and stability and to make them easier to cut into thin slices.

[00171] Permeabilization involves treatment of cells with (usually) a mild surfactant. This treatment will dissolve the cell membranes, and allow larger dye molecules access to the cell's interior.

[00172] In additional embodiments, a sample may be stained before or after contact with one or more padlock probes. In some instances, a sample is stained with a cytological stain such as hematoxylin and eosin (H & E), gram staining, Ziehl-Neelsen staining, Papanicolaou staining, period acid-Schiff (PAS), Masson's trichrome, Romanowsky stains, Wright's stain, Jenner's stain, May-Grunwald stain, Leishman stain, Giemsa stain, silver staining, Sudan staining or Conklin's staining. In certain embodiments, sample may be stained specifically with one or more of acridine orange, Bismarck brown, carmine, coomassie blue, crystal violet, DAPI, eosin, ethidium bromide, acid fuchsine, Hematoxylin (or haematoxylin), Hoechst stains, iodine, Malachine green, methyl green, methylene blue, neutral red, Nile blue, Nile red, osmium tetroxide, rhodamine, and safranin.

D. Localized *in situ* Detection

[00173] The next step of the method following the RCA step is to determine the presence of the extended product (*i.e.* the RCA product or RCP) in the reaction mixture in order to detect the target RNA in the sample. In other words, the sample is screened, *etc.* (*i.e.*, assayed, assessed, evaluated, tested, *etc.*) for the presence of any resultant RCP in order to detect the presence of the target RNA in the sample being tested. The RCP produced by the methods described herein may, in the broadest sense, be detected using any convenient protocol. The particular detection protocol may vary depending on the sensitivity desired and the application in which the method is being practiced. In one embodiment, the RCP detection protocol may include an amplification component, in which the copy number of the RCA

product (or part thereof) is increased, *e.g.*, to enhance sensitivity of the particular assay, but this is not generally necessary. Thus the RCP may be directly detected without any amplification.

[00174] The localized detection may be viewed as comprising two steps, firstly the development of a detectable signal and secondly the read-out of the signal. With respect to the first step, the following detection methods could be contemplated. The signal may include, but is not limited to a fluorescent, chromogenic, enzymatic, radioactive, luminescent, magnetic, electron density or particle-based signal. Thus, a label directly or indirectly providing such a signal may be used. The signal could be obtained either by incorporating a labeled nucleotide during amplification to yield a labeled RCP, using a complementary labeled oligonucleotide that is capable of hybridization to the RCP (a "detection probe"), or to, in a sequence non-specific manner, label the produced nucleic acid. The label could be direct, (*e.g.* but not limited to: a fluorophore, chromogen, radioactive isotope, luminescent molecule, magnetic particle or Au-particle), or indirect (*e.g.* but not limited to an enzyme or branching oligonucleotide). The enzyme may produce the signal in a subsequent or simultaneous enzymatic step. For example horseradish peroxidase may be provided as a label that generates a signal upon contact with an appropriate substrate. Several methods are well described in the literature and are known to be used to render signals that are detectable by various means (which may be used in the second step), *e.g.* microscopy (bright-field, fluorescent, electron, scanning probe), flow cytometry (fluorescent, particle, magnetic) or a scanning device.

[00175] In a particular embodiment, detection is by means of labeled oligonucleotide probes ("detection probes") which have complementarity, and thereby hybridize, to the RCP. Such labeling may be by any means known in the art, such as fluorescent labeling including ratiolabeling, radiolabeling, labeling with a chromogenic or luminescent substrate or with an enzyme *e.g.* horseradish peroxidase, *etc.* Fluorescently-labeled probes are employed in some embodiments. In other embodiments, a chromogenic label is employed. The signal produced by the labels may be detected by any suitable means, such as visually, including microscopically. In particular embodiments, a signal amplification technique or system may be used in conjunction with label detection. With such techniques or systems, the signal from a label may be amplified. In certain embodiments, signal amplification involves tyramide. Tyramide has been used to amplify chromogenic and fluorescent signals. In certain

embodiments, a tyramide amplification system is used, which is commercially available from Perkin Elmer.

[00176] In some embodiments, detection of the label may be direct or it may involve several additional steps and/or reagents to detect the label indirectly. The label may be an enzyme capable of direct detection in the presence of an additional reagent, such as a substrate. Alternatively, and as discussed in further detail below, the label may be detected indirectly through the addition of one or more substances that bind to or react with the label. In some embodiments, for example, such a substance is an antibody or antibody fragment that specifically binds to the label. The antibody or antibody fragment may itself be labeled or it may be detected by the binding of a secondary antibody that itself may or may not be labeled with a detectable moiety. A person of ordinary skill in the art is well aware of a variety of ways of labeling a detection probe. In some embodiments, the crosslinking technology or linkers discussed above may be employed to join a label to a detection probe.

[00177] As the RCPs are comprised of repeated "monomers" corresponding to the padlock probe (optionally with additional incorporated nucleotides or gap oligonucleotides, as discussed above), the sequences to which the oligonucleotide probes hybridize will be "repeated," i.e. assuming the RCA reaction proceeds beyond a single replication of the template, multiple sites for hybridization of the oligonucleotide probes will exist within each RCP. In this way, the signal intensity from the label on the oligonucleotide probes may be increased by prolonging the RCA reaction to produce a long RCA product containing many hybridization sites. Signal intensity and localization is further increased due to spontaneous coiling of the RCP. The resulting coils, containing multiple hybridized oligonucleotide probes, give a condensed signal which is readily discernible by, for example, microscopic visualization against a background of non-hybridized oligonucleotide probes. Hence, it may be possible qualitatively or quantitatively to detect the RNA(s) in a sample without performing a washing step to remove unhybridized oligonucleotide probes.

[00178] Multiplexed detection may be facilitated by using differently-labeled oligonucleotide probes for different RNAs, wherein the respective oligonucleotide probes are designed to have complementarity to "unique" sequences present only in the RCPs (corresponding to sequences present only in the padlock probes) for the respective RNAs.

Such sequences may be barcode or tag sequences, as discussed above. In a particular embodiment, two or more differentially labeled detection oligonucleotides may be used to detect one or more RCPs, one labeled detection oligonucleotides reporting the wild-type variant of a gene and another labeled detection oligonucleotide(s) reporting one or more mutant variants of the gene. Different fluorophores may be used as the labels. Multiplexed detection can also be achieved by applying in situ sequencing technologies such as sequencing by ligation, sequencing by synthesis, or sequencing by hybridization.

[00179] In certain embodiments, screening of RCP may involve sequencing. In yet further embodiments sequencing methods involve sequencing by ligation. In certain embodiments cDNA is generated by in situ reverse transcription of mRNA using an LNA primer and a padlock probe with a gap region between the complementary arms is employed. DNA polymerase may be utilized to generate the complement of the gap between the complementary arms of the padlock probe and DNA ligase utilized to complete a circular DNA molecule. After generation of RCP using cDNA as primer and circular probe DNA as template, in some embodiments, methods further comprise employing one or more anchor primers that are hybridized next to the target. Methods may further comprise ligating 1, 2, 3, or 4 interrogation nonamer oligonucletides (which may or may not include random positions) labeled with different fluorescent dyes and imaging to decode the sequence by analyzing the fluorescence staining pattern for each RCA product through different sequencing cycles. The sequence of the region in question may be read by successive cycles of hybridization of anchor primer, ligation of nonamer oligonucleotides and imaging.

[00180] In other embodiments, additional cycles of ligation sequencing may be achieved by cleavage of the primary interrogation probe and subsequent ligation of additional interrogation probes with different fluorescent dyes that hybridize downstream of the cleaved first probe and are capable of being ligated to the original interrogation probe through possession of 5' phosphate group. In certain embodiments cleavage of the interrogation probe may be achieved by Endonuclease V.

[00181] In certain embodiments sequencing may occur in the 5' to 3' direction or in the 3' to 5' direction or in both directions simultaneously. In additional embodiments sequencing by ligation may employ mate-paired tags. In still further embodiments, the read length of ligation sequencing may be extended by using multiple primers. In yet further

embodiments RCP sequencing may employ alternate sequencing chemistry, for example, sequencing by synthesis as used by Illumina. In certain embodiments sequencing and image acquisition may be automated by using SOLiD 5500, HiSeq, MySeq or Polonator instruments. In other embodiments, greater depth of RNA content sampling may be achieved by the use of micro-dissection or deep sequencing.

[00182] The present method allows for single nucleotide resolution in the detection of RNA nucleotide sequences. The present method may thus be used for the detection of one or more point mutations in an RNA or indeed any single-nucleotide variant. Thus the method may find utility in the detection of allelic variants or alternative splicing, etc. The superior sensitivity and localization afforded by the method also means that it may be used to detect RNA in single cells. For example, multiplex detection of mRNA transcripts in some embodiments may advantageously be used for expression profiling, including in a single cell.

[00183] In some embodiments, replication or amplification of a padlock probe may involve incorporation of a labeled nucleotide that may be specific to one padlock probe so detection of that specific label identifies the sequence of the padlock probe and the sequence complementary to the padlock probe.

[00184] In certain embodiments, a probe is labelled with a detection moiety that can be specifically recognized and/or bound by another agent or substance, which may be referred to as a detection label substance. Such detection moieties or detection labels involve, but are not limited to, antibodies or antibody fragments, haptens or poly haptens, synthetic peptides, antigenic nucleic acid sequences, sequence-specific DNA binding proteins, sequence specific DNA protein complexes, and PNA/DNA hybrids.

[00185] Antibodies are used in certain embodiments. For instance, in some aspects, a polyclonal rabbit IgG is employed. Rabbit immunoglobulins may be covalently linked to detection oligos that are labelled with amino groups at their 5′and/or 3′ end(s) via crosslinkers with an amine reactive functional groups - such as N- hydroxysucccinmide (NHS) esters. The antibody-labelled oligo can then be detected using a polyenzyme- anti-rabbit antibody conjugate such as goat anti-rabbit poly-alkaline phosphatase (goat α rabbit poly-AP or goat anti-rabbit poly-horseradish peroxidase (goat α rabbit poly-HRP). Suitable crosslinkers are available from suppliers such as Thermofisher or Solulink. Goat anti-

Rabbit polyenzyme conjugates are available from Leica Biosciences. It will be understood to those in the art that antibodies that may be used for detection in human cells and tissues include those from chickens, goats, guinea pigs, hamsters, horses, mice, rats, and sheep. IgG antibodies may be obtained from these animal sources. Other examples can be found in U.S. Patent No 6,942,972, Bioconjugate Techniques, Second Edition, Academic Press/Elsevier by Greg T. Hermanson (ISBN 978-0-12-370501-3), Solulink White Paper: "Protein oligonucleotide conjugate synthesis made easy, efficient and reproducible," all of which are hereby incorporated by reference.

[00186] In some embodiments, a detection tag may be an antibody fragment, such as an IgG Fc fragment. As discussed above, fragments of antibodies from chickens, goats, guinea pigs, hamsters, horses, mice, rats, or sheep may be employed in embodiments discussed herein. For example, a rabbit IgG Fc fragment might be used. This can be employed by covalently attaching the antibody fragment to a detection oligo and using a polyenzyme labelled anti-rabbit Fc fragment-specific antibody to effect detection. This may give a cleaner result than using a whole antibody. Systems could also be designed using alternative antibody fragments such as F(ab′)2 fragments. Antibody fragments include that that may be obtained from the antibodies discussed in the references above. In some embodiments, an antibody fragment is a recombinant antibody fragment. Such a fragment has the potential to give very low background staining. It would be possible to produce a monoclonal antibody and use a fragment of it for detection assays.

[00187] A variety of haptens may be employed, including those in which commercial antibodies are readily available, such as Alexaflour, Biotin, BODIPY (boron-dipyrromethene), Cascade blue, Dansyl, Digoxygenin, Dinitrophenol (DNP), Lucifer Yellow, Oregon Green, Rhodamine, Streptavidin, TAMRA (tetramethyl rhodamine), and Texas Red. For instance, fluorescein can be incorporated at the 3′ and 5′ ends of detection oligos during synthesis. Anti-fluorescein antibodies are commercially available and can be detected using Leica BioSystems standard detection system. Additionally, poly haptens may be employed in conjunction with a detection oligo. Polyethylene glycol (PEG) is an example of a poly hapten For instance, in some embodiments, detection oligos are conjugated to PEG. The repeating units of PEG can then be bound by an anti PEG antibody (which are available commercially from suppliers such as Epitomics and Life Sciences Inc). A degree of amplification can be achieved depending upon how many PEG

repeating units are bound by antibody. Jäschke J, Fürste JP, Nordhoff E, Hillenkamp F, Cech, and Erdmann VA. Synthesis and properties of oligodeoxyribonucleotide – polyethylene glycol conjugates. Nucleic acids Research 1994: 22 (22) 4810-4817, which is hereby incorporated by reference. Other examples include attaching known haptens to a polymer with suitable functional groups in its repeating units (e.g. the repeating amine groups in poly-lysine or the repeating carboxylate groups in poly-glutamic acid). In some embodiments the attachment is polyacrylamide hydrazide or various poly-amino acids (poly-arginine, poly-asparagine, poly-aspartic acid, poly-glutamic acid. poly-glutamine and poly-lysine) as possible backbone/scaffold molecules to which haptens can be attached. A hapten that may be used includes, but is not limited to, dinitrophenol, biotin, digoxygenin, fluorescein and rhodamine, as well as oxazole, pyrazole, thiazole, nitroaryl, benzofuran, triterpen, urea, thioureas, rotenoid, coumarin or cytolignin, any of which might be suitable for attachment to the polymeric backbone.

[00188] Synthetic peptides may also be used in detection methods. Peptides that may be attached to a detection probe would be those that could be specifically recognized and/or bound by a detection label substance. In some embodiments, the synthetic peptide is YPYDVPDYA (from influenza hemaglutinin), HHHHHH (6 x His or His tag), DYKDDDDK (the FLAG® peptide), all of which are recognized by antibodies that are commercially available from Origene. In other embodiments, the synthetic peptide is HHHHHHGS (6x His variant) recognized by Millipore's antibody Clone 4D11 or ATDYGAAIDGF (from Phage M13 Coat protein g3p), which is recognised by Anti-g3p (pIII) available from MoBiTec.

[00189] Peptides can be conjugated to oligonucleotides using the same chemistry as that used for antibodies (see above). A huge variety of peptide specific antibodies are commercially available. Most of those in the LBS range are unsuitable as they are directed at targets that occur in human cells. It would be better to use a non-biological peptide sequence - or a sequence from a protein of plant or prokaryotic origin. However, LBS could design and produce bespoke anti-peptide antibodies for use in detection systems. See Lass-Napiorkowska A, Heyduk E, Tian L and Heyduk T. Detection methodology based on target molecule-induced sequence-specific binding to a single-stranded oligonucleotide. Analytical Chemistry 2012: 84 (7) 3382-3389, which is hereby incorporated by reference.

[00190] In further embodiments, a detection label that is attached to a detection probe is an antigenic DNA sequence. In an aspect of methods described herein, a detection label may be an E2 site 25 from the human papilloma virus(HPV) sequence DNA, i.e. 5′-GTAACCGAAATCGGTTGA-3′ (SEQ ID NO: 113). Raising antibodies that recognize specific DNA sequences may be technically difficult. An approach that has worked involved a highly stable DNA protein complex as an immunogen. Such antibodies can be used in conjunction with a cognate sequence in the context of a detection probe for padlock probes. See Cerutti ML, Centeno Jm Goldbaum FA and de Prat-Gay G. Generation of sequence-specific high affinity anti-DNA antibodies. J. of Biochemistry 2001 276 (16) 12769-12773, which is hereby incorporated by reference.

[00191] In additional embodiments a detection tag may involve a sequence-specific DNA binding protein. In some embodiments, the protein is all or part of the E2 protein from HPV 16, Tet repressor (from Gram negative bacteria), Tet repressor (from Gram negative bacteria), or GAL 4 (from yeast). This can be implemented by incorporating one strand of the recognition sequence for the DNA binding protein into the padlock probe and the complementary sequence into the detection probe. The RCA products are incubated with a cocktail of the DNA binding protein, the detection oligo and antibody that will recognise the DNA-binding protein. Antibodies to Tet repressor are commercially available (from Gen Way Biotech for example). Antibodies to E2 have been described by Cerutti et al. Care should be taken to avoid the use of a protein that has significant non-specific DNA-binding activity. See Cerutti et al., Journal of Biochemistry, 2001, 276 (16) 12769-12773 and Pook et al., European Journal of Biochemistry, 1998, 258 915-922, both of which are hereby incorporated by reference.

[00192] Alternatively, sequence specific DNA protein complexes may be employed in detection methods. For instance the following complexes may be employed: E2 protein/ E2 site 25 DNA – complex; Tet repressor/ tet operator complex. This involves the use of antibodies that specifically recognize the complex between a DNA-binding protein and its cognate sequence, or specifically recognise the conformation of the DNA-bound form of the protein. This approach has the potential to circumvent problems that might be caused by non-specific DNA binding. See Cerutti et al., Journal of Biochemistry, 2001, 276 (16) 12769-12773 and Pook et al., European Journal of Biochemistry, 1998, 258 915-922, both of which are hereby incorporated by reference.

[00193] In certain embodiments, the detection label or tag may involve a PNA/DNA hybrid. Peptide Nucleic Acids (PNA) form extremely stable hybrids with DNA or RNA. Antibodies have been developed that recognise the backbone conformation (rather than the sequence of such hybrids). It is therefore possible to design a system whereby the detection oligo consists of PNA . Once this has hybridized to its target it can be detected by a PNA/DNA –specific antibody, which can in turn be recognized by one of Leica BioSystems standard detection systems. See e.g., U.S. Patent 5,612,458, which is hereby incorporated by reference.

[00194] In addition, in some embodiments a detection moiety may be specifically recognized or bound by non-antibody proteins or protein domains that mediate specific high-affinity interactions. The group includes, for instance, protein structures comprising ankyrin-repeats. Typically, in designed ankyrin-repeat proteins (DARPins) three, four or preferably five repeat ankyrin motifs are present. These may form a stable protein domain with a large potential target interaction surface. Further details may be derived, for example, from Binz et al., 2003, J Mol Biol.; 332(2): 489-503, which is incorporated herein by reference.

[00195] A further example of a specific, highly affine molecule is an affibody molecule, i.e. a protein based on the Z domain (the immunoglobulin G binding domain) of protein A. In contrast to antibodies, affibody molecules are typically composed of alpha helices and lack disulfide bridges. They may be expressed in soluble and proteolytically stable forms in various host cells. Affibody molecules may further be fused with other proteins. Further details may be derived, for example, from Nord et al., 1997, Nat Biotechnol.; 15(8): 772-777, which is incorporated herein by reference.

[00196] The group of highly affine protein interactors also comprises adnectins. Adnectins are based on the structure of human fibronectin, in particular its extracellular type III domain, which has a structure similar to antibody variable domains, comprising seven beta sheets forming a barrel and three exposed loops on each side corresponding to the three complementarity determining regions. Adnectins typically lack binding sites for metal ions and central disulfide bonds. They are approximately 15 times smaller than an IgG type antibody and comparable to the size of a single variable domain of an antibody. Adnectins may be customized in order to generate and/or increase specificity for target molecules by modifying the loops between the second and third beta sheets and between

the sixth and seventh sheets. Further details may be derived, for example, from Koide and Koide, 2007, Methods Mol Biol.; 352: 95-109, which is incorporated herein by reference.

[00197] A further example is the antibody mimetic anticalin, which is derived from human lipocalin. Anticalins typically have the property of binding protein antigens, as well as small molecule antigens. They are composed of a barrel structure formed by 8 antiparallel beta sheets, connected by loops and an attached alpha helix. Mutagenesis of amino acids at the binding site may allow for changing of affinity and selectivity of the molecule. Further details may be derived, for example, from Skerra, 2008, FEBS J., 275 (11): 2677–83, which is incorporated herein by reference.

[00198] Another example is affilin, i.e. a genetically engineered protein with the ability to selectively bind antigens, which is structurally derived from gamma-B crystallin or from ubiquitin. Affilins are typically constructed by modification of near-surface amino acids of gamma-B crystallin or ubiquitin and isolated by display techniques such as phage display. The molecular mass of crystallin and ubiquitin based affilins is typically about one eighth or one sixteenth of an IgG antibody, respectively. This may lead to heat stability up to 90°C and an improved stability towards acids and bases. Further details may be derived, for example, from Ebersbach et al., 2007 J Mol Biol.; 372(1): 172-185 or from Hey et al., 2005, Trends Biotechnol.; 23(10): 514-522, which are incorporated herein by reference.

[00199] The group of highly affine protein interactors also comprises avimers, i.e. artificial proteins that are able to specifically bind to certain antigens via multiple binding sites. Typically, the individual avimer sequences are derived from A domains of various membrane receptors and have a rigid structure, stabilized by disulfide bonds and calcium. Each A domain can bind to a certain epitope of the target molecule. The combination of domains binding to different epitopes of the same target molecule may increases affinity to this target. Further details may be derived, for example, from Silverman et al., 2005, Nat Biotechnol.; 23(12): 1556-61, which is incorporated herein by reference.

[00200] Other embodiments include knottins, i.e. small disulfide-rich proteins characterized by a special disulfide through disulfide knot. This knot is typically obtained when one disulfide bridge crosses the macrocycle formed by two other disulfides and the interconnecting backbone (disulfide III-VI goes through disulfides I-IV and II-V). Knottin peptides could be shown to bind with high affinity (about 10 to 30 nmol/L) to integrin

receptors. The knottin scaffold may accordingly be used for the design of highly affine molecules which are able to bind detection moieties according to the invention. Further details may be derived, for example, from Kimura et al., 2009, Cancer Res., 69; 2435, which is incorporated herein by reference.

[00201] The group of highly affine protein interactors additionally comprises fynomers, i.e. Fyn SH3-derived proteins .Fyn is a 59-kDa member of the Src family of tyrosine kinases. The Fyn SH3 domain comprises 63 residues, and its amino acid sequence is fully conserved among man, mouse, rat, and monkey. Fynomers are typically composed of two antiparallel beta sheets and contain two flexible loops (RT and n-Src loops) to interact with other proteins or targets. Further details may be derived, for example, from Grabulovski et al., 2007, Journal of Biological Chemistry, 282 (5): 3196–3204, which is incorporated herein by reference.

[00202] Yet another example of a specific, highly affine molecule is a phylomer peptide. Phylomer peptides are bioactive fragments of naturally occurring proteins that are encoded in the genomes of evolutionary diverse microbes, which are partially sourced from extreme environments and may have evolved over billions of years, providing a multitude of distinct and stable structures capable of binding to biological molecules. Further details may be derived, for example, from Watt, 2009, Future Med. Chem., 1(2): 257-265, which is incorporated herein by reference.

[00203] The group of highly affine protein interactors also comprises kunitz domain peptides. Kunitz domains are the active domains of Kunitz-type protease inhibitors. They typically have a length of about 50 to 60 amino acids and a molecular weight of 6 kDa. Examples of Kunitz-type protease inhibitors are aprotinin, Alzheimer's amyloid precursor protein (APP), and tissue factor pathway inhibitor (TFPI). Kunitz domains are stable as standalone peptides and are able to recognize specific targets such as protein structure and may accordingly be used for the design of highly affine molecules which are able to bind detection moieties according to the invention. Further details may be derived, for example, from Nixon and Wood, 2006, Curr Opin Drug Discov Devel, 9(2), 261-268, which is incorporated herein by reference.

[00204] Other detection methods may be employed. In some embodiments, Förster (Fluorescence) resonance energy transfer (FRET) may be implemented to detect the

presence or absence of a particular sequence. Examples can be found in Li et al., Biochem Biophys Res Commun. 2008 Sep 5;373(4):457-61, which is hereby incorporated by reference.

[00205] Any probe described herein may be multiply labeled with one or more of the same or different labels. In some cases, a probe may be multiply labeled with the same label. For example, branched probes may be used in which one or more labels is attached on each branch. In some embodiments, a branched DNA (bDNA) signal amplification technique is used involving sets of labeled probes, hybridized sequentially to the target nucleic acid creating comb-like DNA structures, which generate chromogenic or fluorescent signals. See Murphy et al., J Clin Microbiol, 1999 Mar;37(3):812-4, Player et al., J Histochem Cytochem, 2001 May;49(5):603-12, and Collins et al., Nucl. Acids Res., (1997) 25 (15): 2979-2984, which are all incorporated by reference. In further embodiments, tyramide signal amplification (TSA) may be employed. TSA is based on the ability of horseradish peroxidase (HRP) to convert fluorescent or hapten-labeled tyramide molecules into a highly reactive oxidized intermediate that can bind tyrosine at the site of the HRP-probe binding site. See Speel EJ, Nucl. Acids Res. (1997) 25 (15): 2979-2984, Thompson et al., Neuron, 60(6), 1010-1021, Werner et al., Prog Histochem Cytochem., 2001;36(1):3-85, and Qian et al., Diagn Mol Pathol., 2003 Mar;12(1):1-13, all of which are hereby incorporated by reference.

[00206] In certain embodiments, multiple padlock probes or detection probes may be employed. It is contemplated that in some embodiments, probes are detected serially instead of at the same time. In cases where probes are added serially, they may be detected serially and there may be a step in between in which detection of a previously added probe is eliminated after it has already been detected. An example of this might be achieved using photobleaching. For example, a first probe may be detected and then the probe may be photobleached prior to addition of a second probe that is then detected. Photobleaching refers to the photochemical destruction of a fluorophore. In the context of FRET, phobleaching may involve the acceptor or the donor molecule. Additional embodiments may involve fluorescence recovery after photobleaching (FRAP).

[00207] In some embodiments, a labeled nucleotide is incorporated directly into a probe that is either the padlock probe or the detection probe. In certain embodiments, the labeled

nucleotide is incorporated in the detection probe, while in other embodiments labeled nucleotides are incorporated into the amplification product.

[00208] Oligonucleotides and dNTPs may be labelled with a variety of substances including radioactive isotopes (such as 13C, 3H, 32P, 35S), haptens (such as organic fluorescent dyes, biotin, digoxigenin (DIG), dinitrophenyl (DNP)), or enzymes (such as calf intestinal alkaline phosphatase or HRP). Haptens are small molecules that can illicit an immune response only when coupled to a larger carrier molecule, such as a protein. Hapten labels are usually used for indirect detection methods in combination with streptavidin or antibody conjugates (as discussed above). Fluorescent labels are used for direct detection.

[00209] Nucleotide analogs are routinely used to label, isolate, study, and manipulate DNA in a wide variety of applications. These nonradioactive nucleotide analogs are introduced into a DNA strand by chemical and enzymatic 5′ and 3′ end labeling and through internal enzymatic labeling or post-labeling methods. The ability to incorporate modified nucleotides into a growing chain of dNTPs is dependent upon a number of factors including the DNA polymerase (especially its fidelity), size and type of fluorophore, the linker between the nucleotide and the fluorophore, and position for attachment of the linker and the cognate nucleotide. In some embodiments, a polymerase that contains a strong 3′ to 5′ exonuclease activity (proofreading ability) is not used for incorporating nucleotide analogs. (Jon P. Anderson et al, (BioTechniques, Vol. 38, No. 2, February 2005, pp. 257–264, which is incorporated by reference). However, according to patent application US 2011/0244548 A1, which is incorporated by reference, Life Technologies have developed several novel DNA polymerases that have reduced discrimination against the incorporation of one or more fluorescently labeled nucleotides into DNA/polynucleotides. Furthermore, various sequencing by synthesis methods (used by Helicos and Illumina) use the incorporation of fluorescent dye terminators whereby a single fluorophore labeled “terminator/inhibitory” nucleotide is incorporated per cycle. Excitation of individual fluorophores by laser is recorded. Fluorophore and terminator/inhibitory group are then removed allowing addition of the next nucleotide.

[00210] Nucleotide incorporations can be detected through fluorescence resonance energy transfer (FRET) interactions between a fluorophore-bearing polymerase and .gamma.-phosphate-labeled nucleotides, or with zeromode waveguides. The illumination can be

restricted to a zeptoliter-scale volume around a surface-tethered polymerase such that incorporation of fluorescently labeled nucleotides can be observed with low background (Levene, M. J. et al. Zero-mode waveguides for single-molecule analysis at high concentrations. Science 299, 682-686 (2003); Lundquist, P. M. et al. Parallel confocal detection of single molecules in real time. Opt. Lett. 33, 1026-1028 (2008); Korlach, J. et al. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. Proc. Natl. Acad. Sci. USA 105, 1176-1181 (2008), the disclosures of which are incorporated herein by reference in their entireties).

[00211] In addition to what is described above, further embodiments related to detection probe and detection systems are discussed below. U.S. Patent 8,114,973 describes optically labeled nucleotides. U.S. Patent 8,133,702 describes fluorescent dye labeled nucleotides. The Alex Fluor dyes (Molecular Probes) are widely used to fluorescently label nucleotides. U.S. Patent 7,235,361 describes fluorescent semiconductor nanocrystals (quantum dots) which can be associated with compounds, including nucleotides. These U.S. patent cited above are all specifically incorporated by reference. Nucleotides labeled with quantum dots are available from Life Technologies. In other embodiments, detection methods may employ quantum dots, which refer to semiconductor nanocrystals that have broad excitation spectra, narrow emission spectra, tunable emission peaks, long fluorescence lifetimes, negligible photobleaching, and the ability to be conjugated to biomolecules, such as probes. See Barroso, J Histochem Cytochem. 2011 Mar;59(3):237-51, which is hereby incorporated by reference. Other references that describe the use of labeled nucleotides are incorporated by reference; these include the following references discussing radiolabelled nucleotides (Yan et al., Biochem Biophys Res Commun, 2001 Jun 8;284(2):295-300; von Guggenberg et al., Cancer Biother Radiopharm., 2010 Dec;25(6):723-31, Tang et al, Biotechniques, 2006 Jun;40(6):759-63, all of which are incorporated by reference); fluorescently labeled nucleotides (Bethge et al., Org Biomol Chem, 2010 May 21;8(10):2439-48, Linck et al., Photochem Photobiol, 2012 Feb 23. doi: 10.1111/j.1751-1097.2012.01119.x, Knapp et al., Chemistry, 2011 Mar 1;17(10):2903-15, Linck et al., European J. Medicinal Chemistry, 2010;45(12):5561-6, Jarchow-Choy et al., Nucleic Acids Res, 2011 Mar;39(4):1586-94, which are all incorporated by reference); enzyme-conjugated oligonucleotides (Ghosh et al., Bioconjug Chem, 1990 Jan-Feb;1(1):71-6, van de Corput et al., J Histochem Cytochem, 1998 Nov;46(11):1249-59,

which are hereby incorporated by reference); antibody-conjugated oligonucleotides (Kazane et al., Proc Natl Acad Sci USA, 2012 Mar 6;109(10):3731-6. Han et al., Bioconjug Chem, 2010 Dec 15;21(12):2190-6, which are hereby incorporated by reference); biotin- or streptavidin-conjugated nucleotides (Rabe et al., Molecules, 2011 Aug 15;16(8):6916-26, which is hereby incorporated by reference); DIG-conjugated nucleotides (Kriegsmann et al., Histochem Cell Biol, 2001 Sep;116(3):199-204, Jalabi et al., J Histochem Cytochem, 2003 Jul;51(7):913-9, Escarceller et al., Anal Biochem, 1992 Oct;206(1):36-42, Trayhurn et al. Anal Biochem, 1994, Oct;222(1):224-30, which are hereby incorporated by reference); DNP conjugated nucleotides (Horáková et al., Org Biomol Chem, 2011 Mar 7;9(5):1366-71, Keller et al., Anal. Biochem, 1989 Mar;177(2):392-5, all of which are incorporated by reference); quantum dot labelled nucleotides (He et al., Biomaterials, 2011 Aug;32(23):5471-7, Bakalova et al.., J Am Chem Soc, 2005 Aug 17;127(32):11328-35, Ma et al., Chromosoma, 2008 Apr;117(2):181-7, Barrett et al., Nanoscale, 2011 Aug;3(8):3221-7. Zhang et al., Faraday Discuss, 2011;149:319-32; discussion 333-56, which are hereby incorporated by reference); nanoparticle-conjugated nucleotides (Zheng et al., J Am Chem Soc, 2008 Jul 30;130(30):9644-5; Lee et al., "Multiplexed detection of oligonucleotides with bio-barcoded gold nanoparticle probes," in Methods in molecular biology (Clifton, NJ), 2011 – Springer-Verlag, all of which are hereby incorporated by reference); metal-conjugated oligonucleotides (Oser et al., Nucleic Acids Res, 1988 Feb 11;16(3):1181-96, Wei et al., Anal. Bioanal Chem, 2012 Jan;402(3):1057-63, Gasser et al., Dalton Trans 2011 Jul 21;40(27):7061-76, which are all incorporated by reference); and, PNAs (Gasser et al., Dalton Trans, 2012 Feb 28;41(8):2304-13, which is hereby incorporated by reference). More details regarding detection methods can be found in Olesen et al., Biotechniques, 1993 Sep;15(3):480-5, Suzuki et al., Anal. Biochem, 1993 May 1;210(2):277-81, Chai et al., Anal. Chim Acta, 2012 Feb 17;715:86-92, all of which are hereby incorporated by reference.

[00212] A variety of embodiments involve sequencing one or more nucleic acids. In some embodiments, after a padlock probe is amplified or replicated, the rolling circle amplification product is sequenced. In some embodiments, a nucleic acid is sequenced on a slide or in the same physical context that one or more other reactions occurred, such as replication. In other embodiments, a nucleic acid is removed from a slide or from the physical context under which a previous enzymatic reaction (via an exogenously added

enzyme) occurred according to methods described in embodiments, such as transcribing using reverse transcriptase, ligating, or replication. For example, cells on a slide or a tissue sample on a slide may be physically removed from the slide. In some embodiments, the cells or tissue are placed in a tube and are no longer attached or fixed to a physical surface or support. A sequencing reaction may occur on a physical or solid support or it may be performed in solution.

[00213] Sequencing-by-synthesis involves the template-dependent addition of nucleotides to a template/primer duplex. Traditional sequencing-by-synthesis is performed using dye-labeled terminators and gel electrophoresis (so-called "Sanger sequencing"). See, e.g., Sanger, F. and Coulson, A. R., 1975, J. Mol. Biol. 94: 441-448; Sanger, F. et al., 1977, Nature. 265(5596): 687-695; and Sanger, F. et al., 1977, Proc. Natl. Acad. Sci. U.S.A. 75: 5463-5467, both of which are hereby incorporated by reference. In other embodiments, a terminator nucleotide may be employed, which may be labeled. In some embodiments, sequencing is accomplished using high-resolution electrophoretic separation of resulting single-stranded extension products in a capillary-based polymer gel or by mass spectroscopy.

[00214] Single molecule sequencing methods have been proposed that provide increased resolution, throughput, and speed at reduced cost. For example, a sequencing-by-synthesis method that results in sequence determination without consecutive base incorporation, has been proposed by Braslavsky, et al., Proc. Nat'l Acad. Sci., 100: 3960-3964 (2003), which is hereby incorporated by reference. These methods do not rely on the user of terminator nucleotides as in Sanger sequencing. Instead, template/primer duplex is anchored directly, or indirectly (e.g., via a polymerase enzyme) to a surface and labeled nucleotides are added in a template-dependent manner.

[00215] In addition to Sanger sequencing, sequencing may also occur by "iterative cycles of enzymatic manipulation and imaging based data collection." Shendure et al., Nature Biotech., 2008, 26(10):1135-1145, which is hereby incorporated by reference. These second generation technologies may be categorized as follows: (1) microphoretic techniques; (2) sequencing by hybridization; (3) observation of single molecules in real time; and, (4) cyclic array sequencing. Commercial embodiments of high-throughput sequencing technology include 454 sequencing (Roche Applied Sciences), Solexa (Illumina), SOLiD (Applied Biosystems), the Polonator (Dover/Harvard), Helioscope™

Single Molecule Sequencer technology (Helicos), Massively Parallel Signature Sequencing (MPSS) (Lynx Therapeutics), Ion semiconductor sequencing (Ion Torrent), DNA nanoball sequencing (Complete Gneomics), Single molecule SMRT™ sequencing (Pacific Biosciences), Single molecule real time (RNAP) sequencing, and Nanopore DNA sequencing. Any of the sequencing technologies discussed above may be modified or applied so as to obtain sequence information from a sample that has undergone rolling circle amplification of one or more padlock probe to detect or identify nucleic acid sequence(s). One or more aspects of the sequencing technology may be performed while a sample remains on a solid support or it may occur in solution, such as after the sample has been removed from a solid support or has been otherwise disrupted on the solid support.

[00216] It is contemplated that in certain embodiments, any sequencing that is performed in conjunction with padlock probe technology is automated. In other embodiments, while aspects of methods may be automated, others may not be. In some embodiments it is contemplated that sequencing may be performed in situ on a solid support or that enzymatic or chemical reactions prior to sequencing may occur in situ on a solid support (such as addition of chain terminating nucleotides). In other embodiments, separation or isolation of a nucleic acid to be sequenced occurs in solution or not in situ or not on the solid support used for rolling circle amplification. In further embodiments, identification of a sequence does not occur in situ or does not occur on the solid support that was used for rolling circle amplification. In certain embodiments, a solid support used for rolling circle amplification may be moved to a different machinery or location in order to do all or part of a sequencing reaction. In particular embodiments, sequencing involves a machine for electrophoretic separation, mass spectroscopy, fluorescence detection, ion sensor, light detector or other signal detector.

[00217] In some embodiments, it is contemplated that rolling circle amplification products that are sequenced have not been generated based on padlock probes that were ligated after hybridization to PCR products. In specific embodiments, sequencing of rolling circle amplification products depends on replication of padlock probes that hybridize to cDNA that is complementary to an RNA transcript in a sample.

[00218] As will be appreciated by the skilled person, the present method may be used in various diagnostic applications, in particular those that require single nucleotide sensitivity. For example, this method may be used to detect point mutations which are

associated with disease, disease risk or predisposition, or with responsiveness to treatment, etc., e.g. activating mutations in oncogenes.

[00219] Methods may be adapted for automation, for example, by applying procedures as used in conventional automated FISH assays. In certain embodiments, the solutions used for performing one or more reactions may be important for efficiency and the integrity of the detection methods. In some embodiments, an effort to decrease evaporation/dryback during periods of prolonged incubation at elevated temperatures.

[00220] The viscosity is the tendency of the fluid to resist flow. Increasing the concentration of a dissolved or dispersed substance generally gives rise to increasing viscosity (i.e., thickening), as does increasing the molecular weight of a solute (a dissolved substance).

[00221] The relationship between viscosity and concentration is generally linear up to viscosity values of about twice that of water. This dependency means that more extended molecules (e.g., linear polymers) increase the viscosity to greater extents at low concentrations than more compact molecules (e.g., highly branched polymers) of similar molecular weight.

[00222] The following substances may be added to increase viscosity of solutions: alcohols or polyols such as glycerol or glycerine, ethylene glycol or 1,2-ethanediol, poly ethylene glycol (PEG) (an oligomer or polymer of ethylene oxide), diethylene glycol (DEG), or PVA or polyvinyl alcohol (synthetic polymer); saccharides or proteins such as trehalose (naturally occurring disaccharide,) glucose, fructose, dextran sulphate (sulphated polysaccharide), betaine or N,N,N-trimethylglycine (N-trimethylated amino acid); natural hydrocolloids such as botanical, animal and microbial hydrocolloids that include but are not limited to acacia gum (gum Arabic, which is a mixture of polysaccharides and glycoproteins, derived from tree bark), tragacanth (mixture of polysaccharides, derived from shrubs), guar gum (polysaccharide, derived from seed of shrubs), locust bean gum (polysaccharide, derived from seeds), agarose (linear polysaccharide, derived from seaweed), agar (mixture of agarose (linear polysaccharide) and agaropectin, derived from seaweed), carageenan (linear sulfated polysaccharide, derived from seaweed), alginate (anionic polysaccharide, derived from seaweed), cellulose (polysaccharide), xanthan gum (polysaccharide, microbial origin), pectin (heteropolysaccharide), gelatin (mixture of

peptides and proteins, animal origin); semisynthetic hydrocolloids, which are hydrocolloids of natural origin that have been modified by further chemical process, and examples are copolymers of starch or cellulose, such as starch-acrylonitrile graft copolymer (a starch polyacrylate salt, and sulfuric acid), vinyl sulfonate, methacrylic acid, vinyl alcohol, vinyl chloride copolymers, methyl cellulose, (CMC) SodiumCarboxyMethylCellulose, (HMC) HydroxyMethylCellulose, (HEMC) HydroxyEthylMethylCellulose, (HPMC) HydroxyPropylMethylCellulose, (HEC) HydroxyEthylCellulose, and (HPC) HydroxyPropylCellulose; synthetic hydrocolloids such as Carbopol®; surfactants such as Tween (polysorbate), NP-40, Triton X-100, SDS (sodium dodecyl sulphate), pluronics/poloxamers, which block copolymers based on ethylene oxide and propylene oxide; miscellaneous polymers such as polyvinyl pyrrolidone (PVP) (polymer made from the monomer N-vinylpyrrolidone) and carbomers (synthetic high molecular weight polymers of acrylic acid), ammonium salts such as tetramethylammonium chloride (TMAC) or other quaternary ammonium salt; amides such as formamide, n-methylformamide, dimethylformamide, 2-pyrollidone, methylpyrollidone, hydroxyethylpyrollidone, acetamide, methylacetamide, dimethylacetamide, propionamide, isobutyramide; organosulfur compounds such as DMSO; amino alcohols or glycols or polyols such as aminoglycols, aminopolyols, 3-amino-1,2-propandiol, diethanolamine, or triethanolamine; organic borates such as 1-butyl-4methylpyridium tetrafluoroborate; organic sulfates such as 1-butyl-3methylimidazolium 2 ethyl sulphate; organic phosphates; or, clays such as Benonite or Veegum®. Other examples of hydrocolloid compositions include those in U.S. Patent Publication 2009/0317467, which is hereby incorporated by reference.

[00223] Other patents discuss viscosity enhancers such as U.S. Patent 5,405,741, which is hereby incorporated by reference. It describes that suitable organic solvent diluents include, for example, alcohols such as methanol, ethanol, isopropanol, butanol, sec-butyl alcohol, and the like; ethers such as dimethyl ether, ethyl methyl ether, diethyl ether, 1-ethoxypropane, and the like; tetrahydrofuran; glycols such as 1,2-ethanediol, 1,2-propanediol, 1,3-propanediol, and the like; ketones such as acetone, methylethylketone, 3-pentanone, methylisobutylketone, and the like; esters such as ethyl formate, methyl acetate, ethyl acetate, butyl acetate, ethyl propionate, b-ethoxyethylacetate, methyl Cellosolve acetate, and the like; amides such as formamide, acetamide, succinic amide, and the like; alkyl esters of a suitable acid such as phthalic acid, including methyl

phthalate, ethyl phthalate, propyl phthalate, n-butyl phthalate, di-n-butyl phthalate, n-amyl phthalate, isoamyl phthalate, dioctyl phthalate and the like; alkyl amides such as N,N-diethyllaurylamide and the like; trimellitic acid esters including tri-tertoctyl mellitate and the like; phosphoric acid esters including polyphenyl phosphate, tricresyl phosphate, dioctylbutyl phosphate and the like; citric acid esters such as acetyl tributyl citrate and the like; and mixtures thereof. In some embodiments there are aqueous compositions containing at least about 80% by weight of water, at least about 90, or up to about 20% by weight of the composition of an organic solvent, or a maximum of about 10%. In some embodiments, aqueous compositions are free of organic solvent. Suitable hydrophilic colloidal materials include both naturally occurring substances such as proteins, protein derivatives, cellulose derivatives such as cellulose esters; gelatin including alkali-treated and acid-treated gelatin, phthalated gelatin, and the like; polysaccharides such as dextran, gum arabic, zein, casein, pectin, collagen derivatives, collodion, agar-agar, arrowroot, albumin and the like. Generally, it is preferred that the aqueous gelatin composition contains at least about 2% by weight of the composition of gelatin; most preferred are aqueous gelatin compositions in which a gelatin is the hydrophilic colloid. Other hydrophilic colloidal materials that can be used include poly(vinyllactams), acrylamide polymers, polyvinyl alcohol and its derivatives, polyvinyl acetals, polymers of alkyl and sulfoalkyl acrylates and methacrylates, hydrolyzed polyvinyl acetates, polyamides, polyvinyl pyridine, acrylic acid polymers, maleic anhydride copolymers, polyalkylene oxides, methacrylamide copolymers, polyvinyl oxazolidinones, maleic acid copolymers, vinylamine copolymers, methacrylic acid copolymers, acryloyloxyalkylsulfonic acid copolymers, sulfoalkylacrylamide copolymers, polyalkyleneimine copolymers, polyamines, N,N-dialkylaminoalkyl acrylates, vinyl imidazole copolymers, vinyl sulfide copolymers, halogenated styrene polymers, amineacrylamide polymers, polypeptides and the like. Other exemplary colloids are disclosed, for example in U.S. Pat. Nos. 2,691,582; 2,787,545; 2,956,880; 3,132,945; 3,138,846; 3,679,425; 3,706,564; 3,813,251; 3,852,073; 3,879,205; 3,003,879; 3,284,207; 3,748,143; 3,536,491 and the like, the disclosures of which are hereby incorporated herein by reference. A copolymer of any suitable alkali metal or ammonium salt of a sulfonic acid containing monomer with any suitable unsaturated monomer, preferably having a number average molecular weight greater than about 300,000, can be used as the viscosity enhancing agent or thickener. Any suitable method can be used to prepare the viscosity enhancing polymers as is known in the art. For example, any suitable base can be reacted with any suitable ester to form the alkali metal

or ammonium salt of the sulfonic acid containing copolymer including acryloyl-oxymethyl bisulfite, acryloyloxymethyl bisulfate, methacryloyloxymethyl bisulfite, methacryloyloxymethyl bisulfate, acryloyloxyethyl bisulfite, acryloyloxyethyl bisulfate, methacryloyloxyethyl bisulfite, methacryloyloxyethyl bisul-fate, acryloyloxypropyl bisulfite, acryloyloxypropyl bisul-fate, methacryloyloxypropyl bisulfite, methacryloyloxypropyl bisulfate, acryloyloxybutyl bisulfite, acryloyloxybutyl bisulfate, methacryloyloxybutyl bisulfite methacryloylxybutyl bisulfate, and the like. The corresponding salt that is reacted with an unsaturated monomer to prepare one or more copolymers can be obtained by reacting the ester with a base as is well known.

[00224] It will be understood by a person of ordinary skill in the art that methods may involve a number of steps, some of which may be repeated throughout a protocol. It is contemplated that one or more of these steps may be repeated, be repeated at least, or be repeated at most 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 time or more (or any range derivable therein). In some embodiments, methods involve or involve at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more of the following additional steps (above the rolling circle amplification steps) in the same or a similar order to achieve detection, identification or characterization of a nucleic acid sequence in a biological sample, including one that has been prepared for analysis on a slide:

1) dewaxing, such as to remove wax from tissue;
2) alcohol rinsing, such as to remove dewaxing reagents and/or hydrate sample;
3) washing the sample,
4) heating the sample, such as to induce epitope retrieval (HIER);
5) washing the sample;
6) incubating with a peroxidase block, such as hydrogen peroxide or methanol to reduce or eliminate endogenous peroxidase activity in a sample
7) washing the sample;
8) incubating with an antibody or antibody fragment under conditions to allow the antibody or antibody fragment to bind a label on a detection probe;
9) washing the sample;
10) incubating the sample with a secondary antibody that binds any antibody or antibody fragment that binds to a label on a detection probe;
11) washing the sample

12) detecting any label on the secondary antibody

[00225] In particular embodiments, instead of employing an antibody in step 8), a reagent that reacts with a detection probe label is incubated under conditions to detect the label. For instance, in some embodiments, a detection probe is labeled with horse radish peroxidase (HRP) (such as by conjugation), and detection of the label is achieved by incubating the sample with a reagent that allows detection of the label, such as 3,3'-diaminobenzidine tetrahydrochloride (DAB). In other embodiments, a detection probe may be labeled with alkaline phosphatase (AP), which may be detected directly or indirectly (such as with an antibody, that may or may not be labeled but is capable of detection).

[00226] In certain embodiments, all of these steps are employed. A sample may be heated at one or multiple times during a process. Heating a sample may be employed to inactivate one or more enzyme or reagents. In other embodiments, heating is used prior to immunohistochemistry (IHC) or in situ hybridization (ISH) to improve staining. Embodiments involve heating a sample to about, at least about, or at most about 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, or 125 °C (and any range derivable therein) for about, at least about, or at most about 30 seconds, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55 minutes, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 hours (and any range derivable therein). A person of ordinary skill in the art will know how temperatures and times may be varied depending on factors that include but are not limited to the sample, volume of the sample and volume of liquid, surface area, reagents present, target of the heating, pH, etc. It is contemplated that in some embodiments, one or more additional enzymes may be employed in conjunction with heating, such as to inactivate one or more enzymes the sample or to make a sample more accessible. In some embodiments, an enzyme that reduces or minimizes protein crosslinking in the sample is employed.

E. *KRAS*

[00227] As described in more detail herein, methods may be used to detect a point mutation in the mRNA sequence that codes for *KRAS*. *KRAS* is one of the most frequently activated oncogenes. As used herein, "*KRAS*" refers to v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog. *KRAS* is also known in the art as *NS3, KRASl, KRAS2, RASK2, KI-RAS, C-K-RAS, K-RAS2A, K-RAS2B, K-RAS4A* and *K-RAS4B*. This gene, a Kirsten ras oncogene homolog

from the mammalian ras gene family, encodes a protein that is a member of the small GTPase superfamily. A single amino acid substitution can be responsible for an activating mutation. The transforming protein that results can be implicated in various malignancies, including lung cancer, colon cancer, thyroid cancer and pancreatic cancer and is strongly associated with resistance to epidermal growth factor receptor (EGFR) inhibitor therapy. For example, in metastatic colorectal cancer the presence of mutations in the KRAS gene is routinely analyzed, and a positive mutation status indicates that the tumor will not respond to EGFR antibody therapy. In lung adenocarcinoma KRAS mutations are associated with smoking, poor prognosis and non-responsiveness to EGFR tyrosine kinase inhibitors (TKI) whereas KRAS wild-type tumors with EGFR mutations are linked to non-smoking, better prognosis and response to EGFR-TKI therapy.

[00228] A tumor may have one or more mutations in KRAS (e.g., an activating mutation), unwanted expression of KRAS (e.g., overexpression over wild type), KRAS deficiency, and/or amplification of KRAS gene (e.g., having more than two functional copies of KRAS gene). There are seven point mutations in codon 12 and 13 that together account for more than 95% of all KRAS mutations. Conventional KRAS analysis is based on DNA extracted from crude tumor tissue, and after PCR amplification of the hot spot region on exon 1 the sequence aberrations in codon 12 and 13 are characterized by direct dideoxy sequencing or by more sensitive targeted assays such as Pyrosequencing or allele-specific PCR. Thus, all different cell types present in a tumor sample – normal parenchymal cells, stromal cells, inflammatory cells, different pre-neoplastic and neoplastic sub-clones – contribute their wild-type and mutated KRAS alleles to these assays. In the routine diagnostic setting tumor cells can be enriched for by manual microdissection, but in order to annotate a mutation to a certain tumor sub compartment the required dissection is laborious. Still, single cell resolution is extremely difficult to achieve. This might not be a problem in colorectal cancer as activating KRAS mutations are considered to be early events in tumorigenesis and presumably homogenously distributed in the tumor. However, for other types of cancer, and for mutations in other oncogenes, very little is known about heterogeneity among cancer sub-clones and its impact on tumor biology and treatment response. Therefore, methods which offer genotyping directly on tissue sections are highly warranted. Hence there is a requirement for sensitive KRAS mutation analysis to determine the most suitable treatment for the patients.

[00229] As described herein the present method may be used in a genotyping assay that targets KRAS-mutations in codon 12 and 13 in situ on tissue samples by the use of multiple mutation specific padlock probes and rolling-circle amplification. Such an in situ technique offers single transcript analysis directly in tissues and thus circumvents traditional DNA extraction from heterogeneous tumor tissues. In addition, or alternatively, mutations in codon 61 and/or codon 146 of KRAS may be targeted (for specific information see also Loupakis et al., 2009, Br J Cancer, 101(4): 715-21, which is incorporated herein by reference in its entirety). Furthermore, mutations in the 3' UTR of KRAS transcripts may be targeted (for specific information see also Graziano et al., 2010, Pharmacogenomics J., doi 10.1038/tpj.2010.9, which is incorporated herein by reference in its entirety). These mutations may be detected in combination with a detection of codon 12, 13, 61 and/or 146 mutations, or they may be detected alone, or in combination with codon 12 mutations, or with codon 13 mutations, or with codon 61 mutations, or with codon 146 mutation, or with any subgrouping of codon 12, 13, 61 and 146 mutations. Methods may be carried out in fresh frozen or formalin-fixed, paraffin-embedded (FFPE) tissue, or in tissues in touch imprint samples. In some embodiments, tissue samples may be cancer tissue, e.g. colon or lung tissues.

[00230] In some embodiments, methods and compositions concern KRAS mutations, particularly those mutations that have been found in cancer cells. The term "KRAS mutation associated with cancer" or "KRAS mutant associated with tumor development" refers to a mutation in the KRAS gene or a corresponding mutant, that has been identified in the Sanger database as of February 15, 2012 as associated with cancer or precancer (on the world wide web at sanger.ac.uk). In certain embodiments, the methods and compositions concern detecting a plurality of mutations. In some embodiments a plurality of mutations refers to at least or at most the following percentage of mutations in that gene associated with cancer: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 or 100%, or any range derivable therein.

F. EGFR

[00231] The epidermal growth factor receptor (EGFR) is an important target in the treatment of some cancers. The combination of anti-EGFR antibodies with chemotherapy is thus commonly used in the treatment of these cancers. The *KRAS* protein is an important mediator in the signal transduction cascade regulated by the EGFR. Mutations in the *KRAS* gene are a

very important factor in the selection of molecular biological treatment options targeted against EGFR. Studies have shown that if the mutation is present, anti-EGFR medications such as cetuximab (Erbitux) and panitumumab (Vectibix) are not sufficiently effective to warrant their use. Thus, as discussed herein, the present method may advantageously be used to detect the presence or absence of a point mutation in the mRNA which codes for *KRAS*, wherein the identification of *KRAS* wild-type mRNA indicates that the cancer may be treated with EGFR inhibitors.

[00232] In addition, the present method may be used to detect the presence or absence of a mutation in the mRNA which codes for the EGFR. Examples of EGFR mutations that may be detected according to various embodiments are shown in Table 7.

G. ***Braf, APC, PTEN, PI3K***

[00233] Methods may be further used to detect one or more point mutations in the mRNA sequence that codes for *Braf, APC, PTEN* or *PI3K*. Suitable *Braf* mutations are known to the skilled person and are described in Rajagopalan et al., 2002, Nature, 418 (29), 934 and Monticone et al., 2008, Molecular Cancer, 7(92), which are incorporated herein by reference in their entirety. Particularly preferred is the detection of mutation V600E. In some embodiments, methods further involve the detection of one or more point mutations in KRAS and Braf. Braf and KRAS mutations are described as being mutually exclusive regarding the function of downstream pathway elements. Thus, by determining mutations in Braf and KRAS at the same time, it may be elucidate whether and pathway functions are compromised by genetic mutations.

[00234] Suitable APC mutations are known to the skilled person and are described, for example, in Vogelstein and Fearon, 1988, N Engl J Med, 319(9): 525-32, which is incorporated herein by reference in its entirety.

[00235] Suitable PTEN mutations are known to the skilled person and are described, for example, in Laurent-Puig et al, 2009, J Clon Oncol, 27(35), 5924-30 or Loupakis et al., 2009, J clin Oncol, 27(16), 2622-9, which are incorporated herein by reference in their entirety.

[00236] Suitable PI3K mutations are known to the skilled person and are described, for example, in Satore-Bianchi et al., 2009, Cancer Res., 69(5), 1851-7 or Prenen et al., 2009, Clin Cancer Res., 15(9), 3184-8, which are incorporated herein by reference in their entirety.

[00237] In some embodiments, methods and compositions concern Braf, APC, PTEN or PI3K mutations. In further embodiments, methods and compositions concern KRAS mutations in combination with Braf mutations, and/or in combination with APC mutations, and/or in combination with PTEN mutations, and/or in combination with PI3K mutations, particularly those mutations that have been found in cancer cells. Such mutations may be derived from suitable literature sources, e.g. those mentioned above, or may be identified according to suitable databases, e.g., the Sanger database as of February 15, 2012 (on the world wide web at sanger.ac.uk). In certain embodiments, the methods and compositions concern detecting a plurality of the mutations. In some embodiments a plurality of mutations refers to at least or at most the following percentage of mutations in that the gene or gene combination associated with cancer: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 or 100%, or any range derivable therein.

[00238] Other embodiments concern characterization, detection, sequencing, and/or analyzing one or more cancer genes shown below. In certain embodiments, there is one or more padlock probes with arms that flank a cancer mutation. In further embodiments, there is one or more padlock probes with the arms discussed above and a sequence(s) at the terminal end of one of the arms that is complementary or identical to a mutation sequence (whether the mutation is a single nucleotide chain or a mutation involving a deletion, insertion, alteration of multiple nucleotides).

Table B

Symbol	Name	GeneID	Chr Band	Tumor Types (Somatic Mutations)	Cancer Syndrome	Tissue Type*	Mutation Type*
ABL1	v-abl Abelson murine leukemia viral oncogene homolog 1	25	9q34.1	CML, ALL, T-ALL		L	T, Mis
ABL2	v-abl Abelson murine leukemia viral oncogene homolog 2	27	1q24-q25	AML		L	T
ACSL3	acyl-CoA synthetase long-chain family member 3	2181	2q36	prostate		E	T
AF15Q14	AF15q14 protein	57082	15q14	AML		L	T
AF1Q	ALL1-fused gene from	10962	1q21	ALL		L	T

	chromosome 1q						
AF3p21	SH3 protein interacting with Nck, 90 kDa (ALL1 fused gene from 3p21)	51517	3p21	ALL		L	T
AF5q31	ALL1 fused gene from 5q31	27125	5q31	ALL		L	T
AKAP9	A kinase (PRKA) anchor protein (yotiao) 9	10142	7q21-q22	papillary thyroid		E	T
AKT1	v-akt murine thymoma viral oncogene homolog 1	207	14q32.32	breast, colorectal, ovarian, NSCLC		E	Mis
AKT2	v-akt murine thymoma viral oncogene homolog 2	208	19q13.1-q13.2	ovarian, pancreatic		E	A
ALDH2	aldehyde dehydrogenase 2 family (mitochondrial)	217	12q24.2	leiomyoma		M	T
ALK	anaplastic lymphoma kinase (Ki-1)	238	2p23	ALCL, NSCLC, Neuroblastoma	Familial neuroblastoma	L, E, M	T, Mis, A
ALO17	KIAA1618 protein	57714	17q25.3	ALCL		L	T
APC	adenomatous polyposis of the colon gene	324	5q21	colorectal, pancreatic, desmoid, hepatoblastoma glioma, other CNS	Adenomatous polyposis coli; Turcot syndrome	E, M, O	D, Mis, N, F, S
ARHGEF12	RHO guanine nucleotide exchange factor (GEF) 12 (LARG)	23365	11q23.3	AML		L	T
ARHH	RAS homolog gene family, member H (TTF)	399	4p13	NHL		L	T
ARID1A	AT rich interactive domain 1A (SWI-like)	8289	1p35.3	clear cell ovarian carcinoma, RCC		E	Mis, N, F, S, D
ARID2	AT rich interactive domain 2	196528	12q12	hepatocellular carcinoma		E	N, S, F
ARNT	aryl hydrocarbon receptor nuclear translocator	405	1q21	AML		L	T
ASPSCR1	alveolar soft part sarcoma chromosome region, candidate 1	79058	17q25	alveolar soft part sarcoma		M	T

ASXL1	additional sex combs like 1	171023	20q11.1	MDS, CMML		L	F, N, Mis
ATF1	activating transcription factor 1	466	12q13	malignant melanoma of soft parts , angiomatoid fibrous histiocytoma		E, M	T
ATIC	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	471	2q35	ALCL		L	T
ATM	ataxia telangiectasia mutated	472	11q22.3	T-PLL	Ataxia-telangiectasia	L, O	D, Mis, N, F, S
ATRX	alpha thalassemia/mental retardation syndrome X-linked	546	Xq21.1	Pancreatic neuroendocrine tumors, paediatric GBM		E	Mis, F, N
BAP1	BRCA1 associated protein-1 (ubiquitin carboxy-terminal hydrolase)	8314	3p21.31-p21.2	uveal melanoma, breast, NSCLC, RCC		E	N, Mis, F, S, O
BCL10	B-cell CLL/lymphoma 10	8915	1p22	MALT		L	T
BCL11A	B-cell CLL/lymphoma 11A	53335	2p13	B-CLL		L	T
BCL11B	B-cell CLL/lymphoma 11B (CTIP2)	64919	14q32.1	T-ALL		L	T
BCL2	B-cell CLL/lymphoma 2	596	18q21.3	NHL, CLL		L	T
BCL3	B-cell CLL/lymphoma 3	602	19q13	CLL		L	T
BCL5	B-cell CLL/lymphoma 5	603	17q22	CLL		L	T
BCL6	B-cell CLL/lymphoma 6	604	3q27	NHL, CLL		L	T, Mis
BCL7A	B-cell CLL/lymphoma 7A	605	12q24.1	BNHL		L	T
BCL9	B-cell CLL/lymphoma 9	607	1q21	B-ALL		L	T
BCOR	BCL6 corepressor	54880	Xp11.4	retinoblastoma, AML, APL(translocation)			F, N, S, T
BCR	breakpoint cluster region	613	22q11.21	CML, ALL, AML		L	T

BHD	folliculin, Birt-Hogg-Dube syndrome	201163	17p11.2		Birt-Hogg-Dube syndrome	E, M	Mis. N, F
BIRC3	baculoviral IAP repeat-containing 3	330	11q22-q23	MALT		L	T
BLM	Bloom Syndrome	641	15q26.1		Bloom Syndrome	L, E	Mis, N, F
BMPR1A	bone morphogenetic protein receptor, type IA	657	10q22.3		Juvenile polyposis	E	Mis, N, F
BRAF	v-raf murine sarcoma viral oncogene homolog B1	673	7q34	melanoma, colorectal, papillary thyroid, borderline ov, Non small-cell lung cancer (NSCLC), cholangiocarcinoma, pilocytic astrocytoma		E	Mis, T, O
BRCA1	familial breast/ovarian cancer gene 1	672	17q21	ovarian	Hereditary breast/ovarian cancer	E	D, Mis, N, F, S
BRCA2	familial breast/ovarian cancer gene 2	675	13q12	breast, ovarian, pancreatic	Hereditary breast/ovarian cancer	L, E	D, Mis, N, F, S
BRD3	bromodomain containing 3	8019	9q34	lethal midline carcinoma of young people		E	T
BRD4	bromodomain containing 4	23476	19p13.1	lethal midline carcinoma of young people		E	T
BRIP1	BRCA1 interacting protein C-terminal helicase 1	83990	17q22		Fanconi anaemia J, breast cancer susceptiblity	L, E	F, N, Mis
BTG1	B-cell translocation gene 1, anti-proliferative	694	12q22	BCLL		L	T
BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	701	15q15		Mosaic variegated aneuploidy	M	Mis, N, F, S
C12orf9	chromosome 12 open reading frame 9	93669	12q14.3	lipoma		M	T
C15orf21	chromosome 15 open reading frame 21	283651	15q21.1	prostate		E	T

C15orf55	chromosome 15 open reading frame 55	256646	15q14	lethal midline carcinoma		E	T
C16orf75	chromosome 16 open reading frame 75	116028	16p13.13	PMBL, Hodgkin Lymphona,		L	T
C2orf44	chromosome 2 open reading frame 44	80304	2p23.3	NSCLC		E	T
CAMTA1	calmodulin binding transcription activator 1	611501	1p36.31-p36.23	epitheliod hemangioendothelioma		M	T
CANT1	calcium activated nucleotidase 1	124583	17q25	prostate		E	T
CARD11	caspase recruitment domain family, member 11	84433	7p22	DLBCL		L	Mis
CARS	cysteinyl-tRNA synthetase	833	11p15.5	ALCL		L	T
CBFA2T1	core-binding factor, runt domain, alpha subunit 2;translocated to, 1 (ETO)	862	8q22	AML		L	T
CBFA2T3	core-binding factor, runt domain, alpha subunit 2; translocated to, 3 (MTG-16)	863	16q24	AML		L	T
CBFB	core-binding factor, beta subunit	865	16q22	AML		L	T
CBL	Cas-Br-M (murine) ecotropic retroviral transforming	867	11q23.3	AML, JMML, MDS		L	T, Mis S, O
CBLB	Cas-Br-M (murine) ecotropic retroviral transforming sequence b	868	3q13.11	AML		L	Mis S
CBLC	Cas-Br-M (murine) ecotropic retroviral transforming sequence c	23624	19q13.2	AML		L	M
CCDC6	coiled-coil domain containing 6	8030	10q21	NSCLC		E	T
CCNB1IP1	cyclin B1 interacting protein 1, E3 ubiquitin protein ligase	57820	14q11.2	leiomyoma		M	T
CCND1	cyclin D1	595	11q13	CLL, B-ALL,		L, E	T

				breast			
CCND2	cyclin D2	894	12p13	NHL,CLL		L	T
CCND3	cyclin D3	896	6p21	MM		L	T
CCNE1	cyclin E1	898	19q12	serous ovarian		E	A
CD273	programmed cell death 1 ligand 2	80380	9p24.2	PMBL, Hodgkin Lymphona,		L	T
CD274	CD274 molecule	29126	9p24	PMBL, Hodgkin Lymphona,		L	T
CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	972	5q32	NSCLC		E	T
CD79A	CD79a molecule, immunoglobulin-associated alpha	973	19q13.2	DLBCL		L	O, S
CD79B	CD79b molecule, immunoglobulin-associated beta	974	17q23	DLBCL		L	Mis, O
CDH1	cadherin 1, type 1, E-cadherin (epithelial) (ECAD)	999	16q22.1	lobular breast, gastric	Familial gastric carcinoma	E	Mis, N, F, S
CDH11	cadherin 11, type 2, OB-cadherin (osteoblast)	1009	16q22.1	aneurysmal bone cysts		M	T
CDK12	cyclin-dependent kinase 12	51755	17q12	serous ovarian		E	Mis, N, F
CDK4	cyclin-dependent kinase 4	1019	12q14		Familial malignant melanoma	E	Mis
CDK6	cyclin-dependent kinase 6	1021	7q21-q22	ALL		L	T
CDKN2A	cyclin-dependent kinase inhibitor 2A (p16(INK4a)) gene	1029	9p21	melanoma, multiple other tumour types	Familial malignant melanoma	L, E, M, O	D, Mis, N, F, S
CDKN2a(p14)	cyclin-dependent kinase inhibitor 2A-- p14ARF protein	1029	9p21	melanoma, multiple other tumour types	Familial malignant melanoma	L, E, M, O	D, S
CDKN2C	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	1031	1p32	glioma, MM		O, L	D

CDX2	caudal type homeo box transcription factor 2	1045	13q12.3	AML		L	T
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha	1050	19q13.1	AML, MDS		L	Mis, N, F
CEP1	centrosomal protein 1	11064	9q33	MPD, NHL		L	T
CHCHD7	coiled-coil-helix-coiled-coil-helix domain containing 7	79145	8q11.2	salivary adenoma		E	T
CHEK2	CHK2 checkpoint homolog (S. pombe)	11200	22q12.1		familial breast cancer	E	F
CHIC2	cysteine-rich hydrophobic domain 2	26511	4q11-q12	AML		L	T
CHN1	chimerin (chimaerin) 1	1123	2q31-q32.1	extraskeletal myxoid chondrosarcoma		M	T
CIC	capicua homolog	23152	19q13.2	oligodendroglioma		O	Mis, F, S
CIITA	class II, major histocompatibility complex, transactivator	4261	16p13	PMBL, Hodgkin Lymphona,		L	T
CLTC	clathrin, heavy polypeptide (Hc)	1213	17q11-qter	ALCL, renal		L	T
CLTCL1	clathrin, heavy polypeptide-like 1	8218	22q11.21	ALCL		L	T
CMKOR1	chemokine orphan receptor 1	57007	2q37.3	lipoma		M	T
COL1A1	collagen, type I, alpha 1	1277	17q21.31-q22	dermatofibrosarcoma protuberans, aneurysmal bone cyst		M	T
COPEB	core promoter element binding protein (KLF6)	1316	10p15	prostate, glioma		E, O	Mis, N
COX6C	cytochrome c oxidase subunit VIc	1345	8q22-q23	uterine leiomyoma		M	T
CREB1	cAMP responsive element binding protein 1	1385	2q34	clear cell sarcoma, angiomatoid fibrous histiocytoma		M	T
CREB3L1	cAMP responsive element binding protein 3-like 1	90993	11p11.2	myxofibrosarcoma		M	T
CREB3L2	cAMP responsive	64764	7q34	fibromyxoid sarcoma		M	T

	element binding protein 3-like 2						
CREBBP	CREB binding protein (CBP)	1387	16p13.3	ALL, AML, DLBCL, B-NHL		L	T, N, F, Mis, O
CRLF2	cytokine receptor-like factor 2	64109	Xp22.3; Yp11.3	B-ALL, Downs associated ALL		L	Mis, T
CRTC3	CREB regulated transcription coactivator 3	64784	15q26.1	salivary gland mucoepidermoid		E	T
CTNNB1	catenin (cadherin-associated protein), beta 1	1499	3p22-p21.3	colorectal, cvarian, hepatoblastoma, others, pleomorphic salivary adenoma		E, M, O	H, Mis, T
CYLD	familial cylindromatosis gene	1540	16q12-q13	cylindroma	Familial cylindromatosis	E	Mis, N, F, S
D10S170	DNA segment on chromosome 10 (unique) 170, H4 gene (PTC1)	8030	10q21	papillary thyroid, CML		E	T
DAXX	death-domain associated protein	1616	6p21.3	Pancreatic neuroendocrine tumors. Paediatric GBM		E	Mis, F, N
DDB2	damage-specific DNA binding protein 2	1643	11p12		Xeroderma pigmentosum (E)	E	Mis, N
DDIT3	DNA-damage-inducible transcript 3	1649	12q13.1-q13.2	liposarcoma		M	T
DDX10	DEAD (Asp-Glu-Ala-Asp) box polypeptide 10	1662	11q22-q23	AML*		L	T
DDX5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5	1655	17q21	prostate		E	T
DDX6	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6	1656	11q23.3	B-NHL		L	T
DEK	DEK oncogene (DNA binding)	7913	6p23	AML		L	T
DICER1	dicer 1, ribonuclease type III	23405	14q32.13	sex cord-stromal tumour, TGCT, embryonal rhadomyosarcoma	Familial Pleuropulmonary Blastoma	E, M, O	Mis F, N

DNM2	dynamin 2	1785	19p13.2	ETP ALL		L	F, N, Splice, Mis, O
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	1788	2p23	AML		L	Mis, F, N, S
DUX4	double homeobox, 4	22947	4q35	soft tissue sarcoma		M	T
EBF1	early B-cell factor 1	1879	5q34	lipoma		M	T
ECT2L	epithelial cell transforming sequence 2 oncogene-like	345930	6q24.1	ETP ALL		L	N, Splice, Mis
EGFR	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	1956	7p12.3-p12.1	glioma, NSCLC	Familial lung cancer	E, O	A, O, Mis
EIF4A2	eukaryotic translation initiation factor 4A, isoform 2	1974	3q27.3	NHL		L	T
ELF4	E74-like factor 4 (ets domain transcription factor)	2000	Xq26	AML		L	T
ELK4	ELK4, ETS-domain protein (SRF accessory protein 1)	2005	1q32	prostate		E	T
ELKS	ELKS protein	23085	12p13.3	papillary thyroid		E	T
ELL	ELL gene (11-19 lysine-rich leukemia gene)	8178	19p13.1	AL		L	T
ELN	elastin	2006	7q11.23	B-ALL		L	T
EML4	echinoderm microtubule associated protein like 4	27436	2p21	NSCLC		E	T
EP300	300 kd E1A-Binding protein gene	2033	22q13	colorectal, breast, pancreatic, AML, ALL, DLBCL		L, E	T, N, F, Mis, O
EPS15	epidermal growth factor receptor pathway substrate 15 (AF1p)	2060	1p32	ALL		L	T
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblast	2064	17q21.1	breast, ovarian, other tumour types, NSCLC, gastric		E	A, Mis, O

	oma derived oncogene homolog (avian)						
ERCC2	excision repair cross-complementing rodent repair deficiency, complementation group 2 (xeroderma pigmentosum D)	2068	19q13.2-q13.3		Xeroderma pigmentosum (D)	E	Mis, N, F, S
ERCC3	excision repair cross-complementing rodent repair deficiency, complementation group 3 (xeroderma pigmentosum group B complementing)	2071	2q21		Xeroderma pigmentosum (B)	E	Mis, S
ERCC4	excision repair cross-complementing rodent repair deficiency, complementation group 4	2072	16p13.3-p13.13		Xeroderma pigmentosum (F)	E	Mis, N, F
ERCC5	excision repair cross-complementing rodent repair deficiency, complementation group 5 (xeroderma pigmentosum, complementation group G (Cockayne syndrome))	2073	13q33		Xeroderma pigmentosum (G)	E	Mis, N, F
ERG	v-ets erythroblastosis virus E26 oncogene like (avian)	2078	21q22.3	Ewing sarcoma, prostate, AML		M, E, L	T
ETV1	ets variant gene 1	2115	7p22	Ewing sarcoma, prostate		M, E	T
ETV4	ets variant gene 4 (E1A enhancer binding protein, E1AF)	2118	17q21	Ewing sarcoma, Prostate carcinoma		M, E	T
ETV5	ets variant gene 5	2119	3q28	Prostate		E	T

ETV6	ets variant gene 6 (TEL oncogene)	2120	12p13	congenital fibrosarcoma, multiple leukemia and lymphoma, secretory breast, MDS, ALL		L, E, M	T
EVI1	ecotropic viral integration site 1	2122	3q26	AML, CML		L	T
EWSR1	Ewing sarcoma breakpoint region 1 (EWS)	2130	22q12	Ewing sarcoma, desmoplastic small round cell tumor , ALL, clear cell sarcoma, sarcoma, myoepitheliom a		L, M	T
EXT1	multiple exostoses type 1 gene	2131	8q24.1 1-q24.13		Multiple Exostoses Type 1	M	Mis, N, F, S
EXT2	multiple exostoses type 2 gene	2132	11p12-p11		Multiple Exostoses Type 2	M	Mis, N, F, S
EZH2	enhancer of zeste homolog 2	2146	7q35-q36	DLBCL		L	Mis
EZR	ezrin	7430	6q25.3	NSCLC		E	T
FACL6	fatty-acid-coenzyme A ligase, long-chain 6	23305	5q31	AML, AEL		L	T
FAM22A	family with sequence similarity 22, member A	728118	10q23.2	edometrial stromal sarcoma		M	T
FAM22B	family with sequence similarity 22, member B	729262	10q22.3	edometrial stromal sarcoma		M	T
FAM46C	family with sequence similarity 46, member C	54855	1p12	MM		L	Mis, F, O
FANCA	Fanconi anemia, complementation group A	2175	16q24.3		Fanconi anaemia A	L	D, Mis, N, F, S
FANCC	Fanconi anemia, complementation group C	2176	9q22.3		Fanconi anaemia C	L	D, Mis, N, F, S
FANCD2	Fanconi anemia, complementation group D2	2177	3p26		Fanconi anaemia D2	L	D, Mis, N, F
FANCE	Fanconi anemia, complementation group E	2178	6p21-p22		Fanconi anaemia E	L	N, F, S

FANCF	Fanconi anemia, complementation group F	2188	11p15		Fanconi anaemia F	L	N, F
FANCG	Fanconi anemia, complementation group G	2189	9p13		Fanconi anaemia G	L	Mis, N, F, S
FBXO11	F-box protein 11	80204	2p16.3	DLBCL		L	Mis, F, D
FBXW7	F-box and WD-40 domain protein 7 (archipelago homolog, Drosophila)	55294	4q31.3	colorectal, endometrial, T-ALL		E, L	Mis, N, D, F
FCGR2B	Fc fragment of IgG, low affinity IIb, receptor for (CD32)	2213	1q23	ALL		L	T
FEV	FEV protein - (HSRNAFEV)	54738	2q36	Ewing sarcoma		M	T
FGFR1	fibroblast growth factor receptor 1	2260	8p11.2 -p11.1	MPD, NHL		L	T
FGFR1OP	FGFR1 oncogene partner (FOP)	11116	6q27	MPD, NHL		L	T
FGFR2	fibroblast growth factor receptor 2	2263	10q26	gastric. NSCLC, endometrial		E	Mis
FGFR3	fibroblast growth factor receptor 3	2261	4p16.3	bladder, MM, T-cell lymphoma		L, E	Mis, T
FH	fumarate hydratase	2271	1q42.1		hereditary leiomyomatosis and renal cell cancer	E, M	Mis, N, F
FHIT	fragile histidine triad gene	2272	3p14.2	pleomorphic salivary gland adenoma		E	T
FIP1L1	FIP1 like 1 (S. cerevisiae)	81608	4q12	idiopathic hypereosinophilic syndrome		L	T
FLI1	Friend leukemia virus integration 1	2313	11q24	Ewing sarcoma		M	T
FLJ27352	BX648577, FLJ27352 hypothetical LOC145788	145788	15q21.3	PMBL, Hodgkin Lymphona,		L	T
FLT3	fms-related tyrosine kinase 3	2322	13q12	AML, ALL		L	Mis, O
FNBP1	formin binding protein 1 (FBP17)	23048	9q23	AML		L	T
FOXL2	forkhead box L2	668	3q23	granulosa-cell tumour of the ovary		O	Mis
FOXO1A	forkhead box O1A (FKHR)	2308	13q14.1	alveolar rhabdomyosarcomas		M	T

FOXO3A	forkhead box O3A	2309	6q21	AL		L	T
FOXP1	forkhead box P1	27086	3p14.1	ALL		L	T
FSTL3	follistatin-like 3 (secreted glycoprotein)	10272	19p13	B-CLL		L	T
FUBP1	far upstream element (FUSE) binding protein 1	8880	1p13.1	oligodendroglioma		O	F, N
FUS	fusion, derived from t(12;16) malignant liposarcoma	2521	16p11.2	liposarcoma, AML, Ewing sarcoma, angiomatoid fibrous histiocytoma, fibromyxoid sarcoma		M, L	T
FVT1	follicular lymphoma variant translocation 1	2531	18q21.3	B-NHL		L	T
GAS7	growth arrest-specific 7	8522	17p	AML*		L	T
GATA1	GATA binding protein 1 (globin transcription factor 1)	2623	Xp11.23	megakaryoblastic leukemia of Downs Syndrome		L	Mis, F
GATA2	GATA binding protein 2	2624	3q21.3	AML(CML blast transformation)		L	Mis
GATA3	GATA binding protein 3	2625	10p15	breast		E	F, N, S
GMPS	guanine monphosphate synthetase	8833	3q24	AML		L	T
GNA11	guanine nucleotide binding protein (G protein), alpha 11 (Gq class)	2767	19p13.3	uveal melanoma		E	Mis
GNAQ	guanine nucleotide binding protein (G protein), q polypeptide	2776	9q21	uveal melanoma		E	Mis
GNAS	guanine nucleotide binding protein (G protein), alpha stimulating activity polypeptide 1	2778	20q13.2	pituitary adenoma		E	Mis
GOLGA5	golgi autoantigen, golgin subfamily a, 5 (PTC5)	9950	14q	papillary thyroid		E	T
GOPC	golgi associated	57120	6q21	glioblastoma		O	O

	PDZ and coiled-coil motif containing						
GPC3	glypican 3	2719	Xq26.1		Simpson-Golabi-Behmel syndrome	O	T, D, Mis, N, F, S
GPHN	gephyrin (GPH)	10243	14q24	AL		L	T
GRAF	GTPase regulator associated with focal adhesion kinase pp125(FAK)	23092	5q31	AML, MDS		L	T, F, S
H3F3A	H3 histone, family 3A	3020	1q42.12	glioma		O	Mis
HCMOGT-1	sperm antigen HCMOGT-1	92521	17p11.2	JMML		L	T
HEAB	ATP_GTP binding protein	10978	11q12	AML		L	T
HERPUD1	homocysteine-inducible, endoplasmic reticulum stress-inducible, ubiquitin-like domain member 1	9709	16q12.2-q13	prostate		E	T
HEY1	hairy/enhancer-of-split related with YRPW motif 1	23462	8q21	mesenchymal chondrosarcoma		M	T
HIP1	huntingtin interacting protein 1	3092	7q11.23	CMML		L	T
HIST1H4I	histone 1, H4i (H4FM)	8294	6p21.3	NHL		L	T
HLF	hepatic leukemia factor	3131	17q22	ALL		L	T
HLXB9	homeo box HB9	3110	7q36	AML		L	T
HMGA1	high mobility group AT-hook 1	3159	6p21	microfollicular thyroid adenoma, various benign mesenchymal tumors,		E, M	T
HMGA2	high mobility group AT-hook 2 (HMGIC)	8091	12q15	lipoma, leiomyoma, pleiomorphic salivary gland adenoma		M	T
HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1	3181	7p15	prostate		E	T
HOOK3	hook homolog 3	84376	8p11.21	papillary thyroid		E	T
HOXA11	homeo box A11	3207	7p15-p14.2	CML		L	T
HOXA13	homeo box A13	3209	7p15-p14.2	AML		L	T
HOXA9	homeo box A9	3205	7p15-	AML*		L	T

			p14.2				
HOXC11	homeo box C11	3227	12q13.3	AML		L	T
HOXC13	homeo box C13	3229	12q13.3	AML		L	T
HOXD11	homeo box D11	3237	2q31-q32	AML		L	T
HOXD13	homeo box D13	3239	2q31-q32	AML*		L	T
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	3265	11p15.5	infrequent sarcomas, rare other types	Costello syndrome	E, L, M	Mis
HRPT2	hyperparathyroidism 2	3279	1q21-q31	parathyroid adenoma	Hyperparathyroidism-jaw tumor syndrome	E, M	Mis, N, F
HSPCA	heat shock 90kDa protein 1, alpha	3320	14q32.31	NHL		L	T
HSPCB	heat shock 90kDa protein 1, beta	3326	6p12	NHL		L	T
IDH1	isocitrate dehydrogenase 1 (NADP+), soluble	3417	2q33.3	gliobastoma		O	Mis
IDH2	socitrate dehydrogenase 2 (NADP+), mitochondrial	3418	15q26.1	GBM		M	M
IGH@	immunoglobulin heavy locus	3492	14q32.33	MM, Burkitt lymphoma, NHL, CLL, B-ALL, MALT, MLCLS		L	T
IGK@	immunoglobulin kappa locus	50802	2p12	Burkitt lymphoma, B-NHL		L	T
IGL@	immunoglobulin lambda locus	3535	22q11.1-q11.2	Burkitt lymphoma		L	T
IKZF1	IKAROS family zinc finger 1	10320	7p12.2	ALL, DLBCL		L	D,T
IL2	interleukin 2	3558	4q26-q27	intestinal T-cell lymphoma		L	T
IL21R	interleukin 21 receptor	50615	16p11	NHL		L	T
IL6ST	interleukin 6 signal transducer (gp130, oncostatin M receptor)	3572	5q11	hepatocellular ca		E	O
IL7R	interleukin 7 receptor	146661	5p13	ALL, ETP ALL		L	Mis, O
IRF4	interferon regulatory factor 4	3662	6p25-p23	MM		L	T
IRTA1	immunoglobulin superfamily receptor	83417	1q21	B-NHL		L	T

	translocation associated 1						
ITK	IL2-inducible T-cell kinase	3702	5q31-q32	peripheral T-cell lymphoma		L	T
JAK1	Janus kinase 1	3716	1p32.3-p31.3	ALL		L	Mis
JAK2	Janus kinase 2	3717	9p24	ALL, AML, MPD, CML		L	T, Mis, O
JAK3	Janus kinase 3	3718	19p13.1	acute megakaryocytic leukemia, ETP ALL		L	Mis
JAZF1	juxtaposed with another zinc finger gene 1	221895	7p15.2-p15.1	endometrial stromal tumours		M	T
JUN	jun oncogene	3725	1p32-p31	sarcoma		M	A
KDM5A	lysine (K)-specific demethylase 5A, JARID1A	5927	12p11	AML		L	T
KDM5C	lysine (K)-specific demethylase 5C ***(JARID1C***)	8242	Xp11.22-p11.21	clear cell renal carcinoma		E	N, F, S
KDM6A	lysine (K)-specific demethylase 6A, ***UTX***	7403	Xp11.2	renal, oesophageal SCC, MM		E, L	D, N, F, S
KDR	vascular endothelial growth factor receptor 2	3791	4q11-q12	NSCLC, angiosarcoma		E	Mis
KIAA1549	KIAA1549	57670	7q34	pilocytic astrocytoma		O	O
KIF5B	kinesin family member 5B	3799	10p11.22	NSCLC		E	T
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	3815	4q12	GIST, AML, TGCT, mastocytosis, mucosal melanoma	Familial gastrointestinal stromal tumour	L, M, O, E	Mis, O
KLK2	kallikrein-related peptidase 2	3817	19q13.41	prostate		E	T
KRAS	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog	3845	12p12.1	pancreatic, colorectal, lung, thyroid, AML, others		L, E, M, O	Mis
KTN1	kinectin 1 (kinesin receptor)	3895	14q22.1	papillary thryoid		E	T
LAF4	lymphoid nuclear protein related to AF4	3899	2q11.2-q12	ALL, T-ALL		L	T
LASP1	LIM and SH3 protein 1	3927	17q11-q21.3	AML		L	T
LCK	lymphocyte-specific protein tyrosine kinase	3932	1p35-p34.3	T-ALL		L	T
LCP1	lymphocyte cytosolic protein 1 (L-	3936	13q14.1-	NHL		L	T

	plastin)		q14.3				
LCX	leukemia-associated protein with a CXXC domain	80312	10q21	AML		L	T
LHFP	lipoma HMGIC fusion partner	10186	13q12	lipoma		M	T
LIFR	leukemia inhibitory factor receptor	3977	5p13-p12	salivary adenoma		E	T
LMO1	LIM domain only 1 (rhombotin 1) (RBTN1)	4004	11p15	T-ALL, neuroblastoma		L	T, A
LMO2	LIM domain only 2 (rhombotin-like 1) (RBTN2)	4005	11p13	T-ALL		L	T
LPP	LIM domain containing preferred translocation partner in lipoma	4026	3q28	lipoma, leukemia		L, M	T
LRIG3	leucine-rich repeats and immunoglobulin-like domains 3	121227	12q14.1	NSCLC		E	T
LYL1	lymphoblastic leukemia derived sequence 1	4066	19p13.2-p13.1	T-ALL		L	T
MADH4	Homolog of Drosophila Mothers Against Decapentaplegic 4 gene	4089	18q21.1	colorectal, pancreatic, small intestine	Juvenile polyposis	E	D, Mis, N, F
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog	4094	16q22-q23	MM		L	T
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	9935	20q11.2-q13.1	MM		L	T
MALT1	mucosa associated lymphoid tissue lymphoma translocation gene 1	10892	18q21	MALT		L	T
MAML2	mastermind-like 2 (Drosophila)	84441	11q22-q23	salivary gland mucoepidermoid		E	T
MAP2K4	mitogen-activated protein kinase	6416	17p11.2	pancreatic, breast, colorectal		E	D, Mis, N

	kinase 4						
MDM2	Mdm2 p53 binding protein homolog	4193	12q15	sarcoma, glioma, colorectal, other		M, O, E, L	A
MDM4	Mdm4 p53 binding protein homolog	4194	1q32	GBM, bladder, retinoblastoma		M	A
MDS1	myelodysplasia syndrome 1	4197	3q26	MDS, AML		L	T
MDS2	myelodysplastic syndrome 2	259283	1p36	MDS		L	T
MECT1	mucoepidermoid translocated 1	94159	19p13	salivary gland mucoepidermoid		E	T
MED12	mediator complex subunit 12	9968	Xq13	uterine leiomyoma		M	M, S
MEN1	multiple endocrine neoplasia type 1 gene	4221	11q13	parathyroid tumors, Pancreatic neuroendocrine tumors	Multiple Endocrine Neoplasia Type 1	E	D, Mis, N, F, S
MET	met proto-oncogene (hepatocyte growth factor receptor)	4233	7q31	papillary renal, head-neck squamous cell	Familial Papillary Renal Cancer	E	Mis
MITF	microphthalmia-associated transcription factor	4286	3p14.1	melanoma		E	A
MKL1	megakaryoblastic leukemia (translocation) 1	57591	22q13	acute megakaryocytic leukemia		L	T
MLF1	myeloid leukemia factor 1	4291	3q25.1	AML		L	T
MLH1	E.coli MutL homolog gene	4292	3p21.3	colorectal, endometrial, ovarian, CNS	Hereditary non-polyposis colorectal cancer, Turcot syndrome	E, O	D, Mis, N, F, S
MLL	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila)	4297	11q23	AML, ALL		L	T, O
MLL2	myeloid/lymphoid or mixed-lineage leukemia 2	8085	12q12-q14	medulloblastoma, renal		O, E	N, F, Mis
MLL3	myeloid/lymphoid or mixed-lineage leukemia 3	58508	7q36.1	medulloblastoma		O	N

MLLT1	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 1 (ENL)	4298	19p13.3	AL		L	T
MLLT10	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 10 (AF10)	8028	10p12	AL		L	T
MLLT2	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 2 (AF4)	4299	4q21	AL		L	T
MLLT3	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 3 (AF9)	4300	9p22	ALL		L	T
MLLT4	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 4 (AF6)	4301	6q27	AL		L	T
MLLT6	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 6 (AF17)	4302	17q21	AL		L	T
MLLT7	myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, Drosophila); translocated to, 7 (AFX1)	4303	Xq13.1	AL		L	T
MN1	meningioma (disrupted in	4330	22q13	AML, meningioma		L, O	T

	balanced translocation) 1						
MPL	myeloproliferative leukemia virus oncogene, thrombopoietin receptor	4352	p34	MPD	Familial essential thrombocythemia	L	Mis
MSF	MLL septin-like fusion	10801	17q25	AML*		L	T
MSH2	mutS homolog 2 (E. coli)	4436	2p22-p21	colorectal, endometrial, ovarian	Hereditary non-polyposis colorectal cancer	E	D, Mis, N, F, S
MSH6	mutS homolog 6 (E. coli)	2956	2p16	colorectal	Hereditary non-polyposis colorectal cancer	E	Mis, N, F, S
MSI2	musashi homolog 2 (Drosophila)	124540	17q23.2	CML		L	T
MSN	moesin	4478	Xq11.2-q12	ALCL		L	T
MTCP1	mature T-cell proliferation 1	4515	Xq28	T cell prolymphocytic leukemia		L	T
MUC1	mucin 1, transmembrane	4582	1q21	B-NHL		L	T
MUTYH	mutY homolog (E. coli)	4595	1p34.3-1p32.1		Adenomatous polyposis coli	E	Mis
MYB	v-myb myeloblastosis viral oncogene homolog	4602	6q22-23	adenoid cystic carcinoma		E	T
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	4609	8q24.12-q24.13	Burkitt lymphoma, amplified in other cancers, B-CLL		L, E	A, T
MYCL1	v-myc myelocytomatosis viral oncogene homolog 1, lung carcinoma derived (avian)	4610	1p34.3	small cell lung		E	A
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	4613	2p24.1	neuroblastoma		O	A
MYD88	myeloid differentiation primary response gene (88)	4615	3p22	ABC-DLBCL		L	Mis

MYH11	myosin, heavy polypeptide 11, smooth muscle	4629	16p13.13-p13.12	AML		L	T
MYH9	myosin, heavy polypeptide 9, non-muscle	4627	22q13.1	ALCL		L	T
MYST4	MYST histone acetyltransferase (monocytic leukemia) 4 (MORF)	23522	10q22	AML		L	T
NACA	nascent-polypeptide-associated complex alpha polypeptide	4666	12q23-q24.1	NHL		L	T
NBS1	Nijmegen breakage syndrome 1 (nibrin)	4683	8q21		Nijmegen breakage syndrome	L, E, M, O	Mis, N, F
NCOA1	nuclear receptor coactivator 1	8648	2p23	alveolar rhadomyosarcoma		M	T
NCOA2	nuclear receptor coactivator 2 (TIF2)	10499	8q13.1	AML, Chondrosarcoma		L	T
NCOA4	nuclear receptor coactivator 4 - PTC3 (ELE1)	8031	10q11.2	papillary thyroid		E	T
NDRG1	N-myc downstream regulated 1	10397	8q24.3	prostate		E	T
NF1	neurofibromatosis type 1 gene	4763	17q12	neurofibroma, glioma	Neurofibromatosis type 1	O	D, Mis, N, F, S, O
NF2	neurofibromatosis type 2 gene	4771	22q12.2	meningioma, acoustic neuroma, renal	Neurofibromatosis type 2	O	D, Mis, N, F, S, O
NFE2L2	nuclear factor (erythroid-derived 2)-like 2 (NRF2)	4780	2q31	NSCLC, HNSCC		E	Mis
NFIB	nuclear factor I/B	4781	9p24.1	adenoid cystic carcinoma, lipoma		E	T
NFKB2	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)	4791	10q24	B-NHL		L	T
NIN	ninein (GSK3B interacting protein)	51199	14q24	MPD		L	T
NKX2-1	NK2 homeobox 1	7080	14q13	NSCLC		E	A
NONO	non-POU domain containing, octamer-binding	4841	Xq13.1	papillary renal cancer		E	T

NOTCH1	Notch homolog 1, translocation-associated (Drosophila) (TAN1)	4851	9q34.3	T-ALL		L	T, Mis, O
NOTCH2	Notch homolog 2	4853	1p13-p11	marginal zone lymphoma, DLBCL		L	N, F, Mis
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	4869	5q35	NHL, APL, AML		L	T, F
NR4A3	nuclear receptor subfamily 4, group A, member 3 (NOR1)	8013	9q22	extraskeletal myxoid chondrosarcoma		M	T
NRAS	neuroblastoma RAS viral (v-ras) oncogene homolog	4893	1p13.2	melanoma, MM, AML, thyroid		L, E	Mis
NSD1	nuclear receptor binding SET domain protein 1	64324	5q35	AML		L	T
NTRK1	neurotrophic tyrosine kinase, receptor, type 1	4914	1q21-q22	papillary thyroid		E	T
NTRK3	neurotrophic tyrosine kinase, receptor, type 3	4916	15q25	congenital fibrosarcoma, Secretory breast		E, M	T
NUMA1	nuclear mitotic apparatus protein 1	4926	11q13	APL		L	T
NUP214	nucleoporin 214kDa (CAN)	8021	9q34.1	AML, T-ALL		L	T
NUP98	nucleoporin 98kDa	4928	11p15	AML		L	T
OLIG2	oligodendrocyte lineage transcription factor 2 (BHLHB1)	10215	21q22.11	T-ALL		L	T
OMD	osteomodulin	4958	9q22.31	aneurysmal bone cysts		M	T
P2RY8	purinergic receptor P2Y, G-protein coupled, 8	286530	Xp22.3; Yp11.3	B-ALL, Downs associated ALL		L	T
PAFAH1B2	platelet-activating factor acetylhydrolase, isoform Ib, beta subunit 30kDa	5049	11q23	MLCLS		L	T
PALB2	partner and localizer of BRCA2	79728	16p12.1		Fanconi anaemia N, breast cancer susceptibili	L, O, E	F, N, Mis

					ty		
PAX3	paired box gene 3	5077	2q35	alveolar rhabdomyosarcoma		M	T
PAX5	paired box gene 5 (B-cell lineage specific activator protein)	5079	9p13	NHL, ALL, B-ALL		L	T, Mis, D, F, S
PAX7	paired box gene 7	5081	1p36.2-p36.12	alveolar rhabdomyosarcoma		M	T
PAX8	paired box gene 8	7849	2q12-q14	follicular thyroid		E	T
PBRM1	polybromo 1	55193	3p21	clear cell renal carcinoma, breast		E	Mis, N, F, S, D, O
PBX1	pre-B-cell leukemia transcription factor 1	5087	1q23	pre B-ALL, myoepithelioma		L, M	T
PCM1	pericentriolar material 1 (PTC4)	5108	8p22-p21.3	papillary thyroid, CML, MPD		E, L	T
PCSK7	proprotein convertase subtilisin/kexin type 7	9159	11q23.3	MLCLS		L	T
PDE4DIP	phosphodiesterase 4D interacting protein (myomegalin)	9659	1q12	MPD		L	T
PDGFB	platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)	5155	22q12.3-q13.1	DFSP		M	T
PDGFRA	platelet-derived growth factor, alpha-receptor	5156	4q11-q13	GIST, idiopathic hypereosinophilic syndrome, paediatric GBM		L, M, O	Mis, O, T
PDGFRB	platelet-derived growth factor receptor, beta polypeptide	5159	5q31-q32	MPD, AML, CMML, CML		L	T
PER1	period homolog 1 (Drosophila)	5187	17p13.1-17p12	AML, CMML		L	T
PHF6	PHD finger protein 6	84295	Xq26.3	ETP ALL		L	F, N, Splice, Mis

PHOX2B	paired-like homeobox 2b	8929	4p12	neuroblastoma	familial neuroblastoma	O	Mis, F
PICALM	phosphatidylinositol binding clathrin assembly protein (CALM)	8301	11q14	TALL, AML,		L	T
PIK3CA	phosphoinositide-3-kinase, catalytic, alpha polypeptide	5290	3q26.3	colorectal, gastric, gliobastoma, breast		E, O	Mis
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	5295	5q13.1	gliobastoma, ovarian, colorectal		E, O	Mis, F, O
PIM1	pim-1 oncogene	5292	6p21.2	NHL		L	T
PLAG1	pleiomorphic adenoma gene 1	5324	8q12	salivary adenoma		E	T
PML	promyelocytic leukemia	5371	15q22	APL, ALL		L	T
PMS1	PMS1 postmeiotic segregation increased 1 (S. cerevisiae)	5378	2q31-q33		Hereditary non-polyposis colorectal cancer	E	Mis, N
PMS2	PMS2 postmeiotic segregation increased 2 (S. cerevisiae)	5395	7p22		Hereditary non-polyposis colorectal cancer, Turcot syndrome	E	Mis, N, F
PMX1	paired mesoderm homeo box 1	5396	1q24	AML		L	T
PNUTL1	peanut-like 1 (Drosophila)	5413	22q11.2	AML		L	T
POU2AF1	POU domain, class 2, associating factor 1 (OBF1)	5450	11q23.1	NHL		L	T
POU5F1	POU domain, class 5, transcription factor 1	5460	6p21.31	sarcoma		M	T
PPARG	peroxisome proliferative activated receptor, gamma	5468	3p25	follicular thyroid		E	T
PPP2R1A	protein phosphatase 2, regulatory subunit A, alpha	5518	19q13.41	clear cell ovarian carcinoma		E	Mis
PRCC	papillary renal cell carcinoma (translocation-	5546	1q21.1	papillary renal		E	T

	associated)						
PRDM1	PR domain containing 1, with ZNF domain	639	6q21	DLBCL		L	D, N, Mis, F, S
PRDM16	PR domain containing 16	63976	1p36.23-p33	MDS, AML		L	T
PRF1	perforin 1 (pore forming protein)	5551	10q22			L	M
PRKAR1A	protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue specific extinguisher 1)	5573	17q23-q24	papillary thyroid	Carney complex	E, M	T, Mis, N, F, S
PRO1073	PRO1073 protein (ALPHA)	29005	11q31.1	renal cell carcinoma (childhood epithelioid)		E	T
PSIP2	PC4 and SFRS1 interacting protein 2 (LEDGF)	11168	9p22.2	AML		L	T
PTCH	Homolog of Drosophila Patched gene	5727	9q22.3	skin basal cell, medulloblastoma	Nevoid Basal Cell Carcinoma Syndrome	E, M	Mis, N, F, S
PTEN	phosphatase and tensin homolog gene	5728	10q23.3	glioma, prostate, endometrial	Cowden Syndrome, Bannayan-Riley-Ruvalcaba syndrome	L, E, M, O	D, Mis, N, F, S
PTPN11	protein tyrosine phosphatase, non-receptor type 11	5781	12q24.1	JMML, AML, MDS		L	Mis
RAB5EP	rabaptin, RAB GTPase binding effector protein 1 (RABPT5)	9135	17p13	CMML		L	T
RAD51L1	RAD51-like 1 (S. cerevisiae) (RAD51B)	5890	14q23-q24.2	lipoma, uterine leiomyoma		M	T
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	5894	3p25	pilocytic astrocytoma		M	T
RALGDS	ral guanine nucleotide dissociation stimulator	5900	9q34.3	PMBL, Hodgkin Lymphona,		L	T
RANBP17	RAN binding protein 17	64901	5q34	ALL		L	T
RAP1GDS1	RAP1, GTP-GDP dissociation stimulator 1	5910	4q21-q25	T-ALL		L	T

RARA	retinoic acid receptor, alpha	5914	17q12	APL		L	T
RB1	retinoblastoma gene	5925	13q14	retinoblastoma, sarcoma, breast, small cell lung	Familial retinoblastoma	L, E, M, O	D, Mis, N, F, S
RBM15	RNA binding motif protein 15	64783	1p13	acute megakaryocytic leukemia		L	T
RECQL4	RecQ protein-like 4	9401	8q24.3		Rothmund-Thompson Syndrome	M	N, F, S
REL	v-rel reticuloendotheliosis viral oncogene homolog (avian)	5966	2p13-p12	Hodgkin Lymphoma		L	A
RET	ret proto-oncogene	5979	10q11.2	medullary thyroid, papillary thyroid, pheochromocytoma, NSCLC	Multiple endocrine neoplasia 2A/2B	E, O	T, Mis, N, F
ROS1	v-ros UR2 sarcoma virus oncogene homolog 1 (avian)	6098	6q22	glioblastoma, NSCLC		O, E	T
RPL22	ribosomal protein L22 (EAP)	6146	1p36.31	AML, CML		L	T
RPN1	ribophorin I	6184	3q21.3-q25.2	AML		L	T
RUNDC2A	RUN domain containing 2A	84127	16p13.13	PMBL, Hodgkin Lymphona,		L	T
RUNX1	runt-related transcription factor 1 (AML1)	861	21q22.3	AML, preB-ALL, T-ALL		L	T
RUNXBP2	runt-related transcription factor binding protein 2 (MOZ/ZNF220)	7994	8p11	AML		L	T
SBDS	Shwachman-Bodian-Diamond syndrome protein	51119	7q11		Schwachman-Diamond syndrome	L	Gene Conversion
SDC4	syndecan 4	6385	20q12	NSCLC		E	T
SDH5	chromosome 11 open reading frame 79	54949	11q12.2		Familial paraganglioma	M	M
SDHB	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	6390	1p36.1-p35		Familial paraganglioma	O	Mis, N, F

SDHC	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa	6391	1q21		Familial paraganglioma	O	Mis, N, F
SDHD	succinate dehydrogenase complex, subunit D, integral membrane protein	6392	11q23		Familial paraganglioma	O	Mis, N, F, S
SEPT6	septin 6	23157	Xq24	AML		L	T
SET	SET translocation	6418	9q34	AML		L	T
SETD2	SET domain containing 2	29072	3p21.31	clear cell renal carcinoma		E	N, F, S, Mis
SF3B1	splicing factor 3b, subunit 1, 155kDa	23451	2q33.1	myelodysplastic syndrome		L	Mis
SFPQ	splicing factor proline/glutamine rich(polypyrimidine tract binding protein associated)	6421	1p34.3	papillary renal cell		E	T
SFRS3	splicing factor, arginine/serine-rich 3	6428	6p21	follicular lymphoma		L	T
SH3GL1	SH3-domain GRB2-like 1 (EEN)	6455	19p13.3	AL		L	T
SIL	TAL1 (SCL) interrupting locus	6491	1p32	T-ALL		L	T
SLC34A2	solute carrier family 34 (sodium phosphate), member 2	10568	4p15.2	NSCLC		E	T
SLC45A3	solute carrier family 45, member 3	85414	1q32	prostate		E	T
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	6597	19p13.2	NSCLC		E	F, N, Mis
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	6598	22q11	malignant rhabdoid	Rhabdoid predisposition syndrome	M	D, N, F, S

SMO	smoothened homolog (Drosophila)	6608	7q31-q32	skin basal cell		E	Mis
SOCS1	suppressor of cytokine signaling 1	8651	16p13.13	Hodgkin Lymphoma, PMBL		L	F, O
SOX2	SRY (sex determining region Y)-box 2	6657	3q26.3-q27	NSCLC, oesophageal squamous carcinoma		E	A
SRGAP3	SLIT-ROBO Rho GTPase activating protein 3	9901	3p25.3	pilocytic astrocytoma		M	T
SRSF2	serine/arginine-rich splicing factor 2	6427	17q25	MDS, CLL		L	Mis
SS18	synovial sarcoma translocation, chromosome 18	6760	18q11.2	synovial sarcoma		M	T
SS18L1	synovial sarcoma translocation gene on chromosome 18-like 1	26039	20q13.3	synovial sarcoma		M	T
SSH3BP1	spectrin SH3 domain binding protein 1	10006	10p11.2	AML		L	T
SSX1	synovial sarcoma, X breakpoint 1	6756	Xp11.23-p11.22	synovial sarcoma		M	T
SSX2	synovial sarcoma, X breakpoint 2	6757	Xp11.23-p11.22	synovial sarcoma		M	T
SSX4	synovial sarcoma, X breakpoint 4	6759	Xp11.23	synovial sarcoma		M	T
STK11	serine/threonine kinase 11 gene (LKB1)	6794	19p13.3	NSCLC, pancreatic	Peutz-Jeghers syndrome	E, M, O	D, Mis, N, F, S
STL	Six-twelve leukemia gene	7955	6q23	B-ALL		L	T
SUFU	suppressor of fused homolog (Drosophila)	51684	10q24.32	medulloblastoma	Medulloblastoma predisposition	O	D, F, S
SUZ12	suppressor of zeste 12 homolog (Drosophila)	23512	17q11.2	endometrial stromal tumours		M	T
SYK	spleen tyrosine kinase	6850	9q22	MDS, peripheral T-cell lymphoma		L	T
TAF15	TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa	8148	17q11.1-q11.2	extraskeletal myxoid chondrosarcomas, ALL		L, M	T

TAL1	T-cell acute lymphocytic leukemia 1 (SCL)	6886	1p32	lymphoblastic leukemia/biphasic		L	T
TAL2	T-cell acute lymphocytic leukemia 2	6887	9q31	T-ALL		L	T
TCEA1	transcription elongation factor A (SII), 1	6917	8q11.2	salivary adenoma		E	T
TCF1	transcription factor 1, hepatic (HNF1)	6927	12q24.2	hepatic adenoma, hepatocellular ca	Familial Hepatic Adenoma	E	Mis, F
TCF12	transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)	6938	15q21	extraskeletal myxoid chondrosarcoma		M	T
TCF3	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)	6929	19p13.3	pre B-ALL		L	T
TCF7L2	transcription factor 7-like 2	6934	10q25.3	colorectal		E	T
TCL1A	T-cell leukemia/lymphoma 1A	8115	14q32.1	T-CLL		L	T
TCL6	T-cell leukemia/lymphoma 6	27004	14q32.1	T-ALL		L	T
TET2	tet oncogene family member 2	54790	4q24	MDS		L	Mis N, F
TFE3	transcription factor binding to IGHM enhancer 3	7030	Xp11.22	papillary renal, alveolar soft part sarcoma, renal		E	T
TFEB	transcription factor EB	7942	6p21	renal (childhood epithelioid)		E,M	T
TFG	TRK-fused gene	10342	3q11-q12	papillary thyroid, ALCL, NSCLC		E, L	T
TFPT	TCF3 (E2A) fusion partner (in childhood Leukemia)	29844	19q13	pre-B ALL		L	T
TFRC	transferrin receptor (p90, CD71)	7037	3q29	NHL		L	T
THRAP3	thyroid hormone receptor associated protein 3 (TRAP150)	9967	1p34.3	aneurysmal bone cysts		M	T
TIF1	transcriptional intermediary factor 1 (PTC6,TIF1A)	8805	7q32-q34	APL		L	T
TLX1	T-cell	3195	10q24	T-ALL		L	T

	leukemia, homeobox 1 (HOX11)						
TLX3	T-cell leukemia, homeobox 3 (HOX11L2)	30012	5q35.1	T-ALL		L	T
TMPRSS2	transmembrane protease, serine 2	7113	21q22.3	prostate		E	T
TNFAIP3	tumor necrosis factor, alpha-induced protein 3	7128	6q23	marginal zone B-cell lymphomas, Hodgkin's lymphoma, primary mediastinal B cell lymphoma		L	D, N, F
TNFRSF14	tumor necrosis factor receptor superfamily, member 14 (herpesvirus entry mediator)	8764	1p36.32	follicular lymphoma		L	Mis, N, F
TNFRSF17	tumor necrosis factor receptor superfamily, member 17	608	16p13.1	intestinal T-cell lymphoma		L	T
TNFRSF6	tumor necrosis factor receptor superfamily, member 6 (FAS)	355	10q24.1	TGCT, nasal NK/T lymphoma, skin squamous cell ca -burn scar-related		L, E, O	Mis
TOP1	topoisomerase (DNA) I	7150	20q12-q13.1	AML*		L	T
TP53	tumor protein p53	7157	17p13	breast, colorectal, lung, sarcoma, adrenocortical, glioma, multiple other tumour types	Li-Fraumeni syndrome	L, E, M, O	Mis, N, F
TPM3	tropomyosin 3	7170	1q22-q23	papillary thyroid, ALCL, NSCLC		E, L	T
TPM4	tropomyosin 4	7171	19p13.1	ALCL		L	T
TPR	translocated promoter region	7175	1q25	papillary thyroid		E	T
TRA@	T cell receptor alpha locus	6955	14q11.2	T-ALL		L	T
TRB@	T cell receptor beta locus	6957	7q35	T-ALL		L	T
TRD@	T cell receptor delta locus	6964	14q11	T-cell leukemia		L	T
TRIM27	tripartite motif-containing 27	5987	6p22	papillary thyroid		E	T
TRIM33	tripartite motif-containing 33 (PTC7,TIF1G)	51592	1p13	papillary thyroid		E	T
TRIP11	thyroid	9321	14q31-	AML		L	T

	hormone receptor interactor 11		q32				
TSC1	tuberous sclerosis 1 gene	7248	9q34		Tuberous sclerosis 1	E, O	D, Mis, N, F, S
TSC2	tuberous sclerosis 2 gene	7249	16p13.3		Tuberous sclerosis 2	E, O	D, Mis, N, F, S
TSHR	thyroid stimulating hormone receptor	7253	14q31	toxic thyroid adenoma		E	Mis
TTL	tubulin tyrosine ligase	150465	2q13	ALL		L	T
U2AF1	U2 small nuclear RNA auxiliary factor 1	7307	21q22.3	CLL, MDS		L	Mis
USP6	ubiquitin specific peptidase 6 (Tre-2 oncogene)	9098	17p13	aneurysmal bone cysts		M	T
VHL	von Hippel-Lindau syndrome gene	7428	3p25	renal, hemangioma, pheochromocytoma	von Hippel-Lindau syndrome	E, M, O	D, Mis, N, F, S
VTI1A	vesicle transport through interaction with t-SNAREs homolog 1A	143187	10q25.2	colorectal		E	T
WAS	Wiskott-Aldrich syndrome	7454	Xp11.23-p11.22		Wiskott-Aldrich syndrome	L	Mis, N, F, S
WHSC1	Wolf-Hirschhorn syndrome candidate 1(MMSET)	7468	4p16.3	MM		L	T
WHSC1L1	Wolf-Hirschhorn syndrome candidate 1-like 1 (NSD3)	54904	8p12	AML		L	T
WIF1	WNT inhibitory factor 1	11197	12q14.3	pleomorphic salivary gland adenoma		E	T
WRN	Werner syndrome (RECQL2)	7486	8p12-p11.2		Werner Syndrome	L, E, M, O	Mis, N, F, S
WT1	Wilms tumour 1 gene	7490	11p13	Wilms, desmoplastic small round cell tumor	Denys-Drash syndrome, Frasier syndrome, Familial Wilms tumor	O	D, Mis, N, F, S
WTX	family with sequence similarity 123B (FAM123B)	139285	Xq11.1	Wilms tumour		O	F, D, N, Mis

WWTR1	WW domain containing transcription regulator 1	607392	3q23-q24	epitheliod hemangioendothelioma		M	T
XPA	xeroderma pigmentosum, complementation group A	7507	9q22.3		Xeroderma pigmentosum (A)	E	Mis, N, F, S
XPC	xeroderma pigmentosum, complementation group C	7508	3p25		Xeroderma pigmentosum (C)	E	Mis, N, F, S
XPO1	exportin 1 (CRM1 homolog, yeast)	7514	2p15	CLL		L	Mis
YWHAE	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide (14-3-3 epsilon)	7531	17p13.3	edometrial stromal sarcoma		M	T
ZNF145	zinc finger protein 145 (PLZF)	7704	11q23.1	APL		L	T
ZNF198	zinc finger protein 198	7750	13q11-q12	MPD, NHL		L	T
ZNF278	zinc finger protein 278 (ZSG)	23598	22q12-q14	Ewing sarcoma		M	T
ZNF331	zinc finger protein 331	55422	19q13.3-q13.4	follicular thyroid adenoma		E	T
ZNF384	zinc finger protein 384 (CIZ/NMP4)	171017	12p13	ALL		L	T
ZNF521	zinc finger protein 521	25925	18q11.2	ALL		L	T
ZNF9	zinc finger protein 9 (a cellular retroviral nucleic acid binding protein)	7555	3q21	aneurysmal bone cysts		M	T
ZRSR2	zinc finger (CCCH type), RNA-binding motif and serine/arginine rich 2	8233	Xp22.1	MDS, CLL		L	F, S, Mis

[00239] The following abbreviations are used in the table above: A, amplification; AEL, acute eosinophilic leukemia; AL, acute leukemia; ALCL, anaplastic large-cell lymphoma; ALL, acute lymphocytic leukemia; AML, acute myelogenous leukemia; AML*, acute myelogenous leukemia (primarily treatment associated); APL, acute promyelocytic leukemia; B-ALL, B-cell acute lymphocytic leukaemia; B-CLL, B-cell Lymphocytic leukemia; B-NHL,

B-cell Non-Hodgkin Lymphoma; CLL, chronic lymphatic leukemia; CML, chronic myeloid leukemia; CMML, chronic myelomonocytic leukemia; CNS, central nervous system; D, large deletion; DFSP, dermatofibrosarcoma protuberans; DLBCL, diffuse large B-cell lymphoma; DLCL, diffuse large-cell lymphoma; Dom, dominant; E, epithelial; F frameshift; GIST, gastrointestinal stromal tumour; JMML, juvenile myelomonocytic leukemia; L, leukemia/lymphoma; M, mesenchymal; MALT, mucosa-associated lymphoid tissue lymphoma; MDS, myelodysplastic syndrome; Mis, Missense; MLCLS mediastinal large cell lymphoma with sclerosis; MM, multiple myeloma; MPD, Myeloproliferative disorder; N, nonsense; NHL, non-Hodgkin lymphoma; NK/T, natural killer T cell; NSCLC, non small cell lung cancer; O, other; PMBL, primary mediastinal B-cell lymphoma; pre-B All, pre-B-cell acute lymphoblastic leukaemia; Rec, recessive; S, splice site; T, translocation; T-ALL, T-cell acute lymphoblastic leukemia; T-CLL, T-cell chronic lymphocytic leukaemia; TGCT, testicular germ cell tumour; T-PLL, T cell prolymphocytic leukaemia. It is contemplated that a padlock probe can be designed in order to detect a mutation in a cancer gene listed in the table above. A padlock probe can have a sequence that is complementary to a mutation, which may be a substitution of one or more nucleotides for one or more wild-type nucleotides in a mRNA sequence, a deletion of one or more nucleotides compared to a wild-type mRNA sequence, an addition of one or more nucleotides compared to a wild-type mRNA sequence, a sequence inversion, a sequence translocation, a frameshift, or other mutation. With many cancer mutations, the mutation has been previously characterized, including with inversions and translocations such that sequence design is possible.

H. Kits

[00240] Also provided are kits for use in methods described herein. The kit may comprise at least one (species of) padlock probe, as defined above, specific for a particular cDNA. Such a kit may also comprise RT primer(s), an RT enzyme, a ribonuclease, a DNA polymerase, a ligase and/or means of detection of RCA product.

[00241] The kit may optionally further comprise one or more gap oligonucleotides with complementarity to the portion of the target cDNA which lies between non-adjacently-hybridized padlock probe ends or may comprise reagents for otherwise filling any gap present when the ends of the padlock probe are hybridized to the cDNA, such as a polymerase, nucleotides and necessary co-factors. In some embodiments, the kit may further

comprises a primer oligonucleotide for priming RCA of the padlock probe. In certain aspects, the primer hybridizes to the padlock probe at a location other than the region(s) of the padlock probe that is complementary to the target cDNA.

[00242] Alternatively or additionally, the kit may comprise a ligase for circularizing the padlock probe(s) (which may or may not be present in the kit) or a polymerase such as phi29 polymerase (and optionally necessary cofactors, as well and nucleotides) for effecting RCA. Reagents for detecting the RCA product may also be included in the kit. Such reagents may include a labeled oligonucleotide hybridization probe having complementarity to a portion of a padlock probe, or to a portion of a gap oligonucleotide, present in the kit.

[00243] The kit may be designed for use in multiplex embodiments of the different method embodiments, and accordingly may comprise combinations of the components defined above for more than target RNA. If probes having binding specificity respectively for a plurality of cDNA species are present in the kit, the kit may additionally comprise components allowing multiple RNA detection in parallel to be distinguished. For example, the kit may contain padlock probes for different cDNA targets, wherein the cDNA targets have "unique" sequences for hybridization only to a particular species of probe. Such padlock probes may for example carry different tag or identifier sequences allowing the detection of different RNAs to be distinguished.

[00244] The kit may be designed for use in the detection of an mRNA coding for KRAS. In some embodiments, the kit may contain one or more padlock probes that target cDNA reverse transcribed from the wild-type KRAS mRNA and/or one or more padlock probes which target cDNA reverse transcribed from a KRAS mRNA molecule comprising a point-mutation.

[00245] In addition to the above components, the kit may further include instructions for practicing method embodiments. These instructions may be present in the kit in a variety of forms, one or more of which may be present in the kit. One form in which these instructions may be present is as printed information on a suitable medium or substrate, *e.g.*, a piece or pieces of paper on which the information is printed, in the packaging of the kit, in a package insert, *etc.* Yet another means would be a computer readable medium, *e.g.*, diskette, CD, *etc.*, on which the information has been recorded. Yet another means that may be present is a

website address which may be used via the internet to access the information at a remote site. Any convenient means may be present in the kit.

[00246] Thus, in a further aspect a kit is provided for use in the localized *in situ* detection of a target RNA in a sample, the kit comprising one or more components selected from the list comprising:

(i) a padlock probe comprising 3′ and 5′ terminal regions having complementarity to cDNA transcribed from the target RNA (such regions can alternatively be defined as corresponding in sequence to regions of the target RNA, which regions as defined above may be adjacent or non-adjacent);

(ii) a reverse transcriptase primer capable of hybridizing to the target RNA (*e.g.* capable of hybridizing specifically to the RNA);

(iii) a reverse transcriptase;

(iv) a ribonuclease;

(v) a ligase;

(vi) a polymerase having 3′ exonuclease activity;

(vii) a gap oligonucleotide capable of hybridizing to a portion of a cDNA transcribed from the target RNA;

(viii) a detection probe capable of hybridizing to a complement of a padlock probe of (i); or to a complement of a gap oligonucleotide of (vii);

(ix) nucleotides for incorporation *e.g.* dNTPs.

[00247] The detection probe of (viii) may be a labeled detection oligonucleotide capable of hybridizing to the amplification product (which will contain a complement of a padlock probe of (i) or a complement of a gap oligonucleotide of (vii)). For example the detection oligonucleotide may be fluorescently labeled or may be labeled with a horse radish-peroxidase.

[00248] In one embodiment the kit may contain the padlock probe of (i) and optionally one or more further components selected from any one of (ii) to (ix). Other combinations of kit components are also possible. For example the kit may contain the padlock probe of (i) and at least one of the reverse transcriptase primer of (ii), the reverse transcriptase of (iii) and the ribonuclease of (iv), optionally with one or more further components selected from any one of (ii) or (iii) or (iv) to (ix). Other representative kits may include the reverse transcriptase primer of (ii), and at least one of components (iii) to (ix), more particularly the primer of (ii) with at least one of components ((iii) to (vi), and optionally with one or more further

components selected from any one of (i) or (vii) to (ix). Also included by way of representative example is a kit comprising at least two, or at least three, or all four, of components (iii) to (vi), optionally together with one or more further components selected from (i), (ii), or (vii) to (ix). All possible combinations of 2 or 3 components selected from (iii) to (iv) are covered. For example, such an embodiment may include (iii), (iv) and (v), or (iii), (v) and (vi), or (iii), (iv) and (vi) and so on.

[00249] In additional embodiments, kits may contain one or more reagents for sequencing after rolling circle amplification.

I. Tables

Table 1: Oligonucleotide sequences

	Name	Sequence	SEQ NOS:
cDNA primers	P-βe1[a]	A+TC+AT+CC+AT+GG+TG+AGCTGGCGGCGG	32
	P-βhum[a]	C+TG+AC+CC+AT+GC+CC+ACCATCACGCCC	33
	P-βmus	C+TG+AC+CC+AT+TC+CC+ACCATCACACCC	34
	P-βe6	T+TA+GA+GA+GA+AG+TG+GGGTGGCTTTTA	35
	P-cMyc[a]	G+CG+TC+CT+TG+CT+CG+GGTGTTGTAAGTTCCAG	36
	P-HER2[a]	G+AG+CT+GG+GT+GC+CT+CGCACAATCCGCAGCCT	37
	P-TERT[a]	A+GG+AC+AC+CT+GG+CG+GAAGGAGGGGGCGGCGG	38
	P-α1βmus[a]	A+CT+CG+TC+AT+AC+TC+CTGCTTGCTGATCCACA	39
	P-γ1mus[a]	G+CC+TC+AG+GA+AA+TC+CTGGAAGTCTGC	40
Padlock probes (detection probe)	PLP-βe1[b] (DP-1)	AGCCTCGCCTTTGCCTTCCTTTTACGACCTCAATGCTG CTGCTGTACTA CTCTTCGCCCCGCGAGCACAG	41
	PLP-βhum[a] (DP-4)	GCCGGCTTCGCGGGCGACGATTCCTCTATGATTACTG ACCTATGCGTCTATTTAGTGGAGCCTCTTCTTTACGGC GCCGGCATGTGCAAG	42
	PLP-βmus[a] (DP-5)	GCCGGCTTCGCGGGCGACGATTCCTCTATGATTACTG ACCTAAGTCGGAAGTACTACTCTCTTCTTCTTTACGGC GCCGGCATGTGCAAA	43
	PLP-βe6[a] (DP-1)	TACAGGAAGTCCCTTGCCATTTCCTCTATGATTACTGA CCTACCTCAATGCTGCTGCTGTACTACTCTTCCCAAAG ATGAGATGCGTTGT	44
	PLP2-βmus[c] (DP-3)	CTGTCCACCTTCCAGAGAGTGTACCGACCTCAGTAAG TAGCCGTGACTATCGACTTCCAGCCTGGCCTCA	45
	PLP-α1mus[c] (DP-2)	CTGTCCACCTTCCAGCCTTTCCTACGACCTCAATGCAC ATGTTTGGCTCCTCTTCTCCAGCCTGGCCTCG	46
	PLP-γ1mus[a] (DP-1)	CCCCAGCCTGGTGGAAGCTAGCTACCTCAATGCTGCT GCTGTACTACTATGACTGCTGGAGATGAGAAAG	47
	PLP-cMyc[c] (DP-4)	CGAAACTTTGCCCATAGCAGATTGGAACGTTTAAATG CGTCTATTTAGTGGAGCCGAGACAATCTTACATCGCA ACCCTTGCCGCATCCA	48
	PLP-HER2[c] (DP-5)	TGCCAGCCTGTCCTTCCTGCATCGTCTTAATCACTAGT CGGAAGTACTACTCTCTTACGCTTACAACTAGCTCAC CTACCTGCCCACCAA	49

	Name	Sequence	SEQ NOS:
	PLP-TERT[c] (DP-2)	GGTGTGCGTGCCCTGGGACGACTTTCTATGATTACTG ACCTACCTCAATGCACATGTTTGGCTCCTCTTCGCGCT GGTGGCCCAGTGCCT	50
	PLP-*KRAS*-wtGGT[3] (DP-3)	GGCGTAGGCAAGAGTTCCTGTAGTAAAGTAGCCGTG ACTATCGACTGAATCTAAGGTAGTTGGAGCTGGT	51
	PLP-*KRAS*-mutGAT[3] (DP-5)	GGCGTAGGCAAGAGTGTAAGTCATCAAGTCGGAAGT ACTACTCTCTGAATCTAAGGTAGTTGGAGCTGTT	52
Detection probes	DP-1Cy3[d]	Cy3-CCTCAATGCTGCTGCTGTACTAC	53
	DP-1Cy3.5[a]	TexasRed-CCTCAATGCTGCTGCTGTACTAC	53
	DP-2FITC[a]	FITC- CCTCAATGCACATGTTTGGCTCC	54
	DP-2Cy5[9]	Cy5-CCTCAATGCACATGTTTGGCTCC	54
	DP-3[d]	Cy3- AGTAGCCGTGACTATCGACT	55
	DP-4[c]	Cy3- TGCGTCTATTTAGTGGAGCC	56
	DP-5[d]	Cy5-AGTCGGAAGTACTACTCTCT	57
qPCR primers	ACTBfwd[b]	CTGGAACGGTGAAGGTGACA	58
	ACTBrev[b]	CGGCCACATTGTGAACTTTG	59
Oligonucleotides are given in 5′-3′ order + symbol denotes the LNA bases Oligonucleotides were purchased from Integrated DNA Technologies[3], DNA technology A/S[b], Biomers[c] and Eurogentec[d]			

Table 2: Sequences of cDNA primers for LNA content investigation

Primer	LNA content	Sequence	SEQ NOS:
P-unmod	No LNA	ATCATCCATGGTGAGCTGGCGGCGG	32
P-LNA1 or P-βe1	7 LNA, every 2nd	A+TC+AT+CC+AT+GG+TG+AGCTGGCGGCGG	32
P-LNA2	7 LNA, every 2nd	A+TC+AT+CC+AT+GG+TG+AGCTGGCGGCGGGTGTG	60
P-LNA3	9 LNA, every 2nd	A+TC+AT+CC+AT+GG+TG+AG+CT+GGCGGCGGGTGTG	60
P-LNA4	5 LNA, every 2nd	A+TC+AT+CC+AT+GGTGAGCTGGCGGCGGGTGTG	60
P-LNA5	5 LNA, every 3rd	AT+CAT+CCA+TGG+TGA+GCTGGCGGCGGGTGTG	60

Oligonucleotides are given in 5′-3′ order. + symbol denotes the LNA bases.

All LNA containin Oligonucleotides were purchased from Integrated DNA Technology. The unmodified primer was purchased from Biomers.

Table 3: Sequences of cDNA primers for investigation of cDNA synthesis length

Primer	Sequence	SEQ NOS:
P-93 nt or P-βe1	A+TC+AT+CC+AT+GG+TG+AGCTGGCGGCGG	32
P-141 nt	G+GC+CT+TG+CA+CA+TG+CCGGAGCCGTTGTCGAC	61
P-231 nt or P-βhum	C+TG+AC+CC+AT+GC+CC+ACCATCACGCCC	33
P-261 nt	C+TG+GG+CC+TC+GT+CG+CCCACATAGGAATCCTT	62
P-501 nt	C+AC+AG+CC+TG+GA+TA+GCAACGTACATGGCTGG	63

Oligonucleotides are given in 5′-3′ order. + symbol denotes the LNA base

Oligonucleotides were purchased from Integrated DNA Technology.

The primer name indicates the maximum length o the produced cDNA for each respective cDNA primer.

Table 4: Oligonucleotide sequences for genotyping of *KRAS* mutations

Oligonucleotide sequences			SEQ ID NOs
Primers		**Sequences (5´ - 3´)**	
P-*KRAS*-c12/13[b]		T+GT+AT+CG+TC+AA+GG+CACTCTT	64
P-*KRAS*-c12/13-II[a]		C+CT+CT+AT+TG+TT+GG+ATCATATTCGTC	65
P-*KRAS*-Q61H[b]		T+AT+TC+GT+CC+AC+AA+AATGATTCTGAA	66
P-EGFR-L858R[b]		T+CT+TT+CT+CT+TC+CG+CACCCAG	67
P-EGFR-S768I[b]		G+GC+GG+CA+CA+CGTGGGGGTTG	68
P-EGFR-G719C/A[b]		C+CT+TA+TA+CA+CC+GT+GCCGAAC	69
P-TP53-S127F[b]		A+GT+TG+GC+AA+AA+CA+TCTTGTTGAGGG	70
P-TP53-P190S[b]		T+TC+CT+TC+CA+CT+CG+GATAAGATGCTG	71
P-ACTB[b]		G+TG+GA+CG+GG+CG+GC+GGATCGGCAAAG	72
P-ACTB-II[b]		A+TC+AT+CC+AT+GG+TG+AGCTGGCGGCGG	73
Padlock probes		**Sequences (5´ - 3´)**	
PP-*KRAS*-wt1[a] *(DP-1)*		GTGGCGTAGGCAAGATCCTAGTAATC*AGTAGCCGTGACTATCGAC T*GGTTCAAAGTGGTAGTTGGAGCTG	74
PP-*KRAS*-G12S[a] *(DP-2)*		GTGGCGTAGGCAAGATTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAAGTGGTAGTTGGAGCTA	75
PP-*KRAS*-G12R[a] *(DP-2)*		GTGGCGTAGGCAAGATTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAAGTGGTAGTTGGAGCTC	76
PP-*KRAS*-G12C[a] *(DP-2)*		GTGGCGTAGGCAAGATTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAAGTGGTAGTTGGAGCTT	77
PP-*KRAS*-wt2[a] *(DP-1)*		TGGCGTAGGCAAGAGTCCTAGTAATC*AGTAGCCGTGACTATCGAC T*GGTTCAAAGGGTAGTTGGAGCTGG	78
PP-*KRAS*-G12D[a] *(DP-2)*		TGGCGTAGGCAAGAGTTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAAGGGTAGTTGGAGCTGA	79
PP-*KRAS*-G12V[a] *(DP-2)*		TGGCGTAGGCAAGAGTTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAAGGGTAGTTGGAGCTGT	80
PP-*KRAS*-G12A[a] *(DP-2)*		TGGCGTAGGCAAGAGTTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAAGGGTAGTTGGAGCTGC	81
PP-*KRAS*-wt3[a] *(DP-1)*		CGTAGGCAAGAGTGCTCCTAGTAATC*AGTAGCCGTGACTATCGAC T*GGTTCAAAGAGTTGGAGCTGGTGG	82
PP-*KRAS*-G13D[a] *(DP-2)*		CGTAGGCAAGAGTGCTTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAAGAGTTGGAGCTGGTGA	83
PP-*KRAS*-wt4[a] *(DP-1)*		GAGGAGTACAGTGCATCCTAGTAAT*CAGTAGCCGTGACTATCGAC T*GGTTCAAAGGACACAGCAGGTCAA	84

PP-*KRAS*-Q61H[a] *(DP-2)*		GAGGAGTACAGTGCACGCTAGATC*CCTCAATGCACATGTTTGGCT CC*GGTTCAAGGACACAGCAGGTCAT	85
PP-EGFR-wt1[a] *(DP-1)*		GGCCAAACTGCTGGGTCCTAGTAAT*CAGTAGCCGTGACTATCGAC T*GGTTCAAAGCACAGATTTTGGGCT	86
PP-EGFR-L858R[a] *(DP-3)*		GGCCAAACTGCTGGGTTCTAGATA*CCTCAATGCTGCTGCTGTACTA C*GGTTCAAGCACAGATTTTGGGCG	87
PP-EGFR-wt2[a] *(DP-1)*		CGTGGACAACCCCCATCCTAGTAAT*CAGTAGCCGTGACTATCGAC* TGGTTCAAAGCTACGTGATGGCCAG	88
PP-EGFR-S768I[a] *(DP-3)*		CGTGGACAACCCCCATTCTAGATA*CCTCAATGCTGCTGCTGTACTA C*GGTTCAAGCTACGTGATGGCCAT	89
PP-EGFR-wt3[a] *(DP-1)*		GCTCCGGTGCGTTCGTCCTAGTAAT*CAGTAGCCGTGACTATCGACT* GGTTCAAAGAGATCAAAGTGCTGG	90
PP-EGFR-G719C[a] *(DP-2)*		GCTCCGGTGCGTTCGTTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAAGAGATCAAAGTGCTGT	91
PP-EGFR-wt4[a] *(DP-1)*		CTCCGGTGCGTTCGGTCCTAGTAAT*CAGTAGCCGTGACTATCGAC*T GGTTCAAAGGATCAAAGTGCTGGC	92
PP-EGFR-G719A[a] *(DP-2)*		CTCCGGTGCGTTCGGTTCTAGATC*CCTCAATGCACATGTTTGGCTC C*GGTTCAATGATCAAAGTGCTGGG	93
PP-TP53-wt1[a] *(DP-3)*		CCCTGCCCTCAACAATTCCTTTTACGA*CCTCAATGCTGCTGCTGTA CTAC*TCTTCGACTTGCACGTACTC	94
PP-TP53-S127F[a] *(DP-4)*		CCCTGCCCTCAACAACTAGTATCTG*AGTCGGAAGTACTACTCTCT*T GTGCCATAAGACTTGCACGTACTT	95
PP-TP53-wt2[a] *(DP-3)*		CTCCTCAGCATCTTATTCCTTTTACGA*CCTCAATGCTGCTGCTGTA CTAC*TCTTCGCGATGGTCTGGCCC	96
PP-TP53-P190S[a] *(DP-4)*		CTCCTCAGCATCTTACTAGTATCTG*AGTCGGAAGTACTACTCTCT*T GTGCCATAAGCGATGGTCTGGCCT	97
PP-ACTB[a] *(DP-3)*		AGCCTCGCCTTTGCCTTCCTTTTACGA*CCTCAATGCTGCTGCTGTA C*TACTCTTCGCCCCGCGAGCACAG	98
PP-ACTB-II[a] *(DP-2)*		AGCCTCGCCTTTGCCTTCCTTTTACGA*CCTCAATGCACATGTTTGG CTCC*TCTTCGCCCCGCGAGCACAG	99
Detection probes		**Sequences (5′ - 3′)**	
DP-1[d]		AGTAGCCGTGACTATCGACT	55
DP-2[d]		CCTCAATGCACATGTTTGGCTCC	54
DP-3[c]		CCTCAATGCTGCTGCTGTACTAC	53
DP-4[a]		AGTCGGAAGTACTACTCTCT	57
+ = LNA-modified base, underline = target complementary sequence, italic = detection probe complementary sequence			
Oligonucleotides were purchased from Integrated DNA Technologies[a], Exiqon[b], Biomers[c] and Eurogentec[d].			

Table 5: Summary of samples that were genotyped for *KRAS* mutations

Mutation analysis of fresh frozen, FFPE and tumor imprint samples						
Sample ID	Sample Type	Target	1. Pyrosequencing		2. *In situ* padlock probe mutation detection	
			Mutants/Total	Mutations	Mutants/Total	Concordance
1-5	Fresh frozen colon	*KRAS*	4/5	1xG12D, 1xG12C, 1xG13D, 1xG12A	4/5	100%
6-10	Fresh frozen lung	*KRAS*	4/5	1xG12D, 1xG12V, 1xG12C, 1xG12S	4/5	100%
11-24	FFPE colon	*KRAS*	14/14	2xG12D, 3xG12V, 2xG12C, 3xG13D, 2xG12S, 1xG12A	14/14	100%
25-26	FFPE lung	*KRAS*	2/2	2xQ61H	2/2	100%
27-35	FFPE lung	*EGFR*	8/9	8xL858R	8/9	100%
36	FFPE lung	*EGFR*	1/1	1xG719C,1xS768I	1/1	100%
37	FFPE lung	*EGFR* / *TP53*	1/1	1xG719A, 1xS127F	1/1	100%
38	FFPE lung	*KRAS* / *TP53*	1/1	1xG12C, 1xP190S	1/1	100%

Mutation analysis of prospective FFPE and tumor imprint samples						
			1. *In situ* padlock probe mutation detection	2. Pyrosequencing		
Sample ID	Sample Type	Target	Mutants/Total	Mutants/Total	Mutations	Concordance
39-46	FFPE lung	*KRAS*	3/8	3/8	2xG12C, 1xG12R	100%
47-54	Colon tumor imprint	*KRAS*	2/8	2/8	1xG12D, 1xG12R	100%
55-79	FFPE colon (from TMA)	*KRAS*	11/25	11/25	6xG12V, 2xG12S, 2xG13D, 1xG12A	100%

Table 6: Summary of oligonucleotides used to sequence KRAS mutations

Oligonucleotides on samples											
Sample ID	**Primers**		**Padlock probes**				**Detection probes**			
1	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12D	PP-ACTB		DP-1	DP-2	DP-3	
2	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt1	PP-*KRAS*-G12C	PP-ACTB		DP-1	DP-2	DP-3	
3	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt3	PP-*KRAS*-G13D	PP-ACTB		DP-1	DP-2	DP-3	
4	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12A	PP-ACTB		DP-1	DP-2	DP-3	
5	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12A	PP-ACTB		DP-1	DP-2	DP-3	
6	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12D	PP-ACTB		DP-1	DP-2	DP-3	
7	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12V	PP-ACTB		DP-1	DP-2	DP-3	
8	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt1	PP-*KRAS*-G12C	PP-ACTB		DP-1	DP-2	DP-3	
9	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt1	PP-*KRAS*-G12S	PP-ACTB		DP-1	DP-2	DP-3	
10	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12A	PP-ACTB		DP-1	DP-2	DP-3	

11-12	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12D	PP-ACTB		DP-1	DP-2	DP-3	
13-15	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12V	PP-ACTB		DP-1	DP-2	DP-3	
16-17	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt1	PP-*KRAS*-G12C	PP-ACTB		DP-1	DP-2	DP-3	
18-20	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt3	PP-*KRAS*-G13D	PP-ACTB		DP-1	DP-2	DP-3	
21-22	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt1	PP-*KRAS*-G12S	PP-ACTB		DP-1	DP-2	DP-3	
23-24	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt2	PP-*KRAS*-G12A	PP-ACTB		DP-1	DP-2	DP-3	
25-26	P-*KRAS*-Q61H	P-ACTB	PP-*KRAS*-wt4	PP-*KRAS*-Q61H	PP-ACTB		DP-1	DP-2	DP-3	
27-35	P-EGFR-L858R	P-ACTB	PP-EGFR-wt1	PP-EGFR-L858R	PP-ACTB-II		DP-1	DP-3	DP-2	
36	P-EGFR-S768I	P-EGFR-G719C/A	PP-EGFR-wt2	PP-EGFR-S768I	PP-EGFR-wt3	PP-EGFR-G719C	DP-1	DP-3	DP-2	
37	P-EGFR-G719C/A	P-TP53-S127F	PP-EGFR-wt4	PP-EGFR-G719A	PP-TP53-wt1	PP-TP53-S127F	DP-1	DP-2	DP-3	DP-4
38	P-*KRAS*-c12/13	P-TP53-P190S	PP-*KRAS*-wt1	PP-*KRAS*-G12C	PP-TP53-wt2	PP-TP53-P190S	DP-1	DP-2	DP-3	DP-4
39-79	P-*KRAS*-c12/13	P-ACTB	PP-*KRAS*-wt1	PP-*KRAS*-G12S	PP-*KRAS*-G12R	PP-*KRAS*-G12C	DP-1	DP-2	DP-3	
			PP-*KRAS*-	PP-*KRAS*-	PP-*KRAS*-	PP-*KRAS*-				

			wt2	G12D	G12V	G12A					
			PP-*KRAS*-wt3	PP-*KRAS*-G13D	PP-ACTB						
Cell lines	**Primers**		**Padlock probes**					**Detection probes**			
ONCO-DG-1	P-*KRAS*-c12/13-II	P-ACTB-II	PP-*KRAS*-wt2	PP-*KRAS*-G12A	PP-ACTB			DP-1	DP-2	DP-3	
A427	P-*KRAS*-c12/13-II	P-ACTB-II	PP-*KRAS*-wt2	PP-*KRAS*-G12D	PP-ACTB			DP-1	DP-2	DP-3	
SW480	P-*KRAS*-c12/13-II	P-ACTB-II	PP-*KRAS*-wt2	PP-*KRAS*-G12V	PP-ACTB			DP-1	DP-2	DP-3	
HCT-15	P-*KRAS*-c12/13-II	P-ACTB-II	PP-*KRAS*-wt3	PP-*KRAS*-G13D	PP-ACTB			DP-1	DP-2	DP-3	
A549	P-*KRAS*-c12/13-II	P-ACTB-II	PP-*KRAS*-wt1	PP-*KRAS*-G12S	PP-ACTB			DP-1	DP-2	DP-3	
HUPT3	P-*KRAS*-c12/13-II	P-ACTB-II	PP-*KRAS*-wt1	PP-*KRAS*-G12R	PP-ACTB			DP-1	DP-2	DP-3	

Table 7: *EGFR* mutations and prevalence based on cases in the COSMIC database.

#	Mutation	Prevalence	
1	L858R	1258	45,48084%
2	2335_2349del15	560	20,24584%
3	2336_2350del15	314	11,35213%
4	2340_2357del18	110	3,97686%
5	T790M	104	3,75994%
6	2339_2348TTAAGAGAAG>C	71	2,56688%
7	2337_2355>T	43	1,55459%
8	2340_2354del15	41	1,48228%
9	L861Q	34	1,22921%
10	2339_2356del18	28	1,01229%
11	G719S	24	0,86768%
12	G719A	23	0,83153%
13	S768I	22	0,79537%
14	2339_2351>C	19	0,68691%
15	2337_2351del15	18	0,65076%
16	2339_2347del9	18	0,65076%
17	2339_2353del15	18	0,65076%
18	G719C	16	0,57845%
19	2307_2308ins9	8	0,28923%
20	2339_2358>CA	7	0,25307%
21	2340_2351del12	7	0,25307%
22	2310_2311insGGT	4	0,14461%
23	2337_2354del18	4	0,14461%
24	2338_2355del18	4	0,14461%
25	2338_2348>GC	4	0,14461%
26	2319_2320insCAC	2	0,07231%
27	2335_2352>AAT	2	0,07231%
28	2338_2352>GCA	2	0,07231%
29	2336_2353del18	1	0,03615%

J. Examples

[00250] Embodiments will now be further described with reference to the following non-limiting Examples. It should be understood that these Examples, while indicating embodiments, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. All documents referenced herein are incorporated by reference.

Materials and Methods

Cell culture:

[00251] The cell lines GM08402 (Coriell Cell Repositories) and BJhTERT were cultured in MEM without phenol red and l-glutamine (Gibco) supplemented with 10% FBS (Sigma), 1× nonessential amino acids (Gibco), 2 mM l-glutamine (Sigma) and 1× penicillin-streptomycin (PEST, Sigma). Mouse embryonic fibroblasts (MEF) were cultured in DMEM without phenol red and l-glutamine (Gibco) supplemented with 10% FBS, 2 mM l-glutamine and 1× PEST. ONCO-DG-1, SW-480, A-427 and HCT-15 (all four from DSMZ), SKOV3 and SKBR3 were cultured in RPMI culture medium (Sigma) supplemented with 10% FBS, 2 mM l-glutamine and 1× PEST. A-549 (DSMZ) was cultured in DMEM without phenol red and L-Glutamine (Gibco) supplemented with 10% FBS, 2 mM L-Glutamine and 1x PEST. HUP-T3 (DSMZ) was cultured in MEM-Eagle culture medium (Sigma) supplemented with 10% FBS, 2 mM L-glutamine and 1x PEST.

Preparation of tissue sections:

[00252] Fresh frozen 9-μm sections of E14.5 mouse embryos were placed on Superfrost Plus Gold slides (Thermo Scientific). Fully anonymized fresh frozen human tissue sections from a HER2-positive breast cancer were obtained from the Fresh Tissue Biobank at the Department of Pathology, Uppsala University Hospital, in accordance with the Swedish Biobank Legislation. Breast tissue sections of 4 μm thickness were placed on Starfrost microscope slides (Instrumedics).

Sample pretreatment for *in situ* experiments:

[00253] Cells were seeded on Superfrost Plus slides (Thermo Scientific) and allowed to attach. When the cells reached the desired confluency they were fixed in 3% (w/v) paraformaldehyde (Sigma) in PBS for 30 min at room temperature (20–23 °C). After fixation, slides were washed twice in DEPC-treated PBS (DEPC-PBS) and dehydrated through a series of 70%, 85% and 99.5% ethanol for 3 min each. The molecular reactions were performed in Secure-seals (Grace Bio-Labs, 9 mm in diameter and 0.8 mm deep) attached to the slides. A 50-µl reaction volume was used for each sample. To make the RNA more readily available for cDNA synthesis, 0.1 M HCl was applied to the cells for 10 min at room temperature. This was followed by two brief washes in DEPC-PBS. Tissues were treated similarly to cell lines, with a few exceptions. Tissue fixation was performed in 2% (w/v) paraformaldehyde in PBS. The tissue was then permeabilized with 0.01% pepsin (Sigma) in 0.1 M HCl at 37 °C for 2 min. Molecular reactions were carried out with a reaction volume of 100 µl in Secure-seals (13 mm in diameter, 0.8 mm deep; Grace Bio-Labs) mounted over the tissue. Reverse transcription was carried out overnight and incubation times for ligation, RCA and detection probe hybridization were doubled. For the mouse tissue, ligation was carried out with T4 DNA ligase.

Oligonucleotide sequences:

[00254] Oligonucleotide sequences (Tables 1–3) were designed using GenBank accession numbers NM_001101.3 (*ACTB*), NM_007393.3 (*Actb*), NM_198253.2 (*TERT*), NM_002467 (*MYC*), NM_001005862.1 (*ERBB2*), NM_009606 (*Acta1*), NM_009609 (*Actg1*) and NM_033360 (*KRAS*). All padlock probes were 5′-phosphorylated at a concentration of 2 µM with 0.2 U μl^{-1} T4 polynucleotide kinase (Fermentas) in the manufacturer's buffer A plus 1 mM ATP for 30 min at 37 °C, followed by 10 min at 65 °C. For β-actin transcript detection in cultured cells, primer P-βe1 was used for detection with padlock probe PLP-βe1, primer P-βe6 with padlock probe PLP-βe6, primer P-βhum with padlock probe PLP-βhum and primer P-βmus with padlock probe PLP-βmus unless otherwise indicated. *TERT* was detected with primer P-TERT and padlock probe PLP-TERT, cMyc with primer P-cMyc and padlock probe PLP-cMyc and *HER2* with primer P-HER2 and padlock probe PLP-HER2. For detection of transcripts in mouse tissue, primer P-βmus was used with padlock probe PLP-βmus for β-actin, primer P-α1mus with padlock probe PLP-α1mus for α1-actin and primer P-γ1mus with padlock probe PLP-γ1mus for γ1-actin. For *KRAS* genotyping, primer P-*KRAS* was used in combination with the padlock probes PLP-*KRAS*-wtGGT, PLP-*KRAS*-mutGTT and PLP-*KRAS*-mutGAT.

Sample preparation for *KRAS* genotyping experiments:

[00255] Cell lines ONCO-DG-1, A-427, SW-480, HCT-15, A-549 and HUP-T3 (all DSMZ) were seeded on Collagen I 8-well CultureSlides (BD BioCoat), and allowed to attach. When the cells reached the desired confluency they were fixed in 3% (w/v) paraformaldehyde (Sigma) in DEPC-treated PBS (DEPC-PBS) for 30 min at room temperature (20–23 °C). After fixation slides were washed twice in DEPC-PBS and the plastic wells were removed from the slides. The slides were thereafter dehydrated through an ethanol series of 70%, 85% and 99.5% ethanol for 1 min each.

[00256] Fresh frozen and FFPE human tumor tissues from colorectal- and lung cancer patients were obtained from the Biobank at the Department of Pathology and Cytology (Botling and Micke, 2011), Uppsala University Hospital, in accordance with the Swedish Biobank Legislation and Ethical Review Act (Uppsala Ethical Review Board approval, reference numbers 2006/325 and 2009/224).

[00257] Tape transferred fresh frozen tissue sections (4 µm) on Starfrost microscope slides (Instrumedics) were prepared from fresh frozen tumor samples stored at -80 °C. The slides were fixed in 3% (w/v) paraformaldehyde in DEPC-PBS for 45 min at room temperature and then permeabilized with 0.01% pepsin (Sigma, #P0609) in 0.1 M HCl at 37 °C for 2 min followed by a brief wash in DEPC-PBS.

[00258] Touch imprints, prepared on Superfrost Plus microscope slides, were obtained from fresh surgical colorectal and lung cancer specimens. After slide preparation the slides were air-dried for 1 min and thereafter stored at -80 °C. The slides were fixed in 3% (w/v) paraformaldehyde in DEPC-PBS for 30 min at room temperature followed by a brief wash in DEPC-PBS.

[00259] FFPE tissue sections (4 µm) were placed on Superfrost Plus microscope slides (Menzel Gläser and baked for 30 min at 60 °C. The slides were then deparaffinized by immersion in xylene for 15 + 10 min and thereafter gradually rehydrated through an ethanol series (2 x 2 min in 100%, 2 × 2 min in 95%, 2 × 2 min in 70%, and finally for 5 min in DEPC-H_2O). The slides were washed in DEPC-PBS for 2 min before fixation with 4% (w/v) paraformaldehyde in DEPC-PBS for 10 min at room temperature which was followed by another DEPC-PBS wash for 2 min. The FFPE tissue slides were then permeabilized in 2 mg ml-1 Pepsin (Sigma #P7012) in 0.1 M HCl at 37 °C for 10 min. The digestion was stopped by

a wash in DEPC-treated H_2O (DEPC- H_2O) for 5 min followed by a wash in DEPC-PBS for 2 min. Finally, the slides were fixed a second time with 4% (w/v) paraformaldehyde in DEPC-PBS for 10 min at room temperature and washed in DEPC-PBS for 2 min. After completed pretreatments of tissues, the slides were dehydrated through an ethanol series of 70%, 85% and 99.5% ethanol for 1 min each.

[00260] The *KRAS* mutation status of the tissues was analyzed by Pyrosequencing (Pyromark Q24 *KRAS*, Qiagen GmbH, Hilden, Germany) as described previously (Sundstrom *et al.,* 2010).

***In situ* cDNA detection procedure:**

[00261] Samples were preincubated in M-MuLV reaction buffer. Then 1 μM of cDNA primer was added to the slides with 20 U μl^{-1} of RevertAid H minus M-MuLV reverse transcriptase (Fermentas), 500 nM dNTPs (Fermentas), 0.2 μg μl^{-1} BSA (NEB) and 1 U μl^{-1} RiboLock RNase Inhibitor (Fermentas) in the M-MuLV reaction buffer. Slides were incubated for 3 h to overnight at 37 °C. After incubation, slides were washed briefly in PBS-T (DEPC-PBS with 0.05% Tween-20 (Sigma)), followed by a postfixation step in 3% (w/v) paraformaldehyde in DEPC-PBS for 30 min at room temperature. After postfixation, the samples were washed twice in PBS-T. To make the target cDNA strands available for padlock probe hybridization, the RNA portion of the created RNA-DNA hybrids was degraded with ribonuclease H. This was performed in the same step as the padlock probe hybridization and ligation. For most reactions, Ampligase (Epicentre) was used for ligation. Samples were first preincubated in Ampligase buffer (20 mM Tris-HCl, pH 8.3, 25 mM KCl, 10 mM $MgCl_2$, 0.5 mM NAD and 0.01% Triton X-100). Ligation was then carried out with 100 nM of each padlock probe in a mix of 0.5 U μl^{-1} Ampligase, 0.4 U μl^{-1} RNase H (Fermentas), 1 U μl^{-1} RiboLock RNase Inhibitor, Ampligase buffer, 50 mM KCl and 20% formamide. Incubation was performed first at 37 °C for 30 min, followed by 45 min at 45 °C. For detection of actin transcript isoforms in mouse embryonic tissue sections, ligation was instead carried out using T4 DNA ligase (Fermentas). Samples were then first preincubated in T4 DNA ligase buffer (Fermentas). Then 100 nM of each padlock probe was added with 0.1 U μl^{-1} T4 DNA ligase, 0.4 U μl^{-1} RNase H, 1 U μl^{-1} RiboLock RNase Inhibitor and 0.2 μg μl^{-1} BSA in T4 DNA ligase buffer supplemented with 0.5 mM ATP and 250 mM NaCl. Slides were then incubated at 37 °C for 30 min. After ligation with Ampligase or T4 DNA ligase, slides were washed in DEPC-treated 2× SSC with 0.05% Tween-20 at 37 °C for 5 min

and rinsed in PBS-T. Slides were preincubated briefly in Φ29 DNA polymerase buffer (Fermentas). RCA was then performed with 1 U μl^{-1} Φ29 DNA polymerase (Fermentas) in the supplied reaction buffer, 1 U μl^{-1} RiboLock RNase Inhibitor, 250 μM dNTPs, 0.2 $\mu g\ \mu l^{-1}$ BSA and 5% glycerol. Incubation was carried out for 60 min at 37 °C. The incubation was followed by a wash in PBS-T. RCPs were visualized using 100 nM of each corresponding detection probe in 2× SSC and 20% formamide at 37 °C for 30 min. Slides were then washed in PBS-T, the Secure-seals were removed and the slides were dehydrated using a series of 70%, 85% and 99.5% ethanol for 3 min each. The dry slides were mounted with Vectashield (Vector), containing 100 ng ml^{-1} DAPI to counterstain the cell nuclei. The protocol for counterstaining of cell membranes in Figure 5 is described under "WGA Staining" below.

WGA staining:

[00262] For counterstaining of cytoplasms 2.5 $\mu g\ ml^{-1}$ WGA 488 (Invitrogen) diluted in 1× PBS was added for 60 min at room temperature. This was followed by two washes in PBS-T and dehydration before mounting and nuclear staining with DAPI as described before.

Single-cell quantification:

[00263] For single-cell quantification in Figure 6, a custom made MatLab script was used for marking individual cells and counting RCPs within the marked areas. The quantification of RCPs in MatLab differs in how an RCP is defined compared to the BlobFinder software used for quantification in other Examples herein. As a consequence the results show ~30% fewer RCPs compared to the BlobFinder analysis.

Image acquisition and analysis:

[00264] Images of cultured cells were acquired using an Axioplan II epifluorescence microscope (Zeiss) equipped with a 100 W mercury lamp, a CCD camera (C4742-95, Hamamatsu), and a computer-controlled filter wheel with excitation and emission filters for visualization of DAPI, FITC, Cy3, Cy3.5 and Cy5. A ×20 (Plan-Apocromat, Zeiss), ×40 (Plan-Neofluar, Zeiss) or ×63 (Plan-Neofluar, Zeiss) objective was used for capturing the images. Images were collected using the Axiovision software (release 4.3, Zeiss). Exposure times for cell images were 260–340 ms (at ×20 magnification), 10–80 ms (×40) or 220 ms (×63) for DAPI; 40 ms (×40) or 220 ms (×63) for FITC; 560–640 ms (×20), 110–160 ms (×40) or 200 ms (×63) for Cy3; 110 ms (×40) or 250 ms (×63) for Texas Red; and 6,350 ms (×20), 180 ms (×40) or 350 ms (×63) for Cy5. For SKBR3 and SKOV3 cells, images were

collected as z-stacks to ensure that all RCPs were imaged. The imaging of α1-actin and β-actin in fresh frozen mouse embryonic tissue sections in Example 3 was imaged using a Mirax Midi slide scanner (3D Histech) equipped with a CCD camera (AxioCam MRm, Zeiss) and a ×20 Plan-Apochromat objective. Exposure times in the slide scanner were 45 ms for DAPI, 270 ms for Cy3, 340 ms for Texas Red and 3,200 ms for Cy5. For quantification, the numbers of RCPs and cell nuclei in images were counted digitally using BlobFinder software (version 3.0_beta). For cultured cells, the quantification was done on five 20× microscope images (approximately 20–30 cells for each sample). The total number of RCPs was divided by the number of nuclei for each image. The average for each sample was then calculated from the result of the five images and is reported as RCPs per cell. The procedure for single-cell quantification used in Figure 6 is described "Single cell quantification" above.

qPCR for β-actin transcript quantification in cells:

[00265] Two separate passages of the cell line GM08402 were collected after counting of cells, and total RNA was purified from the cells using the PARIS kit (Ambion) with the protocol for RNA isolation from total cell lysate. Traces of DNA were removed from the purified RNA using the DNA-free kit (Ambion). First-strand cDNA synthesis was carried out with 700 ng of template RNA in a mix containing 20 U RevertAid H minus M-MuLV reverse transcriptase (Fermentas) in the corresponding enzyme buffer, 0.5 μg oligo(dT) primer (20-mer), 1 mM dNTPs and 1 U μl−1 RiboLock RNase Inhibitor. Samples were incubated at 37 °C for 5 min, followed by 42 °C for 60 min. The reaction was stopped by heating to 70 °C for 10 min. A preparative PCR was carried out to synthesize template for standard curve creation. For this PCR, 1 μl of cDNA from one of the cell passages was amplified in a mix of 0.02 U μl−1 Platinum Taq DNA polymerase (Invitrogen), PCR buffer, 2 mM MgCl2, 200 μM dNTPs, 200 nM ACTBfwd primer and 200 nM ACTBrev primer in a total volume of 50 μl. PCR was carried out with 2 min at 95 °C, followed by cycling 45 times (95 °C for 15 s, 50 °C for 15 s, and 72 °C for 1 min) and finishing with 72 °C for 5 min. The PCR product was purified using the Illustra GFX PCR and gel band purification kit (GE Healthcare) according to the protocol for purification of DNA from solution. The concentration of the purified PCR product was measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific) and the number of molecules per microliter was calculated. qPCR was run with 2 μl of template cDNA, or diluted standard curve PCR product, with SYBR Green (Invitrogen), 0.02 U μl−1 Platinum Taq DNA polymerase, PCR buffer, 2 mM MgCl2, 200 μM dNTPs, 200 nM ACTBfwd primer and 200 nM ACTBrev primer in a total volume of 30 μl. The qPCR was

run using the same program as for the preparative PCR. Standard curve samples were run in duplicates of the same sample and cDNA samples from the two passages of cells were run in triplicates. Calculations of transcript copy numbers for the two cell passages were based on the number of counted cells at harvest. The average β-actin mRNA copy number for the cell line was then determined. The protocol for efficiency estimation by qPCR for the in situ multiplex detection experiment is as follows:

[00266] The cell lines GM08402, SKBR3 and BJhTERT were harvested after counting of cells and total RNA was purified from the cells using the RiboPure kit (Ambion). DNA traces were removed from the purified RNA using the DNA-free kit (Ambion). RNA concentration and quality was investigated on an Agilent Bioanalyzer using a RNA 6000 Pico chip (Agilent). First strand cDNA synthesis was carried out using the High capacity cDNA reverse transcription kit (Applied Biosystems). The prepared cDNA was diluted 4× before analysis with TaqMan qPCR. PCR primers and TaqMan probes were purchased as validated 20× TaqMan Gene Expression Assays from Applied Biosystems (assay no Hs99999903_m1 for β-actin, Hs00972650_m1 for TERT, and Hs99999005_mH for HER2). Templates for standard curves for the different genes were created by PCR. For this PCR, 1 µl of cDNA from the BJhTERT cell line was amplified in a mix of 0.02 U µl-1 Platinum Taq DNA polymerase (Invitrogen), 1× PCR buffer, 2 mM MgCl2, 200 µM dNTP, and 0.01× of each primer mix (0.2 µM of each primer) in separate reactions for the different genes. The total PCR volume was 50 µl and the PCR was carried out with 2 min at 95 °C, followed by cycling 45 × (95 °C for 15 s and 60 °C for 1 min), and finished with 60 °C for 5 min. The PCR products were purified using the Illustra GFX PCR and gel band purification kit (GE Healthcare). The concentration of the purified PCR products was measured using a Nanodrop 1000 spectrophotometer (Thermo Scientific) and the number of molecules per µl was calculated. The qPCR was run with 4 µl of template cDNA, or standard curve PCR product in 1 × TaqMan Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems) with 1 × TaqMan Gene Expression Assay primer and probe mix in a total volume of 20 µl. The qPCR program was run with 10 min at 95 °C, followed by cycling 40 × with 95 °C for 15 s and 60 °C for 1 min. All samples were run in duplicates and featured serial dilutions of the standard curves, serial diluted cDNA samples, RNA controls from the cell lines, and no template controls. Calculations of transcript copy numbers were based on the number of counted cells at harvest.

***In situ* genotyping of *KRAS* on cell lines and tissues:**

[00267] All the molecular *in situ* reactions were carried out in Secure-seals (Grace Bio-Labs Inc.) and the reaction volumes for tissues or imprints were either 100 µl (size 13 mm diameter, 0.8 mm deep) or 350 µl (size 22 mm diameter, 0.8 mm deep) depending on the size of the sample. The Secure-seals that were used for cells had a total volume of 50 µl (size 9 mm diameter and 0.8 mm deep). The Secure-Seals were mounted over the cells or tissues and the wells were dehydrated by a brief flush with PBS-T (DEPC-PBS with 0.05% Tween-20 (Sigma)).

[00268] The samples were thereafter treated in the same way with just the following exceptions. Post-fixation of fresh frozen and FFPE tissues was performed for 45 min compared to 30 min for cell lines imprints. Also, the RCA time on tissues was longer (8 h) compared to cultured cells and tumor imprints (2 h). For all reactions slides were incubated in humid chambers.

Oligonucleotide sequences for *KRAS* genotyping experiments:

[00269] Oligonucleotides sequences (Table 4) were designed using GenBank accession numbers NM_033360 (*KRAS*), NM_005228 (*EGFR*), NM_001126114.1 (*TP53*) and NM_001101.3 (*ACTB*). All padlock probes were 5' phosphorylated at a concentration of 10 µM with 0.2 U µl-1 T4 PNK (Fermentas) in PNK buffer A and 1 mM ATP for 30 min at 37 °C, followed by 10 min at 65 °C. The primers, padlock probes and detection probes applied on the different tissue samples and cell lines are summarized in Table 6.

[00270] One µM of cDNA primer was added to the slides with 20 U µl-1 of RevertAid H minus M-MuLV reverse transcriptase (Fermentas), 500 µM dNTP (Fermentas), 0.2 µg µl-1 BSA (NEB), and 1 U µl-1 RiboLock RNase Inhibitor (Fermentas) in the M-MuLV reaction buffer. Slides were incubated for 3 hours at 37 °C.

[00271] After incubation slides were washed briefly by flushing the wells in PBS-T, followed by a post-fixation step in 3% paraformaldehyde (w/v) in DEPC-PBS for 45 (fresh frozen and FFPE tissues) or 30 (imprints) minutes at room temperature. After post-fixation the samples were washed by flushing the Secure-seals chambers with PBS-T.

RNase H digestion, padlock probe hybridization and ligation for *KRAS* genotyping experiments:

[00272] To create single-stranded target cDNA available for padlock probe hybridization, the RNA part of the created RNA-DNA hybrids was degraded with RNase H. This was performed in the same step as hybridization and ligation of the padlock probes. The reaction was carried out with 100 nM of each padlock probe in a mix of 1 U μl^{-1} Ampligase (Epicentre), 0.4 U μl^{-1} RNase H (Fermentas), 1 U μl^{-1} RiboLock RNase Inhibitor, 50 mM KCl, 20% formamide in Ampligase buffer,. Incubation was performed first at 37 °C for 30 min, followed by 45 min at 45 °C. After ligation, slides were washed flushing the chambers with PBS-T. For prospective *KRAS* mutation detection of unknown tissue samples a cocktail of all *KRAS* codon 12 and 13 padlock probes was mixed with a final concentration of 10 nM.

Amplification and detection of circularized padlock probes for *KRAS* genotyping experiments:

[00273] RCA was performed with 1 U μl^{-1} Φ29 DNA polymerase (Fermentas) in the supplied reaction buffer with 1 U μl^{-1} RiboLock RNase Inhibitor, 250 μM dNTP, 0.2 $\mu g\ \mu l^{-1}$ BSA, and 5% glycerol. Incubation was carried out for 2 h for tumor imprints as well as for cell lines and approximately 5 h for fresh frozen and FFPE tissues at 37 °C. After RCA the samples were washed flushing the Secure-seals chambers with PBS-T. RCPs were visualized using 100 nM of each corresponding detection probe in 2× SSC and 20% formamide at 37 °C for 15 min. Slides were then washed again by flushing the chambers in PBS-T, the Secure-seals were removed and the slides were dehydrated using a series of 70%, 85%, and 99.5% ethanol for 30 sec each. The dry slides were mounted with Vectashield (Vector), containing 100 ng ml^{-1} DAPI to counterstain the cell nuclei.

Image acquisition and analysis for *KRAS* genotyping experiments:

[00274] Images were acquired using an AxioplanII epifluorescence microscope (Zeiss), equipped with a 100 W mercury lamp, a CCD camera (C4742-95, Hamamatsu), and a computer-controlled filter wheel with excitation and emission filters for visualization of DAPI, FITC, Cy3 and Cy5. For capturing the images, a ×10 (Plan-Apocromat, Zeiss) objective was used for fresh frozen and FFPE tissues, a ×20 (Plan-Apocromat, Zeiss) objective for tumor imprints and finally a ×63 (Plan-neofluar, Zeiss) objective was used for the cells. Images were collected using the Axiovision software (Release 4.8, Zeiss). Images displayed for illustrations were processed using image editing software for clarity in print.

The threshold for different color channels was set using ImageJ 1.42q and for clearer visualization of the *KRAS* signals in Cy3 and Cy5, a maximum filter was applied.

Example 1:
Detection of β-actin (*ACTB*) transcripts in cultured human cells using padlock probes

[00275] To detect β-actin (*ACTB*) transcripts in cultured human cells, two different padlock probes were used targeting sequences in the first and last exon, respectively. Many bright, spot-like signals localized to the cytoplasm of cells were visualized, consistent with previous observations of this transcript consistent with previous reports regarding this transcript. The detection efficiency was similar for the two padlock probes, indicating that in this case detection was not highly dependent on target position along the transcript (FIG. 5). In contrast, when reverse transcriptase was omitted from the cDNA synthesis reaction, no signals were detected, verifying that the signals were cDNA dependent (FIG. 5). It was estimated that the overall *in situ* detection efficiency to be ~30% of available transcripts, on the basis of a comparison to quantitative PCR (qPCR) data for β-actin mRNA in the GM08402 cell line (2,000 copies per cell). There was considerable variation in the number of signals among cells (FIG. 6), consistent with other reports of intercellular variation in β-actin mRNA expression.

Example 2:
Detection of single-nucleotide variants of transcripts in cultured cells *in situ*

[00276] To demonstrate high selectivity of detection, an assay was used to detect of single-nucleotide variants of transcripts *in situ*. Expressed polymorphisms are rare in β-actin, therefore a single-base difference between the human and mouse β-actin sequences was used as genotyping target. Co-cultured human and mouse fibroblast cells were subjected to *in situ* genotyping of cDNA using padlock probes PLP-βhum (human) and PLP-βmus (Mus musculus) and target-primed RCA. There was a clear-cut distinction observed between the two subpopulations of cells in the co-culture. The preference for perfectly matched padlock probes at the circularization step ensures distinction between the two targets by the ligase.

Example 3:
Detection of single-nucleotide variants of transcripts in fresh frozen tissue *in situ*

[00277] To test the method in fixed tissue sections, closely related skeletal muscle α1-actin (*Acta1*) and cytoplasmic β-actin (*Actb*) transcripts were targeted in fresh frozen tissue from an E14.5 mouse embryo cross sectioned at the level of the neck. The two actin transcripts were successfully detected in the tissue using padlock probes designed with target sequences differing by a single base. The α1-actin signals were mainly distributed to skeletal muscles, whereas β-actin signals were widely distributed but showed slightly more signals in the non-muscular tissue. The ability to distinguish three transcripts from the same gene family was demonstrated by including a probe specific for the cytoplasmic α1-actin (*Acta1*) transcript.

Example 4:
Detection of transcripts for expression profiling

[00278] To test the method's ability for multiplex detection of transcripts for expression profiling, padlock probes were designed for the three cancer-related transcripts *HER2* (also known as *ERBB2*), *cMyc* (also known as *MYC*) and *TERT*. Using β-actin as a reference transcript, these transcripts were assayed in four cell lines (a human ovarian carcinoma cell line, a human breast carcinoma cell line, a *TERT*-immortalized human foreskin fibroblast cell line and a primary fibroblast cell culture). The levels of expression of the cancer-related genes differed among the cell lines (FIG. 2a–d). The ovarian and breast carcinoma cell lines showed similar patterns of expression of the *HER2* and *cMyc* transcripts, whereas the *TERT*-immortalized fibroblast was the only cell type with a detectable level of the *TERT* transcript. All four cell lines expressed β-actin, and in the normal fibroblasts this was the only investigated transcript expressed at a detectable level. These results were compared to qPCR data and to available literature and good correlation with the expected relative expression levels in the different cell lines and a notable consistency in detection efficiency among the different transcripts was found (Example 5). Large cell-to-cell variation in expression for all investigated transcripts was noticed, which is consistent with previous studies of expression in single cells in cultures.

Example 5:

Expression of cancer-related transcripts in human cell lines

[00279] The three cancer-related transcripts *TERT, HER2*, and *cMyc* were assayed in four cell lines as described in Example 4. All cell lines expressed the housekeeping gene β-actin, but differed in the expression of the cancer-related transcripts according to the *in situ* data. qPCR measurements were then performed to quantify the different transcripts in the GM08402, BJhTERT and SKBR3 cell lines to be able to evaluate the variation in detection efficiency in the *in situ* experiments. qPCR measurements of *TERT* expression showed relatively high expression in the BJhTERT cell line (247 molecules/cell), as well as low expression in the SKBR3 breast carcinoma cell line (6 molecules/cell). No TERT expression was detected in the normal primary fibroblasts by qPCR. The qPCR data for TERT correlates well with the mRNA expression level for TERT found in the literature (220 molecules/cell for BJhTERT and 0.57 molecules/cell for SKBR3 (Yi *et al.,* 2001). The *in situ* result of 39 RCPs detected per cell in BJhTERT thus corresponds to a detection efficiency of 16% based on the qPCR data. The HER2 transcript is known to be overexpressed in the ovarian and breast carcinoma cell lines. In the SKBR3 cell line the number of *HER2* mRNAs/cell is reported to be 168-336 molecules, the qPCR measurement described herein ended at 177 molecules/cell. The number of *HER2* mRNAs/cell detected *in situ* was 25. This gives a detection efficiency of 14% for the *HER2*-transcript in SKBR3 cells. *HER2* expression could not be detected by qPCR in the normal primary fibroblasts or in the BJhTERT cell line. The expression level of *cMyc* in SKOV3 cells is estimated to about one quarter of the number of *HER2* transcripts in the same cell line, which correlates well with the *in situ* measurement described herein. When assayed alone, the detection efficiency for β-actin in cultured fibroblast cells was estimated to be 30% based on qPCR measurements and *in situ* detection of the transcript. In these multiplex experiments the detection efficiency is slightly lower, about 15%, based on the same qPCR estimation. A similar effect is observed among the cancer transcripts that show detection efficiencies of about 15% in multiplex, while they perform better individually. It is likely that the lower detection efficiency observed for targets in multiplex experiments are due to interactions between different padlock probes and/or cDNA primers, especially since the LNA modified bases of the cDNA primers have the capacity to bind very strongly to each other. The detection protocol for multiplex reactions can be improved by optimizing the concentration of the probes and/or primers. Further qPCR measurements show good correlation with the *in situ* measurements for the relative β-actin

expression level between the cell lines. Taken together these data indicate that the relative levels of RCPs in the different cell types are good estimates of the true relative transcript levels in the cell populations. Thus it is believed that the method is suitable for relative expression profiling in different samples. Although as the reverse transcription reaction is known to introduce variation in mRNA quantification by qPCR, this is likely to be the case also for reverse transcription *in situ*.

Example 6:

Detection of transcript distribution in a fresh frozenHER2-positive human breast cancer tissue

[00280] The technique of this embodiments was also used to assess *HER2* transcript distribution in a fresh frozen *HER2*-positive human breast cancer tissue section. Expression varied widely among the cells, consistent with the expected presence of cancer cells and normal stroma in the tumor tissue.

Example 7:

Genotyping of a *KRAS* point mutation in *KRAS* wild-type and mutant cells

[00281] The method of the embodiment was also used to genotype a *KRAS* point mutation in *KRAS* wild-type and mutant cells (FIG. 7). The different cell types could be clearly distinguished on the basis of the color of their corresponding RCPs. Activating mutations of the KRAS oncogene are found in 17%–25% of all human tumors, and assays to monitor these mutations and other tumor cell–specific markers in tissue specimens *in situ* could be of great value for clinical pathology investigations. The potential for studies of allelic expression was further investigated by analyzing 77 cells from a cell line heterozygous for a point mutation in *KRAS*. An average allelic ratio of 48% wild-type transcripts was observed, with considerable cell-to-cell variation (FIGs. 6b and 7), indicating a balanced allelic transcription. In this experiment, all heterozygous cells with more than seven RCPs showed signals from both alleles. For cells showing fewer than seven signals, it will be difficult to determine the potential for biallelic expression and extent of unbalanced allelic expression in single cells.

Example 8:
Effect of LNA base incorporation in the cDNA primer

[00282] To increase the efficiency of the reverse transcription step, a RT primer with incorporated locked nucleic acid (LNA)-bases was used. LNA modified oligonucleotides have previously been used for FISH, with DNA/LNA mixmers with every second or third base substituted for LNA performing the best. In addition to the increased hybridization efficiency to the targets, the LNA content of the primers can be designed to protect the target RNA from breakdown by RNase H. This means that in the present method, the *in situ* synthesized cDNA can maintain the localization to the detected mRNA molecule in the cell via the hybridization of the cDNA primer (FIG. 1). cDNA primers with different LNA substitutions (Table 2) were tested *in situ* for subsequent detection of the PLP-βe1 padlock probe target site. It was found that primers with every second base at the 5′-end substituted with LNA performed better than primers with substitutions of every third base (FIG. 3). Primers with five, seven or nine LNA bases in total were also investigated and it was found that adding nine LNA bases resulted in a small decrease in the amount of signals *in situ*. To ensure that the LNA would not interfere with the ability of the reverse transcriptase to synthesize cDNA from the primer, LNA bases were placed on the 5′-side of the primers, leaving the 3′-end unmodified. It was found that shortening the total length of the primer from 30 to 25 nucleotides did not influence the results, and thus it was concluded that the priming is not disturbed by the presence of LNA bases in the primer.

Example 9:
cDNA synthesis efficiency

[00283] To ensure an optimal distance between hybridizing cDNA primers and target sequences for the padlock probes, the length of the produced cDNA molecules in cells was investigated. An *in situ* detection experiment was set with cDNA primers located at different distances from the 5′-end of the β-actin mRNA. Reverse transcription was then performed *in situ* and the resulting cDNA molecules were detected with PLP-βe1, with a target sequence near the 3′-end of the reverse transcribed cDNA. The number of RCPs formed per cell was then quantified for the different primers. The primers tested were to result in cDNA molecules ranging from approximately 90-500 nt in length, measured from the start of the primer site to the end of the transcript (see Table 3 for primer sequences). It was found that

predominantly short molecules were formed and that the cDNA primer site should be located close to the padlock probe target site (FIG. 4). As well as providing details on how to design primers for reverse transcription, the knowledge about the limited cDNA synthesis length has a practical relevance for the execution of the RCA reaction. In this protocol a target-priming strategy was used that was originally described for endogenous mitochondrial DNA molecules *in situ* (Larsson *et al.,* 2004). Target-priming takes advantage of a 3′-5′ exonuclease activity of the ϕ29 DNA polymerase on single stranded DNA to create a primer from a nearby 3′-end of the target molecule. The efficiency of the RCA reaction has been shown to decrease as the length of the protruding 3′-end of mitochondrial DNA is increased from 0 to 130 nucleotides. As very short cDNA molecules were produced in this method, the target-primed RCA approach can efficiently be applied for signal amplification also for cDNA detection without further preparation of the target strand.

Example 10:

Different ligases for ligation of padlock probes

[00284] There are mainly two enzymes that have been used for padlock probe ligation previously; the ATP dependent T4 DNA ligase and the NAD+ dependent Ampligase™. Both ligases were tested for *in situ* detection of cDNA with padlock probes, good detection efficiencies were obtained. However, when performing the experiments for detection of sequences with single nucleotide resolution in human and mouse cells, it was found that Ampligase resulted in a lower proportion of signals from the non-matched probe. The proportion of correct signals with T4 DNA ligase was 87% (human RCPs/total RCPs) for human cells and 98% (mouse RCPs/total RCPs) for mouse cells. This is in contrast to Ampligase™, which had a much higher selectivity for the human target sequence (98% correct) whereas the mouse target sequence was unchanged compared to T4 DNA ligase. As the transcripts of the different actin isoforms share a high proportion of similarity and many pseudogenes exist, it is believed that some of these unexpected positive signals originate from sequences similar to the padlock probe target sequence that do not show up when performing simple *in silico* sequence analysis. In addition to these observations, Ampligase™ is known to be more specific for matched substrates than T4 DNA ligase.

Example 11:

Assay design for *in situ* mutation detection

[00285] Padlock probes were designed for point mutations of *KRAS* in codon 12 and 13 (G12S, G12R, G12C, G12D, G12A, G12V and G13D) and codon 61 (Q61H), as well as for *EGFR* (G719A, G719C, S768I and L858R) and *TP53* (S127F and P190S). Padlock probes for the wild-type forms of the different targets were designed as well. The mutation-specific padlock probes were designed with identical target sequences except for the last nucleotide in the 3′-end that differ depending on genotype. Mismatches at this position are not accepted by the DNA ligase used and single nucleotide differences, like point mutations, are therefore efficiently discriminated. There are furthermore two different sites for detection probes for wild-type and mutant padlocks to distinguish the RCPs from each other using detection probes labeled with different fluorescence dyes, e.g. green and red. Also detection of the *ACTB* transcript was included in these assays, detected by an additional fluorophore, as an internal reference having a relative constant expression between cell types. A comparison of the *ACTB* signals across samples provided an estimation of the detection efficiency in different samples. The *ACTB* data has been useful during the development phase of this study, but turned out to be dispensable for mutation scoring and tissue classification.

Example 12:

Mutation detection in fresh frozen colon and lung tissues with known *KRAS* status

[00286] The selectivity of the padlock probes was first tested *in situ* on wild-type- and mutation-specific *KRAS* cell lines. After confirmation of the quality of the probes, our *in situ* genotyping method was applied on ten fresh frozen human colon and lung cancer tissues with known *KRAS* status. In this validation phase, each probe-pair (one probe for a particular mutation and one for the corresponding wild-type variant) was tested individually on a collection of fresh-frozen tissue samples with known *KRAS* status. Wild-type probes were designed to generate green fluorescence RCPs and mutation-specific probes to generate red fluorescence RCPs. The samples represented all codon 12 and 13 mutations except for the rarest one, G12R. However, the performance of the padlock probe pair for the G12R mutation was still verified for specificity on one of the tested cell lines. Thus, *KRAS* wild-type tumor tissues could be distinguished from ones having tumors carrying activating *KRAS* mutations by microscopic visualization in a fashion similar to regular fluorescent *in situ*

hybridization (FISH). The colon and lung sections with *KRAS* mutations displayed a mixture of signals originating from both of the probes in the padlock probe pair, whereas the normal tissues showed signals exclusively from the wild-type padlock probe. By visually examining the ten samples variations can clearly be seen in *KRAS* expression levels both within and between the tissues. Overall, a slightly higher expression level of *KRAS* was noticed in lung compared to colon. The results showed that most cases displayed both wild-type and mutant *KRAS* signals in the tumor cell areas indicating heterozygous expression. In contrast, one lung sample almost exclusively displayed mutant signals in the tumor regions while the few existing wild-type signals belonged to the normal surrounding stroma. This could reflect a *KRAS* homozygous mutation or loss-of-heterozygosity (LOH).

Example 13:

Mutation detection in FFPE tissue

[00287] The *in situ* padlock probe technique was tested to evaluate whether it could be applicable on FFPE tissue. The protocol applied on this type of tissue material was essentially the same as for fresh frozen tissues, except for the pretreatment procedure. *KRAS* mutation analysis was performed on a collection of 14 colorectal FFPE cancer tissues with known *KRAS* mutations in codon 12 and 13 applying the respective padlock probe-pair. All tissues displayed a mixture of signals originating from both the wild-type and mutant padlock probe, however variation in the number of signals (for both *KRAS* and *ACTB*) were significant between tissues, which probably reflects the expected difference in tissue quality among FFPE samples. Moreover, the ratio between wild-type and mutant signals was also observed to differ between tissues carrying the same *KRAS* mutation which probably reflects tumor-specific characteristics. Probes were also designed for the most common mutation in codon 61 (Q61H) and tested in two colon tumor FFPE samples that successfully were scored as mutants.

Example 14:

In situ detection of *KRAS* mutations on prospective clinical samples with unknown mutation status

[00288] After the initial verification that the padlock probes are selective, all probes were combined into single reactions that could answer the primary diagnostic question whether a

case is *KRAS* positive or not. This was tested by comparing *in situ* mutation detection using single pairs of *KRAS* mutation-specific padlock probes with a multiplex detection approach using a padlock probe cocktail containing all probes for *KRAS* codon 12 and 13 mutations. The results, based on visual examinations of the tissues, indicated that neither efficiency nor selectivity were lost when multiple probes were in competition for the two-codon target site. The analysis thus provides a rapid answer if the tumor harbors an activating KRAS mutation or not. Nevertheless, if requested there is still a possibility with this technique to reveal the exact sequence alteration by simply testing for all mutations individually on consecutive sections.

[00289] Multiplex mutation detection was thereafter demonstrated on eight prospective lung FFPE tissues with unknown KRAS mutations status. Approximately 15-30% of all lung cancer cases have activating KRAS mutations. After performing mutation analysis with padlock probes and RCA, three of the eight cases were concluded to be mutated. The results were compared with pyrosequencing on the same tissues and the suggested genotypes were confirmed to be correct for every case.

[00290] To test the method in a diagnostic setting involving cytology preparation tumor imprint slides were prepared from eight prospective fresh colon cancer specimens with unknown KRAS mutation status. Multiplex KRAS mutation detection using padlock probes and target-primed RCA were prepared using the protocol for unfixed tissue. By microscopic examination of the imprints, two cases were found to be positive in the in situ mutation assay, while the other six tumor imprints only showed wild-type signals. DNA from corresponding FFPE tumor sections from the same cases were thereafter tested for *KRAS* mutations by pyrosequencing. The pyrosequencing results were completely concordant with the *in situ* assay.

Example 15:

High-throughput mutation screening on tissue microarrays

[00291] Tissue microarrays (TMA) can be used to analyze hundreds of patient FFPE tumor samples on one slide, and have been used to characterize protein expression (by immuno-histochemistry (IHC)) and gene copy number variations (by FISH) in large patient cohorts. Here a TMA containing 25 FFPE colon samples (in duplicates) was assayed for possible *KRAS* codon 12 and 13 mutations. The array consisted of samples from normal colon

mucosa, tubular adenomas, serrated adenomas, primary tumors and matched metastasis, all with unknown mutation status for *KRAS*. Of all samples eleven were found to be *KRAS* positive - two adenomas, one serrated adenoma, four primary tumors and their matched metastasis. Mutation analysis by pyrosequencing on the corresponding FFPE blocks was completely concordant with the *in situ* data.

Example 16:

Differential expression of mutated oncogene alleles related to tumor progression and histological heterogeneity

[00292] Variable expression of a mutated oncogene across a tumor could potentially result in a variable response to targeted therapy in different areas of a single cancer lesion. Therefore, cases were screened with the *in situ* assay for distinct patterns of expressed mutations. In one colon cancer case with a codon 61 mutation, the histological progression from normal colon mucosa to low-grade and high-grade dysplasia and invasive carcinoma could be visualized on a single slide. There was a clear increase in the expression of the mutation along with tumor progression. Thus, one can speculate if the level of resistance to EGFR inhibitors would follow the expression levels in the different neoplastic compartments.

[00293] Also, the EGFR L858R mutation was targeted in a set of nine FFPE lung tissues in which eight were known to be positive. The results from the in situ mutation assay were completely concordant with the DNA sequencing data. Even though some of the lung samples were collected more than a decade ago high detection efficiency was observed with high numbers of signals, especially mutant signals, which might reflect high mRNA expression from amplified *EGFR* in the tumor. In one lung sample a great histological heterogeneity was observed with regard to tumor growth patterns. Wild-type *EGFR* was only expressed in normal bronchial epithelium. In areas with bronchioalveolar/lepidic growth pattern the expression of mutated *EGFR* was low, and equaled the expression of the wild-type allele. The expression of the mutant allele increased in more poorly differentiated glandular areas, both in absolute numbers and relative to the wild-type allele. The expression of mutant *EGFR* peaked in areas with solid growth pattern. Thus, if the expression level of L858R affects the sensitivity of a tumor clone for EGFR-TKI therapy, the poorly differentiated areas of the tumor would be expected to respond better than the well differentiated areas in this individual tumor.

Example 17:

Expression patterns in tumors with multiple mutations

[00294] To further study intra-tumor heterogeneity, probes were designed for tumors that were known to harbor multiple point mutations. Personalized medicine implies therapy tailored to the individual characteristics of a patient. The advent of next-generation sequencing technology is now increasingly providing researchers, and soon probably clinicians, with mutational profiles of individual tumors that taken together may provide improved opportunities for individualized therapy. Sequencing DNA prepared from a part of a tumor will reveal all mutations in that sample but not if they reside in different sub-clones of the tumor. As a proof-of-concept that intergenic tumor heterogeneity can be studied with our technology, individualized in situ mutation assays were set up for screening of FFPE cases carrying unique combinations of mutations in EGFR, KRAS, and TP53.

[00295] One lung cancer case was positive for the activating EGFR mutation G719C as well as the EGFR S768I mutation that is associated with resistance to anti-EGFR therapy. Both mutation variants were successfully detected with the padlock probe-based in situ technique and their individual expression patterns were identified. The expression of the G719C mutation was high compared to the S768I mutation throughout the tumor section. This balance between the expressed mutated alleles might be expected as that this case represents a patient that had not received anti-EGFR therapy so no selection pressure for increased expression of the resistance mutation was present.

[00296] Another lung FFPE sample was assayed for a G719A *EGFR* mutation in combination with a S127F mutation of the tumor suppressor gene *TP53*. The *in situ* analysis of this tissue showed cells in stromal regions that only expressed the wild-type form of *TP53* while no expression of any of the *EGFR* alleles could be detected. Hematoxylin and eosin (HE) staining of this tissue sample confirmed that the cell populations with wild-type TP53 were lymphocytes. The TP53 S127F mutation-positive tumor regions displayed signals from both the wild-type EGFR and G719A padlock probes but none from wild-type TP53 padlock probe, indicating TP53 LOH.

[00297] A set of padlock probes was applied on a FFPE lung tissue sample with reported KRAS G12C and TP53 P190S mutations. In contrast to the previous case, in which the wild-type and mutant TP53 signals were located in different compartments (stroma and tumor

respectively), here the mutant and wild-type TP53 transcripts were expressed in a heterozygous fashion in the tumor compartment. Similarly the wild-type and mutant KRAS signals were evenly distributed across the tumor areas with a higher expression of mutant compared to wild-type KRAS alleles. This difference in expression pattern of the wild-type and mutant alleles in the two cases would not have been identified unless an in situ technique was included as a complement to DNA sequencing. Moreover, since this in situ assay reveals information on a single cell level, unique information (e.g. expression of more than one mutation in the same cell, can be identified and studied in detail. Co-localization of different alleles in the same cell provides strong evidence of their co-existence in cells in the tumor while absence of co-localization does not prove that they are not co-expressed in a certain cell-lineage. Even though all four alleles were not detected in any of these cells, the most likely interpretation of the staining pattern in is that the *KRAS* mutation is carried by all *TP53* mutation-positive cells.

Discussion of Examples 11 to 17

[00298] Examples 11 to 17 document the establishment of a multiplex *in situ* assay that specifically targets point mutations on tumor tissue sections and on cytological preparations. Transcripts, synthesized by reverse transcription of mRNA *in situ*, are targeted with mutant- or wild-type specific padlock probes and amplified to a detectable level with RCA. The resulting wild-type and mutated products are thereafter labeled with fluorophores of different colors. This padlock probe-based assay demonstrates for the first time that mutation analysis for molecular cancer diagnostics can be performed directly on tumor tissue sections. A multiplexed *in situ* assay was developed and validated as a proof-of-concept for the activating point mutations in *KRAS* codon 12 and 13 that are associated with resistance to anti-EGFR therapy in colorectal cancer. The selectivity of the probes was first tested individually. There was a clear-cut distinction between the *KRAS* mutant and wild-type samples and the genotypes were easily determined by simple microscopic visualization of the corresponding fluorescent signals. For multiplex detection, a side-by-side comparison between single corresponding padlock pairs and a cocktail of all codon 12 and 13 *KRAS* padlock probes showed that the two approaches were similar in efficiency and specificity. The padlock strategy was developed on unfixed tissue preparations as fresh frozen tissue contains high quality DNA and RNA and serves as the golden standard for molecular studies. However, implementation of diagnostics on fresh frozen tissue requires substantial and expensive biobanking efforts. As an alternative, unfixed tumor cells were used on touch imprints from

the fresh cut tumor surface. The *KRAS* mutation status could thus be determined on the day of sample arrival and was concordant with our routine pyrosequencing assay.

[00299] FFPE tissue blocks are used globally in routine surgical pathology and can be preserved for years in tissue archives. However, crosslinking of biomolecules induced by formalin results in fragmentation of DNA and RNA. Nevertheless, the short length of the padlock probe, in combination with the requirement of dual recognition sites and ligation makes this assay ideal for fixed histopathology specimens. Using a protocol optimized for formalin-fixed tissues *in situ* detection in routine FFPE sections was achieved and prospective surgical cancer specimens with unknown *KRAS* status were successfully characterized. A promising prospect for this assay is that hundreds of FFPE cancer samples can be screened simultaneously in TMAs for presence of mutations. Thus, for biomarker discovery in retrospective patient cohorts with available TMAs, high-throughput screening for point mutations could be performed along with IHC for protein expression and FISH-analysis for chromosomal aberration. The *in situ* protocol can be adapted for automation as any conventional FISH-assay, facilitating implementation of the assay for routine use. Moreover, the fluorescence readout can be changed to a histochemical staining for brightfield imaging if desired.

[00300] Tumor heterogeneity is a complex concept. One aspect is the variable mixture of cancer cells with acquired somatic mutations and genetically normal stromal and inflammatory cells. A second aspect is the morphological, and possibly genetic, variation within the tumor compartment with regard to pre-neoplastic versus invasive components, high-grade versus low-grade areas, invasion front versus central tumor area, and variable differentiation patterns, e.g. sarcomatoid, glandular, squamous or neuroendocrine etc. A third aspect is that the expression of a mutated allele can be influenced by promoter and splicing mutations, epigenetic alteration, or gene copy number aberrations, e.g. amplifications, deletions and LOH, in different parts of the tumor. These may be challenging to analyze on a genomic level. The described *in situ* technique allows studies of all these challenging features of tumor heterogeneity. Heterozygous and homozygous expression of mutated and wild-type alleles can be appreciated in tumor cells and demonstrate one form of fundamental information about a particular tissue specimen that probably would have gone undetected with PCR-based techniques resulting in an average value of the extracted mixture of mutant tumor and wild-type cells. This assay shows increased expression of a mutated *KRAS* codon

61 allele along with tumor progression in a colon cancer sample. In a case of lung adenocarcinoma, the expression of an activating *EGFR* mutation was demonstrated to be different in areas with distinctive histological architecture. Moreover, the technique allows dissection of how multiple different mutations are distributed and associated across a tumor lesion, as illustrated by two lung cancer cases where mutated *TP53* alleles could be visualized together with activating mutations in *EGFR* and *KRAS* respectively. Thus, mutation analysis *in situ* can help to dissect processes such as cancer initiation, tumor progression and metastasis. For future studies an intriguing application will be studies of the emergence of resistance mutations in response to targeted therapy. One case with a double mutation in *EGFR* was presented where low expression of the resistance mutation was seen in parallel with expression of the mutation associated with treatment response, as might be expected in a patient with a de novo resistance mutation. Analysis of a follow-up sample after EGFR treatment could reveal a patient-specific response on a histological level regarding the expression of the two mutations.

[00301] Despite the fact that the 79 patient samples assayed in this study had been collected at different time points during the last two decades, as well as treated under various conditions, they all qualified as suitable tissue material for this presented method. Furthermore, specifically designed padlock probes were successfully applied for *in situ* detection of totally 14 different point mutations which give confidence that this mutation assay offers robustness and can easily be adapted for detection of other mutations on tissue material from various sources. In conclusion, the presented padlock probe and RCA technology is believed to be an important assay for studies of histologic-genotypic correlations in complex tumor tissues for diagnostic molecular pathology and translational cancer research.

Example 18:

Detection of Braf mutations

[00302] BRAF presents somatic mutations in different sort of tumors, predominantly in malignant melanoma, sporadic colorectal tumors showing mismatch repair defects in microsatellites (MSI), low-grade ovarian serous carcinoma and thyroid papillary cancer. 80% of these mutations correspond to the hotspot transversion mutation T1799A that causes the amino acidic substitution V600E.

Most common mutation is the V600E mutation (Substitution - Missense)

Target cDNA region (mutated base):

5′GCATATACATCTGACTGAAAGCTGTATGGATTTTTATCTTGCATTCTGATGACTT
CTGGTGCCATCCACAAAATGGATCCAGACAACTGTTCAAACTGAT
GGGACCCACTCCATCGAGATTTC**A**CTGTAGCTAGACCAAAATCACCTA-3′ (SEQ ID NO:100)

BRAF padlock probe target region (arms: 15 + 15 nt):

5′-CTCCATCGAGATTTC**A**CTGTAGCTAGACCA-3′ (SEQ ID NO:101)

Example 19:

In situ sequencing of mRNA in intact cells and tissues

[00303] To better understand the molecular and physiological processes of normal and disease tissues, gene expression profiling needs to be studied at cellular level. We describe a method for sequencing individual mRNA molecules directly in situ on cytological slides and tissue sections, providing sequence and spatial information at micrometer resolution. Our approach proves the principle of massively parallel sequencing of individual mRNA molecules in the preserved architecture of heterogeneous tissues.

[00304] Most organism-level functions are executed by the concerted action of several different cell types. The identity and function of each cell is defined by its gene expression program, which in turn is governed by its path of differentiation, and external stimulation from surrounding cells and extracellular matrix. Attempts to study organ function by measuring gene expression in extracts after tissue homogenization will be more or less flawed, since the average gene expression profile across the tissue cannot be used to deduce the molecular state of the different cell types in the tissue (Levsky, 2003). After decades of research on cell cultures we need to move our focus to understand the interplay between distinct cell types in complex organ tissues. To that end, new technological approaches are required.

[00305] Isolated individual cells can be profiled on RNA level after laser capture microdissection (Bonner, 1997) or FACS sorting (Dalerba, 2011), but these techniques are laborious and limited in terms of number of cells analyzed. Moreover, the results are difficult to map back to the histological context as neighbouring cells are destroyed by the dissection process. The application of next generation sequencing technology (NGS) to RNA sequencing had provided a much more comprehensive view of the RNA content of cells than any previous technique (Wang, 2009). However, the currently available technologies for RNA sequencing are based on purified nucleic acids removed from their natural context, limiting the possibility to connect the sequence with spatial information, which is important when analyzing tissues consisting of heterogeneous populations of different cell-types. To address this limitation, single cell sequencing of genomes (Navin, 2011) and transcriptomes (Tang, 2009) has been achieved by isolating individual cells. However, this is a laborious, time-consuming, and costly process, as well as low in throughput in terms of number of cells analyzed.

[00306] Here we show that NGS can be applied *in situ* for sequencing of single mRNA molecules in the unperturbed context of fixed cells and tissues, enabling studies of tissue-specific expression and differentiation, with potential impact on areas such as developmental biology, epigenetics, and tumor- and stem-cell research. Our method is based on padlock probing (Nilsson, 1994), gap-fill polymerization (Lizardi, 1998; Porreca 2007), rolling-circle amplification (RCA), and sequencing by ligation chemistry (SBL) (Shendure, 2005; Drmanac, 2010). The padlock probe and RCA technique has previously been used to detect and genotype individual mRNA molecules *in situ* in single cells and tissue (Larsson , 2010). Here, the RCA generates clonally amplified molecules that form sub- micron-sized clusters that can be used as sequencing substrates (Drmanac, 2010; Pihlak, 2008). The workflow is presented in **FIG. 10**. First, the mRNA is converted into cDNA *in situ* by reverse transcription. A padlock probe is then hybridized to the cDNA such that a gap is formed between the two ends of the probe. The gap, which is the target for sequencing, is thereafter filled by polymerization and ligation, allowing the padlock probe to form a DNA circle. Next, the target cDNA primes RCA of the DNA circle that contains the padlock probe and target sequences, generating molecular clones of up to a thousand copies of the target. Thus, in contrast to the other next generation platforms, where the sequencing substrates are introduced after amplification from purified nucleic acids, we generate an array of amplified DNA single molecules in their natural context.

The amplified DNA single molecules are then subjected to sequencing-by-ligation (Shendure, 2005; Drmanac, 2010). An anchor primer is hybridized right next to the target, followed by ligation of four interrogation nonamer oligonucleotides (including eight random positions) labeled with four different fluorescent dyes; one for each query nucleotide. The sequence is read by cycles of hybridization of anchor primer, ligation of nonamer oligonucleotides and imaging. The sequence is decoded by analyzing the fluorescence staining pattern for each RCA product through the different sequencing cycles. Visual examination of the patterns is possible but very tedious and subject to errors. We have therefore developed a fully automated decoding pipeline in the open source software CellProfiler calling ImageJ plugins for image alignment (Kamentsky, 2011; Thevenaz, 1998). A sequence is assigned to each RCA product and each sequence is related to an individual cell or region of interest. In order to reduce noise and enhance base-calling accuracy, we include a detection oligonucleotide that hybridizes to a sequence common to all probes, thus labeling all RCA products independent of their target sequence. We use this general stain to localize the RCA products and align signals from subsequent sequencing cycles (**Figs. 13, 14**).

[00307] To test the feasibility of our method, we first sequenced a stretch of four different nucleotides in the human β-actin transcript (*ACTB*). We successfully called bases for all four positions and with a base- calling accuracy of 98.6% (19 errors out of 1316 base calls) (**FIG. 12**). High correlation of signals from different imaging cycles indicates that the sequencing substrates do not move significantly between imaging cycles.. To evaluate potential loss of sequencing substrates during repeated cycles we sequenced four bases in *ACTB* followed by re-hybridization of the general detection probe (**FIG. 16**). Virtually no sequencing substrates were lost in this experiment, and the signal intensities measured under similar conditions did not decrease substantially, meaning that the sequencing substrates are firmly attached within the cellular matrix. However, a tendency for increased fluorescence background in the cells over sequencing cycles caused the weakest signals to be lost in the background. Next we targeted two four bases long sequences in human and mouse *ACTB* from a slide where human and mouse fibroblasts (BJhTERT and MEF respectively) had been grown in a mixed culture. The target sequences in the two species were identical except for a single nucleotide variation (SNV) at the second position from the sequencing primer. As summarized in **FIG. 11**, all the reads belonged to the two expected sequences,

indicating 100% sequencing accuracy, of which 97% of the sequences were mapped correctly to individual cells. This is in part due to errors in assigning signals to individual cells by the automated image analysis software. The sequencing accuracy can be improved by adding cytoplasmic staining after the sequencing cycles, thus improving cell segmentation (**FIG. 16**).

[00308] To prove the possibility of parallel sequencing of different transcripts in tissue sections, we sequenced two short sequence-tags of *ACTB* and *HER2* (also known as *ERBB2*) transcripts in a *HER2*- positive human fresh frozen breast cancer sample. **FIG. 12** and **FIG. 17** show the decoded sequences and their frequency in different parts of the tissue section. Both *ACTB* and *HER2* were successfully sequenced by our method and the sequence from the *HER2* transcript was detected in the cancer areas of the section. We generated 298 reads from *ACTB* and 27 reads from *HER2* in one ×40 image. We also observed 4 reads that did not match with the query sequences with one miscalled base per read, which corresponds to a base calling accuracy of 99.7%.

[00309] Finally, we used our technique to sequence 5 bases in codons 12 and 13 of the *KRAS* transcript, which represent hot-spots for activating mutations in many cancer types. The two codons are also routinely analyzed before anti-EGFR therapy in colon cancer since a mutation confers resistance to therapy (Karapetis, 2008). We co-cultured six different cell lines with different mutations in the KRAS gene, and applied in situ sequencing on the fixed cells. All mutation variants were detected among the obtained sequences (FIG. 18).

[00310] The read length of our method at the current state is short. This is due to the fact that we only attempted sequencing by ligation using a single primer. By using multiple primers, sequences of 5+5 bases can be achieved (Dramanac, 2010). We also foresee that by using other sequencing chemistry, i.e., sequencing by synthesis as used by Illumina, it will be possible to generate much longer reads. In the current study the sequencing and image acquisition was performed manually. However, instruments from Life (SOLiD 5500), Illumina (HiSeq and MySeq) and the open source Polonator instrument have the required fluidics and imaging components to implement these steps automatically. A second limitation of the current protocol is an incomplete analysis of the RNA content since not all mRNA molecules are converted to sequence reads, providing a relatively shallow view of the RNA content of the tissue. If greater depth is required, this approach can be complemented

with micro-dissection and deep sequencing, where the in situ sequencing will provide the landmarks to which the detailed sequence information can be mapped.

[00311] The presented in situ sequencing provides a unique tool for selective re-sequencing of regions of interest in the expressed part of the genome in fixed cells and tissue sections. It enables investigation of single molecules at single cell resolution in a large number of cells in parallel, impossible to achieve by any other method. It also enables sequencing of DNA tags in padlock probes, increasing multiplexity otherwise limited by the number of fluorescence spectra that can be resolved. The current padlock design allows for parallel detection of 1024 (45) transcripts, splice variants, or SNVs. In situ sequencing opens new possibilities to study complex biological events such as differential allele expression, gene splicing, and stem cell differentiation in wide populations of cells in their natural context. In conclusion, this is the first time sequencing is carried out *in situ*, and our method opens up new prospects for next generation sequencing technology and provides new opportunities for basic research and clinical diagnostics.

Methods

Cell culture and sample preparation

[00312] The cell lines BJhTERT, MEF and A-549 were cultured in DMEM (Gibco) without phenol red and L- glutamine supplemented with 10% FBS (Sigma), 2 mM L-glutamin (Sigma) and 1x PEST (Sigma). The cell lines ONCO DG-1, A-427, SW-480 and HCT-15 were cultured in RPMI 1640 (Sigma) without phenol red and L-glutamine supplemented with 10% FBS (Sigma), 2 mM L-glutamin (Sigma) and 1x PEST (Sigma). The cell line HUP-T3 was cultured in MEM with Earls salts (Gibco) supplemented with 10% FBS (Sigma), 2 mM L-glutamin (Sigma), 1% nonessential amino acids (Sigma), 1% sodium pyruvate (Sigma) and 1x PEST (Sigma). All the cell lines were incubated at 37 °C, 5% CO_2.

[00313] To prepare cell samples, confluent cell lines were treated with 0.25% (w/v) Trypsin-EDTA (Sigma) and resuspended in culturing medium. Resuspended cells were then dropped on Superfrost plus slides (Thermo) placed in a 150mm x 25mm petri dish (Corning) and culturing medium was added to the final volume of 25 ml. Three ml of resuspended cells were used to seed 5 slides. Cells were incubated in the same previous conditions 12-24 hours before fixation. Slides with co-cultured cell lines were prepared as previously described but

seeding equal amount of human BJhTERT and mouse MEF cell lines or human ONCO DG-1, A-427, SW-480, A-549, HCT-15 and HUP-T3 cell lines. Fixation was performed in paraformaldehyde (Sigma) 3% (w/v) in DEPC-treated PBS for 30 min at RT after removal of the culturing medium and two washes in PBS. After fixation slides were washed twice in DEPC- treated PBS and dehydrated in ethanol series 70%, 85% and 100% 5 min each. All the following reactions were performed in a volume of 50 µl in a 9 mm diameter Secure-Seal™ hybridization chamber (Invitrogen).

Tissue section

[00314] Frozen sections (4 µm) from fully anonymized human fresh frozen breast cancer tissues were obtained from the biobank at the Department of Pathology, Uppsala University Hospital, and are in accordance with the Swedish Biobank Legislation. Tissue samples were stored at -80 °C until fixation. Fixation was performed in paraformaldehyde (Sigma) 3% (w/v) in DEPC treated PBS with 0.05% Tween-20 (Sigma) (DEPC-PBS-T) for 45 min at RT followed by two washes in DEPC-PBS-T. The tissue was permeabilized in 2 mg/ml pepsin (Sigma) in 0.1 M HCl at 37 °C for 90 s. After permeabilization slides were washed twice in DEPC-treated PBS and dehydrated in ethanol series. All the following reactions were performed in a volume of 50 µl in a 9 mm diameter Secure-Seal™ hybridization chamber (Invitrogen).

In situ reversed transcription

[00315] Both cells and tissue samples were first rinsed with DEPC-PBS-T. For the cell samples, this was followed by incubation in 0.1 M HCl (in DEPC H_20) for 5 min and two washes with DEPC-PBS-T. For tissue, no further treatments were performed after the steps described above. Then the reversed transcription mix containing 1 µM of LNA modified cDNA primer (all oligonucleotides sequences are listed in **Table 8**), 20 U/µl of RevertAid H minus M-MuLV reverse transcriptase (Fermentas), 500 µM dNTPs (Fermentas), 0.2 µg/µl BSA (NEB) and 1 U/µl RiboLock RNase Inhibitor (Fermentas) in the M-MuLV reaction buffer, was added on the slides. The incubation was carried out for 3 h at 37 °C. Slides were washed twice with PBS-T, followed by a postfixation step in 3% (w/v) paraformaldehyde in DEPC-PBS for 10 min at room temperature (30 min for tissues). After postfixation, the samples were washed twice in DEPC-PBS-T.

Table 8 Oligonucleotide sequences. Padlock probes and LNA modified primers have been designed based on the sequences available at GenBank database, using the following accession number: NM_001101.3 (ACTB), NM_001005862.1 for (ERBB2), and NM_033360 (KRAS).

Fuction	Name	Sequence (5' to 3' and modifications)
cDNA primers	P-βhum	CT+GA+CC+CA+TG+CC+CA+CCATCACGCCC (SEQ ID NO: 114)
	P-βmus	CT+GA+CC+CA+TT+CC+CA+CCATCACACCC (SEQ ID NO: 115)
	P-HER2	G+AG+CT+GG+GT+GC+CT+CGCACAATCCGCAGCCT (SEQ ID NO: 116)
	P-KRAS	C+CT+CT+AT+TG+TT+GG+ATCATATTCGTC (SEQ ID NO: 117)
Padlock probes	PLP- βactin	CGGCTTCGCGGGCGACGATGCTCCTCTATGATTACTG ACTGCGTCTATTTAGTGGAGCCCTATCTTCTTTCAAC GGCTCCGGCATGTGCA (SEQ ID NO: 118)
	PLP- βactin-4nt	AGGCCGGCTTCGCGGGCGACGGCGACTATGATTACT GACTGCGTCTATTTAGTGGAGCCCTATCTTCTTTCAA CGGCTCCGGCATG (SEQ ID NO: 119)
	PLP- HER2	TGAGGCAGGTCCCACTGCAGATCGTCTTAATCACTA GTCGGAAGTACTACTCTCTTACGCTTACAACTAGACG TGCTCATCGCTCACAAC (SEQ ID NO: 120)
	PLP-KRAS	GTAGGCAAGAGTGCCTCCTCTATGATTACTGACCAA GATCCCTCAATGCTGCTGCTGTACTACGGTTCAAGTG TGGTAGTTGGAGCT (SEQ ID NO: 121)
Detection probes	DO_ACTB_Cy3	Cy3- TGCGTCTATTTAGTGGAGCC - 3' (SEQ ID NO: 122)
	DO_ACTB_U_Cy3	Cy3-UGCGUCUAUUUAGUGGAGCC-3'(SEQ ID NO: 123)
	DO_KRAS_U_Cy3	Cy3-CCUCAAUGCUGCUGCUGUACUAC-3'(SEQ ID NO: 124)
	DO_Her2_U_Cy3	Cy3-AGUCGGAAGUACUACUCUCU-3'(SEQ ID NO: 125)
Interrogation probes	Cy50NA	p-ANNNNNNNN-Cy5
	Cy30NG	p-GNNNNNNNN-Cy3
	TR0NC	p-CNNNNNNNN-Cy3.5
	FITC0NT	p-TNNNNNNNN-FITC
	Cy51NA	p-NANNNNNNN-Cy5
	Cy31NG	p-NGNNNNNNN-Cy3
	TR1NC	p-NCNNNNNNN-Cy3.5
	FITC1NT	p-NTNNNNNNN-FITC
	Cy552NA	p-NNANNNNNN-Cy5
	Cy32NG	p-NNGNNNNNN-Cy3
	2NCTR	p-NNCNNNNNN-Cy3.5
	FITC2NT	p-NNTNNNNNN-FITC
	Cy53NA	p-NNNANNNNN-Cy5
	Cy33NG	p-NNNGNNNNN-Cy3
	TR3NC	p-NNNCNNNNN-Cy3.5
	FITC3NT	p-NNNTNNNNN-FITC
	NAN_DO	p-NNNNANNNN-Cy5
	N4GN	p-NNNNGNNNN-Cy3
	TR4NC	p-NNNNCNNNN-Cy3.5
	FITC4NT	p-NNNNTNNNN-FITC

Padlock probing, gap-fill polymerization and ligation

[00316] All padlock probes were phosphorylated prior to use. A mixture containing 2 μM of padlock probes, 1× PNK buffer A (Fermentas), 1 mM ATP, 0.1 U/μl T4 Polynucleotide Kinase was incubated at 37 °C for 30 min and 60 °C for 20 min. The phosphorylated padlock probes can be stored at -20 °C until used. After reversed transcription, a mix that performs the degradation of RNA, hybridization of padlock probe, copy of target sequence for sequencing, and ligation to form a complete DNA circle was added. The mix contained 1× Ampligase buffer (20 mM Tris-HCl, pH 8.3, 25 mM KCl, 10 mM $MgCl_2$, 0.5 mM NAD and 0.01% Triton X-100), 100 nM of each padlock probe, 50 μM of dNTPs, 0.2 U/μl Stoffel Fragment (Applied Biosystems), 0.5 U/μl Ampligase, 0.4 U/μl RNase H (Fermentas), 1 U/μl RiboLock RNase Inhibitor, 50 mM KCl and 20% formamide. The incubation was carried out at 37 °C for 30 min and followed by 45°C for 45 min. Then the slide was washed with 1x DEPC-PBS-T twice.

RCA and RCA products detection

[00317] A RCA mix containing 1 U/μl phi29 polymerase (Fermentas), 1x phi29 polymerase buffer, 0.25 mM dNTPs, 0.2 μg/μl BSA and 5% Glycerol in DEPC H2O was added to the reaction chamber and incubated for 2 hours to overnight to carry out the RCA. After the incubation, the slide was washed in DEPC-PBS-T 3 times. Finally, 100 nM of each corresponding detection probe (Uracil contained detection probes were used in KRAS sequencing as well as tissue sequencing) in 2× SSC and 20% formamide was applied to the slide and incubated at 37 °C for 30 min. Excess amount of detection probes were then eliminated by 3 times washing with DEPC-PBS-T. After the washing, the secure seals were removed. The slides were then dehydrated through ethanol series. The slides were mounted in Vectashield mounting medium (Vector) that contains 100 ng/ml of DAPI for counterstaining the nuclei and analyzed using an AxioplanII epifluorescence microscope (Zeiss). Exposure time in all the experiments is list in **Table 9**.

Table 9 Exposure time used during image acquisition.

Samples	Reaction	Objective	DAPI (ms)	FITC (ms)	Cy3 (ms)	Cy3.5 (ms)	Cy5 (ms)
Human cell	Detection	×40	100-120	/	100	/	/
	Sequencing		100	300	157	127	2000
Human and Mouse co-cultured cells	Detection	×40	100-120	/	180	/	/
	Sequencing		100	275	130	115	2500
Breast cancer tissue	Detection	×40	800	/	25	/	/
	Sequencing		300-350	120-150	35	25	500-1000
KRAS co-cultured cells	Detection	×20	1000-1200	/	300	/	/
	Sequencing		300	500	150	130	2000

Sequencing-by-Ligation

[00318] Before the sequencing was performed, the detection probes were stripped off. The slides were first incubated through ethanol series to remove the mounting medium, and dried at room temperature. For detection probes without uracils, the samples were first washed with DEPC-PBS-T and then incubated with 65% formamide for 30 s 3 times, followed by washing with DEPC-PBS-T twice. Detection probes that contain uracils were first treated with UNG treating buffer (1x phi29 polymerase buffer (Fermentas), 0.2 μg/μl BSA, 0.02 U/μl UNG (Fermentas)) for 10 min and washed twice with DEPC-PBS-T prior to the formamide incubation.

[00319] A mix containing 500 nM of corresponding anchor primers in 2× SSC and 20% formamide were then added to the sample and incubate at RT for 30 min, followed by two brief washes with DEPC- PBS-T. A ligation mix containing each interrogation probes, 1× T4 ligase buffer (Fermentas), 1 mM ATP (Fermentas) and 0.1 U/μl of T4 ligase (Fermentas) was applied to the samples and incubated for

[00320] 30 min at RT. The unligated probes were washed away by three times 1 min incubation with DEPC- PBS-T. The slides were mounted in Vectashield mounting medium (Vector) that contains 100 ng/ml of DAPI for counterstaining the nuclei. The concentration of each interrogation probe was 100 nM for cell cultures and 500 nM for tissue section.

[00321] After imaging, the slides were then ready for the next sequencing cycle. The slides were then treated with UNG treatment buffer as described above and followed by repeating the hybridization, ligation and imaging processes. To sequence each base, the same procedures were applied. In the case for evaluation of loss of sequencing substrates (RCA products), cell cytoplasm was stained after the last sequencing cycle with Alexa

Fluor® 488 Phalloidin (Invitrogen) in PBS at a final concentration of 0.15 μM, incubating the cells 10 min at RT.

Image acquisition and analysis

[00322] All images were acquired using an AxioplanII epifluorescence microscope (Zeiss). In order to capture signals from RCPs located in slightly different focal planes, a series of images were captured at different focal depths. The stacks of images were thereafter merged to a single image using the Maximum Intensity Projection (MIP) in the Zeiss AxioVision software. The detailed image analysis process is described below.

[00323] Raw image data from each sequencing cycles of co-cultured mouse and human cells is shown in FIG. 13. Images were analyzed using the free and open source software CellProfiler (version 2.0 r10997). A dedicated CellProfiler pipeline was designed and combined with image alignment plug-ins written for ImageJ (version 1.43u) processing package. The complete executable and commented pipeline together with raw images and links to ImageJ plugins is available on the world wide web at cellprofiler.org/examples.shtml. The fully automated sequence decoding is performed as follows: First, images are cropped to a fixed size to remove edge effects from the MIP. Cell nuclei are defined by automated thresholding and separated from each other based on shape (FIG. 14a). Each cell's cytoplasm is defined by using the nucleus as a seed in a seeded watershed approach on an image constructed by merging all image channels from the first hybridization step and relying on sufficient cytoplasmic autofluorescence to distinguish it from empty background (FIG. 14b). In cases where background fluorescence is low and such approach fails, RCPs are assigned to the cell having the closest cell nucleus. Next, the image of the general stain from the first hybridization step is filtered by a TopHat filter to enhance RCPs. Touching RCPs are separated by watershed segmentation and each RCP is given a unique label (FIG. 14c). Finally, the fluorescence intensity from each of the signals representing A, C, T, and G is enhanced by TopHat filtering and extracted using the labeled RCPs as a template. This requires perfect alignment of images, achieved using Fiji function MultiStackReg and finding the optimal transformation between a merged image of all signals (A+C+T+G) and the general stain from the first hybridization step. This Java plugin is called directly from CellProfiler. All intensity information is saved to a csv-file and decoded using a script written in Matlab (also provided together with the CellProfiler pipeline). In short, for each RCP and hybridization step, the RCP is assigned the letter with the highest intensity. A

quality measure is also extracted, defined as the highest intensity divided by the sum of all intensities. A value close to 0.25 means poor quality (signal similar for all letters), while a value close to one means that the assigned letter is strong above a low background. Once all RCPs are detected an intensity threshold as well as a quality thresholds are set, and the frequency of each sequence is extracted for each hybridization step. Number of frequencies is also extracted on a per-cell basis and shown overlaid on the original image (FIG. 14d).

REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

Bonner, R. F. *et al. Science* **278**, 1481-1483, 1997.

Botling and Micke, *Methods Mol. Biol.,* 675:299-306, 2011.

Dalerba, P. *et al. Nat Biotechnol* **29**, 1120-1127, 2011.

Drmanac, R. *et al. Science* **327**, 78-81, 2010.

Kamentsky, L. *et al. Bioinformatics* **27**, 1179-1180, 2011.

Karapetis, C. S. *et al. New Engl J Med* **359**, 1757-1765, 2008.

Lagunavicius *et al.,* RNA, 15:765-771, 2009.

Larsson *et al., Nat. Methods*, 1:227-232, 2004.

Larsson, C., Grundberg, I., Soderberg, O. & Nilsson, M. *Nat. Methods* **7**, 395-397, ,2010.

Levsky, J. M. & Singer, R. H. *Trends Cell Biol* **13**, 4-6, ,2003.

Lizardi *et al.*, *Nat. Genet.*, 19:225-232, 1998.

Mitra and Church, *Nucleic Acids Res.*,27(24), 1999.

Navin, N. *et al. Nature* **472**, 90-94, 2011.

Nilsson *et al.*, *Nat. Biotechnol.*, 18:791-793, 2000.

Nilsson *et al.*, *Nucleic Acid Res.*, 29:578-581, 2001.

Nilsson *et al., Science*, 265:2085-2088, 1994.

Owczarzy *et al., Biochemistry*, 47:5336-5353, 2008.

Pihlak, A. *et al. Nat. Biotechnol.***26**, 676-684, ,2008.

PCT Appln. WO 99/49079

Pena *et al.*, *Nat. Methods*, 6(2):139-141, 2009.

Porreca, G. J. *et al. Nat Methods* **4**, 931-936, 2007.

Sambrook *et al.*, Molecular Cloning--A Laboratory Manual, Cold Spring Harbor Laboratory, NY, 1989.

Shendure, J. *et al. Science* **309**, 1728-1732, 2005.

Stougaard *et al.*, *BMC Biotech.*, 7:69, 2007.

Sundstrom *et al., BMC Cancer*, 10:660, 2010.

Tang, F. *et al. Nat Methods* **6**, 377-382, 2009.

Thevenaz, P., Ruttimann, U. E. & Unser, M. *Ieee T Image Process* **7**, 27-41, 1998.

Wang, Z., Gerstein, M. & Snyder, M. *Nat Rev Genet* **10**, 57-63, 2009.

Wetmur, *Critical Rev. Biochem. Mol. Biol.*, 26(3/4):227-259, 1991.

Yi *et al., Nucleic Acid Res.*, 29:4818-4825, 2001.

CLAIMS

What is claimed:

1. A method for localized *in situ* detection of at least one target RNA in a sample of one or more cells, comprising:

a) hybridizing the target RNA with a complementary nucleic acid that comprises a region complementary to the target RNA;

b) digesting the RNA hybridized to the complementary nucleic acid;

c) contacting the sample with one or more padlock probes, wherein the padlock probe(s) comprise terminal regions complementary to the cDNA and hybridizing the padlock probe to the cDNA at the complementary terminal regions;

d) joining the ends of the padlock probe(s); and,

e) subjecting the circularized padlock probe(s) to rolling circle amplification (RCA).

2. The method of claim 1, further comprising detecting the rolling circle amplification product(s).

3. The method of claim 1, further comprising sequencing all or part of one or more rolling circle amplification product(s).

4. The method of claim 1 wherein the RNA is mRNA.

5. The method of any of claims 1 to 4, further comprising generating the complementary nucleic acid.

6. The method of claim 5, wherein generating the complementary nucleic acid comprising incubating the target RNA with a reverse transcriptase and a reverse transcription primer that is complementary to the target RNA to generate a cDNA.

7. The method of any of claims 1 to 6, wherein the sample is on a solid support.

8. The method of claim 7, wherein the solid support is a slide.

9. The method of claim 8, wherein the slide has a cover.

10. The method of claim 5, wherein the cells are stained.

11. The method of claim 10, wherein the cells are stained with hematoxylin and eosin.

12. The method of any of claims 6 – 11, wherein the reverse transcription primer is modified so as to be capable of immobilization in the cells.

13. The method of claim 12, wherein the reverse transcription primer has a functional moiety capable of binding to or reacting with a cell or cellular component or an affinity binding group capable of binding to a cell or cellular component.

14. The method of any one of claims 12 to 13, wherein the reverse transcription primer comprises at least one nucleotide modified with biotin, an amine group, a lower alkylamine group, an acetyl group, DMTO, fluoroscein, a thiol group, or acridine.

15. The method of claim 14, wherein the reverse transcription primer comprises one or more locked nucleic acid residues.

16. The method of claim 15, wherein the reverse transcription primer comprises 2 or more locked nucleic acids separated by 1 or more natural or synthetic nucleotides in the primer sequence.

17. The method of any of claims 1 to 15, further comprising adding a ribonuclease to digest RNA hybridized to the complementary nucleic acid.

18. The method of claim 17, wherein the ribonuclease is RNase H.

19. The method of any of claims 1 to 18, further comprising incubating the padlock probe with a ligase to join the ends.

20. The method of any one of claims 1 to 18, wherein the rolling circle amplification uses a DNA polymerase having 3'-5' exonuclease activity wherein if necessary the exonuclease activity digests the complementary nucleic acid to generate a free 3' end which acts as a primer for the RCA.

21. The method of claim 20, wherein the DNA polymerase is a Φ29 polymerase.

22. The method of any one of claims 1 to 21, further comprising contacting the sample with an exonuclease to digest the complementary nucleic acid to generate a free 3' end which acts as a primer for the RCA.

23. The method of any of claims 1 to 22, further comprising incubating the amplification product with a detection probe under conditions to allow hybridization between the product and the probe.

24. The method of claim 23, wherein the detection probe has one or more fluorescent labels, enzymatic labels, epitope labels, colorimetric labels, chromogenic labels, radioactive labels, luminescent labels, magnetic labels, or electron-density labels.

25. The method of any of claims 24, further comprising incubating the detection probe with one or more polypeptides that binds the one or more labels.

26. The method of claim 25, wherein at least one or more polypeptides is an antibody.

27. The method of claim 26, further comprising incubating the one or more polypeptides that binds the one or more labels with a secondary polypeptide that binds the label-binding polypeptide.

28. The method of claim 27, wherein the secondary polypeptide is an antibody.

29. The method of claim 26 or 27, wherein the label-binding polypeptide or the secondary polypeptide comprises a detectable label.

30. The method of any of claims 24 to 28, wherein detecting the rolling circle amplification product(s) comprises assaying for the label(s) on the detection probe.

31. The method of any of claims 24 to 30, further comprising incubating the rolling circle amplification product(s) and label(s) on the probe with an enzyme substrate to detect the rolling circle amplification product(s).

32. The method of any of claims 23 to 31, further comprising incubating the amplification product with a detection probe under conditions to allow hybridization between the product and the probe.

33. The method of any of claims 1 to 32, wherein in the contacting step, the sample is contacted with at least a first and a second padlock probe, wherein the first padlock probe comprises terminal regions complementary to immediately adjacent regions on the complementary nucleic acid, and wherein the second padlock probe comprises terminal regions that differ from the terminal regions of the first padlock probe only by a single nucleotide at the 5' or 3' terminus of the second padlock probe.

34. The method of claim 33, wherein the first padlock probe is configured to hybridize to a complementary nucleic acid that is complementary to a wild-type mRNA, and the second padlock probe is configured to hybridize to a complementary nucleic acid that is complementary to a single nucleotide variant of the mRNA.

35. The method of claim 33 or 34, wherein the first padlock probe comprises a first detection probe binding region, and the second padlock probe comprises a second detection probe binding region.

36. The method of claim 35, further comprising contacting the sample with at least a first labeled detection probe comprising a sequence identical to the first detection probe binding region of the first padlock probe, and a second labeled detection probe comprising a sequence identical to the second detection probe binding region of the first padlock probe, and hybridizing the first and second labeled detection probes to the rolling circle amplification products.

37. The method of claim 36, further comprising detecting a signal from the first labeled detection probe hybridized to the rolling circle amplification product, or detecting a signal from the second labeled detection probe hybridized to the rolling circle amplification product, or detecting a signal from both the first and second labeled detection probes hybridized to the rolling circle amplification products, wherein the detection of the first labeled detection probe indicates the presence of the wild-type mRNA in the cell and the detection of the second labeled detection probe indicates the presence of the variant mRNA in the cell.

38. The method of claim 36 or 37, wherein the labeled detection probe comprises different fluorescent labels, enzymatic labels, colorimetric labels, chromogenic labels, radioactive labels, luminescent labels, magnetic labels, or electron-density labels.

39. The method of any one of claims 1 to 38, wherein multiple different RNAs are detected using multiple different padlock probes.

40. The method of any of claims 1 to 34, wherein subjecting the circularized padlock probe(s) to rolling circle amplification comprises adding labeled nucleotides to generate labeled, amplified padlock(s).

41. The method of claim 40, where the nucleotides are labeled with fluorescent labels, enzymatic labels, colorimetric labels, chromogenic labels, radioactive labels, luminescent labels, magnetic labels, or electron-density labels.

42. The method of claim 40, wherein at least two different padlock probes are used.

43. The method of claim 42, wherein the at least two different padlock probes have different backbone sequences.

44. The method of claim 43, wherein a difference in backbone sequence comprises a difference in nucleotide content.

45. The method of claim 44, wherein a difference in backbone sequence comprises a difference in nucleotide content by at least 2x.

46. The method of claim 45, wherein the difference in nucleotide content comprises a difference in the number of G nucleotides in the backbone sequences.

47. The method of claim 45, wherein the difference in nucleotide content comprises a difference in the number of C nucleotides in the backbone sequences.

48. The method of claim 45, wherein the difference in nucleotide content comprises a difference in the number of A nucleotides in the backbone sequences.

49. The method of claim 45, wherein the difference in nucleotide content comprises a difference in the number of T nucleotides in the backbone sequences.

50. The method of any of claims 40 to 44, wherein at least two differentially labeled nucleotides are added.

51. The method of claim 50, wherein at least two different padlock probes are used and the amplified padlock probes are differentially labeled.

52. The method of claim 51, wherein the amplified padlock probes are differentially labeled with different fluorescent labels, enzymatic labels, colorimetric labels, chromogenic labels, radioactive labels, luminescent labels, magnetic labels, or electron-density labels.

53. The method of any of claims 1 to 40, wherein rolling circle amplification product(s) are detected by sequentially adding at least two probes.

54. The method of claim 53, wherein each probe is detected prior to the addition of a next probe.

55. The method of claim 54, further comprising eliminating what is detected prior to the addition of the next probe.

56. The method of claim 55, wherein eliminating what is detected comprises employing photobleaching.

57. The method of any of claims 1-56, wherein rolling circle amplification product(s) are detected with one or more detection probes that comprises one or more branches having one or more labeling moieties on each branch.

58. The method of claim 57, wherein the one or more labeling moieties are fluorescent, enzymatic, epitopic, chromogenic, radioactive, luminescent, magnetic, and/or electron-dense.

59. The method of any of claims 1 to 38, wherein RNA is detected in a single cell.

60. The method of claims 1 to 59, wherein the sample comprises a fixed tissue section, touch imprint samples, a formalin-fixed paraffin-embedded tissue section or a cytological preparation comprising one or more cells.

61. The method of any of claims 1 to 59, wherein the sample comprises a fresh frozen tissue.

62. The method of claim 60, wherein the sample is a formalin-fixed paraffin-embedded tissue section.

63. The method of any one of claims 1 to 62, wherein the sample is derived from a tissue or organ of the body, or from a bodily fluid.

64. The method of claim 63, wherein the sample is a colon, lung, pancreas, prostate, skin, thyroid, liver, ovary, endometrium, kidney, brain, testis, lymphatic fluid, blood, plasma, urinary bladder, or breast sample.

65. The method of any one of claims 1 to 64, wherein the sample is suspected to comprise an RNA found in a cancer cell.

66. The method of claim 65, wherein the cancer is colorectal cancer, lung cancer, pancreas cancer, prostate cancer, skin cancer, thyroid cancer, liver cancer, ovary cancer, endometrium cancer, kidney cancer, cancer of the brain, testis cancer, acute non lymphocytic leukemia, myelodysplasia, urinary bladder cancer, head and neck cancer or breast cancer, or any early development stages thereof.

67. The method of claim any of claims 1 to 66, wherein the RNA is at least one mRNA selected from the group comprising or consisting of mRNAs that code for *KRAS, HER2*, *cMyc*, *TERT*, *APC, Braf, PTEN, PI3K, EGFR* and β-actin.

68. The method of claim 67, wherein the RNA is an mRNA that codes for *KRAS*.

69. The method of claim 68, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant *KRAS* mRNA sequences selected from the group consisting of 12AGT, 12CGT, 12TGT, 12GAT, 12GCT, 12GTT, and 13GAC, mutants of *KRAS* codon 61, mutants of *KRAS* codon 146, and mutants of the 3' untranslated region of *KRAS*.

70. The method of claim 68 or 69, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more of the mutant *KRAS* mRNA sequences.

71. The method of claim 68, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more wild-type *KRAS* mRNA sequences selected from the group consisting of 12GGT and 13GGC, wild-type sequences of *KRAS* codon 61, *KRAS* codon 146, and of the 3' untranslated region of *KRAS* .

72. The method of claim 71, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more of the wild-type *KRAS* mRNA sequences.

73. The method of any one of claims 68 to 72, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant *KRAS* mRNA sequences selected from the group consisting of 12AGT, 12CGT, 12TGT, 12GAT, 12GCT, 12GTT, and 13GAC mutants of *KRAS* codon 61, mutants of *KRAS* codon 146, and mutants of the 3' untranslated region of *KRAS;* and to one or more wild-type *KRAS* mRNA sequences selected from the group consisting of 12GGT and 13GGC, wild-type sequences of *KRAS* codon 61, *KRAS* codon 146, and of the 3' untranslated region of *KRAS.*

74. The method of claim 73, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more of the mutant and wild-type *KRAS* mRNA sequences.

75. The method of claim 67, wherein the RNA is an mRNA that codes for EGFR.

76. The method of claim 75, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more variant *EGFR* mRNA sequences.

77. The method of claim 76, wherein one or more of the variant *EGFR* mRNA sequences comprises an insertion mutation or a deletion mutation.

78. The method of claim 67, wherein the RNA is an mRNA that codes for *Braf, APC, PTEN or PI3K.*

79. The method of claim 78, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

80. The method of claim 79, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more mutant *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

81. The method of claim 78, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more wild-type *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

82. The method of claim 81, comprising detecting the presence or absence of a rolling circle amplification product corresponding to the wild-type *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

83. The method of any one of claims 78 to 82, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant *Braf, PTEN* and/or *PI3K* mRNA sequences, and to one or more wild-type *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

84. The method of claim 83, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more of the mutant and wild-type *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

85. The method of claim 67, wherein the RNA is mRNA that codes for *Braf* and for *KRAS,* or mRNA that codes for *KRAS* and *APC,* or mRNA that codes for *KRAS* and *PTEN*, or mRNA that codes for *KRAS* and *PI3K*.

86. The method of claim 85 wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to wild-type *KRAS* and *Braf* mRNA sequences; or corresponding to wild-type *KRAS* and *APC* mRNA sequences*;* or corresponding to wild-type *KRAS* and *PTEN* mRNA sequences; or corresponding to wild-type *KRAS* and *PI3K* mRNA sequences.

87. The method of claim 86, wherein the RNA is at least one mRNA is encoded by a gene listed in Table B.

88. The method of claim 87, wherein a point mutation has been associated with cancer in the gene.

89. The method of any of claims 1 to 23 or 59 to 86, wherein detecting the rolling circle amplification product(s) comprises sequencing at least a portion of the rolling circle amplification product(s).

90. The method of claim 89, wherein the sample is on a slide and is removed from the slide prior to sequencing.

91. A method for *in situ* localization of a nucleic acid sequence in a cell in a biological sample on a slide comprising:

a) incubating an immobilized biological sample on solid support with reverse trancriptase and a ribonuclease-resistant primer under conditions to generate a nucleic acid molecule that contains the nucleic acid sequence and that hybridizes to a complementary RNA molecule in the cell to form an RNA-DNA hybrid;

b) incubating the RNA-DNA hybrid with a ribonuclease under conditions to digest RNA in the RNA-DNA hybrid;

c) incubating the digested RNA-DNA hybrid under conditions to hybridize a complementing padlock probe to the DNA portion of the digested RNA-DNA hybrid, wherein the padlock probe comprises the nucleic acid sequence and has two terminal ends that are complementary to different but immediately adjacent regions of the DNA,

d) incubating the padlock probe hybridized to the DNA portion of the RNA-DNA hybrid with a ligase under conditions to ligate the terminal ends of the padlock probe;

e) incubating the ligated padlock probe with a polymerase and nucleotides under conditions to create a primer from the DNA that is used to replicate the padlock probe and generate a nucleic acid with multiple copies of the replicated padlock probe; and,

f) incubating the nucleic acid with one or more complementing nucleic acid probes to detect the presence or absence of the specific sequence.

92. The method of claim 91, wherein the one or more complementing nucleic acid probes are labeled with more than one labeling moiety.

93. The method of claim 92, wherein one or more complementing nucleic acid probes is branched.

94. The method of claim 93, wherein the branched probe comprises a labeling moiety on each probe branch.

95. The method of claim 91, wherein multiple complementing nucleic acid probes are incubated with the nucleic acid.

96. The method of claim 95, wherein the multiple probes are fluorescently labeled.

97. The method of any of claims 95 to 96, wherein the multiple probes are added sequentially to the nucleic acid.

98. The method of claim 97, wherein hybridization of an added probe is detected prior to addition of other probes.

99. The method of claim 98, further comprising photobleaching a fluorescently labeled probe after detecting hybridization of an added probe.

100. The method of claim 99, further comprising detecting hybridization of a probe added after photobleaching.

101. The method of any of claims 91 to 100, wherein the sample is a formalin-fixed paraffin-embedded tissue section.

102. The method of any of claims 91 to 101, wherein in step (c) the terminal regions of the padlock probe hybridize to non-contiguous regions of the cDNA such that there is a gap between the terminal regions.

103. The method of claim 102, wherein the gap is a 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 nucleotides gap, or a 11 to 60 nucleotides gap, or a 3 to 40 nucleotides gap.

104. The method of claim 102 or 103, wherein the gap between the terminal regions is filled by a gap oligonucleotide or by extending the 3' end of the padlock probe.

105. The method of any of claims 91 to 104, wherein biological sample is suspected of having cancer cells.

106. The method of any one of claims 1 to 105, wherein the sample may comprises a tissue section, with the proviso that the tissue section is not or does not include a preparation including seeding on Superfrost Plus slides.

107. The method of any one of claims 1 to 106, wherein in step (c) the terminal regions of the padlock probe(s) hybridize to immediately adjacent regions on the cDNA.

108. A method for determining at least one nucleic acid sequence in a target nucleic acid in a biological sample comprising:

a) amplifying the nucleic acid sequence or its complement in the sample using one or more padlock probes, wherein the padlock probe(s) comprise terminal regions complementary to the nucleic acid sequence in the target nucleic acid;

b) determining the nucleic acid sequence in the amplified padlock probe.

109. The method of claim 108, wherein the sample is on a slide.

110. The method of claim 108 and 109, wherein the sample is a tissue section.

111. The method of claim 110, wherein the tissue section is or was frozen, or fixed.

112. The method of claim 119, wherein the tissue section is frozen.

113. The method of claim 111, wherein the tissue section is fixed.

114. The method of claim 113, wherein the fixed tissue section is or was formalin-fixed and embedded in paraffin.

115. The method of any of claims 108 to 114, wherein amplifying the nucleic acid sequence comprises generating a complement of the nucleic acid sequence.

116. The method of claim 115, wherein generating a complement of the nucleic acid sequence comprises generating a cDNA complementary to an RNA in the sample.

117. The method of any of claims 108, wherein amplifying the nucleic acid sequence comprises hybridizing the nucleic acid sequence to one or more padlock probes, wherein the padlock probe(s) comprise one or two terminal regions complementary to the nucleic acid sequence;

ligating the ends of the padlock probe(s);

replicating the nucleic acid sequence using polymerase and the ligated padlock probe(s) as a primer.

118. The method of claim 116, wherein amplifying the nucleic acid sequence comprises

digesting the RNA complementary to the cDNA;

hybridizing the cDNA to one or more padlock probes, wherein the padlock probe(s) comprise terminal regions complementary to the cDNA;

ligating the ends of the padlock probe(s);

replicating the circularized padlock probe(s) using polymerase and rolling circle amplification (RCA); and,

determining the nucleic acid sequence in the replicated padlock probe(s).

119. The method of any of claims 115 to 118, wherein determining the nucleic acid sequence comprises performing mass spectroscopy.

120. The method of any of claims 115 to 118, wherein determining the nucleic acid sequence comprises incubating the replicated padlock probe(s) with one or more labeled terminal nucleotides.

121. The method of any of claims 115 to 118, wherein determining the nucleic acid sequence comprises high throughput sequencing.

122. The method of any of claims 119 to 121, wherein multiple padlock probes are used.

123. The method of any of claims 109 to 122, wherein the sample is on a slide and the sample is removed from the slide prior to determining the nucleic acid sequence.

124. A method for determining in situ a nucleic acid sequence in a tissue section comprising:

generating a cDNA complementary to an RNA containing the nucleic acid sequence in the tissue section;

incubating the cDNA with a ribonuclease to the sample to digest the RNA;

hybridizing one or more padlock probes to the cDNA, wherein the padlock probe(s) comprise one or two terminal regions having the nucleic acid sequence;

incubating the hybridized padlock probes and cDNA with ligase under conditions to ligate the ends of the padlock probe(s);

incubating the replicating the padlock probe(s) using a polymerase to create an amplified product;

sequencing the complement of the nucleic acid sequence in the amplified product to determine the nucleic acid sequence and/or its complement.

125. The method of claim 124, wherein the sequencing comprises performing mass spectroscopy.

126. A method for detecting *in vitro* a genetic mutation in a gene of a cell comprising:

a) incubating the cell with a reverse transcriptase and a primer to hybridize the primer to RNA from the gene to generate a cDNA of all or part of the gene;

b) incubating the cell with a ribonuclease under conditions to digest the RNA;

c) incubating the cDNA with at least one padlock probe under conditions to hybridize the padlock probe to the cDNA, wherein the padlock probe comprises two terminal ends that are complementary to different but adjacent regions of the cDNA;

d) incubating the cDNA and at least one padlock probe with a ligase under conditions to join terminal ends of the padlock probe;

e) incubating the cDNA and the at least one ligated padlock probe with a polymerase and labeled nucleotides under conditions to amplify the at least one padlock probe and generate labeled, amplified padlock probes;

f) assaying for one or more labeled nucleotides in the amplified probes to detect the genetic mutation.

127. The method of claim 126, wherein the labeled nucleotides are chromogenic.

128. The method of claim 126, wherein the labeled nucleotides are fluorescent.

129. The method of claim 126, wherein the labeled nucleotides are colorimetric.

130. The method of any of claims 126 to 129, wherein multiple different padlock probes are incubated with the cDNA.

131. The method of claim 126 or 130, wherein the nucleotides incubated with the ligated padlock probe(s) and polymerase are differentially labeled by nucleotide.

132. The method of claim 131, wherein multiple different padlock probes are incubated with the cDNA.

133. The method of claim 132, wherein at least two of the multiple different padlock probes have different backbone sequences.

134. The method of claim 133, wherein a difference in backbone sequence comprises a difference in nucleotide content by at least 2x.

135. The method of claim 134, wherein the difference in nucleotide content comprises a difference in the number of G nucleotides in the backbone sequences.

136. The method of claim 134, wherein the difference in nucleotide content comprises a difference in the number of C nucleotides in the backbone sequences.

137. The method of claim 134, wherein the difference in nucleotide content comprises a difference in the number of A nucleotides in the backbone sequences.

138. The method of claim 134, wherein the difference in nucleotide content comprises a difference in the number of T nucleotides in the backbone sequences.

128. A method for evaluating a nucleic acid sequence in a tissue section comprising:

hybridizing a cDNA complementary to an RNA containing the nucleic acid sequence in the tissue section;

digesting all or part of the RNA;

hybridizing one or more padlock probes to the cDNA, wherein the padlock probe(s) comprise one or two terminal regions having the nucleic acid sequence;

incubating the hybridized padlock probes and cDNA with ligase under conditions to ligate the ends of the padlock probe(s);

incubating the replicating the padlock probe(s) using a polymerase to create an amplified rolling circle amplification product; and,

sequencing the amplified rolling circle amplification product.

139. The method of claim 128 wherein the RNA is mRNA.

140. The method of any of claims 128 to 139, further comprising generating the complementary nucleic acid.

141. The method of claim 140, wherein generating the complementary nucleic acid comprising incubating the target RNA with a reverse transcriptase and a reverse transcription primer that is complementary to the target RNA.

142. The method of any of claims 128 to 141, wherein the sample is on a solid support.

143. The method of claim 142, wherein the solid support is a slide.

144. The method of claim 143, wherein the slide has a cover.

145. The method of claim 140, wherein the cells are stained.

146. The method of claim 145, wherein the cells are stained with hematoxylin and eosin.

147. The method of any of claims 142-146, wherein the tissue sample is removed from the solid support prior to sequencing.

148. The method of any of claims 141 – 146, wherein the reverse transcription primer is modified so as to be capable of immobilization in the cells.

149. The method of claim 148, wherein the reverse transcription primer has a functional moiety capable of binding to or reacting with a cell or cellular component or an affinity binding group capable of binding to a cell or cellular component.

150. The method of any one of claims 148 to 149, wherein the reverse transcription primer comprises at least one nucleotide modified with biotin, an amine group, a lower alkylamine group, an acetyl group, DMTO, fluoroscein, a thiol group, or acridine. .

151. The method of claim 150, wherein the reverse transcription primer comprises one or more locked nucleic acid residues.

152. The method of claim 151, wherein the reverse transcription primer comprises 2 or more locked nucleic acids separated by 1 or more natural or synthetic nucleotides in the primer sequence.

153. The method of any of claims 128 to 152, further comprising adding a ribonuclease to digest RNA hybridized to the cDNA.

154. The method of claim 153, wherein the ribonuclease is RNase H.

155. The method of any of claims 128 to 154, further comprising incubating the padlock probe with a ligase to join the ends.

156. The method of any one of claims 128 to 154, wherein the rolling circle amplification uses a DNA polymerase having 3'-5' exonuclease activity wherein if necessary the exonuclease activity digests the cDNA to generate a free 3' end which acts as a primer for the RCA.

157. The method of claim 156, wherein the DNA polymerase is a Φ29 polymerase.

158. The method of any of claims 128 to 157, further comprising contacting the sample with an exonuclease to digest the cDNA to generate a free 3' end which acts as a primer for the RCA.

159. The method of any of claims 128 to 158, wherein sequencing involves one or more chain terminating nucleotides.

160. The method of any of claims 128 to 160, wherein sequencing comprises mass spectroscopy.

161. The method of any of claims 128 to 160, wherein in the contacting step, the sample is contacted with at least a first and a second padlock probe, wherein the first padlock probe comprises terminal regions complementary to immediately adjacent regions on the cDNA, and wherein the second padlock probe comprises terminal regions that differ from the terminal regions of the first padlock probe only by a single nucleotide at the 5' or 3' terminus of the second padlock probe.

162. The method of claim 161, wherein the first padlock probe is configured to hybridize to a cDNA complementary to a wild-type mRNA, and the second padlock probe is configured to hybridize to a cDNA complementary to a single nucleotide variant of the mRNA.

163. The method of claim 161 or 162, wherein the first padlock probe comprises a first detection probe binding region, and the second padlock probe comprises a second detection probe binding region.

164. The method of claim 163, further comprising contacting the sample with at least a first labeled detection probe comprising a sequence identical to the first detection probe binding region of the first padlock probe, and a second labeled detection probe comprising a sequence identical to the second detection probe binding region of the first padlock probe, and hybridizing the first and second labeled detection probes to the rolling circle amplification products.

165. The method of claim 164, further comprising detecting a signal from the first labeled detection probe hybridized to the rolling circle amplification product, or detecting a signal from the second labeled detection probe hybridized to the rolling circle amplification product, or detecting a signal from both the first and second labeled detection probes hybridized to the rolling circle amplification products, wherein the detection of the first labeled detection probe indicates the presence of the wild-type mRNA in the cell and the detection of the second labeled detection probe indicates the presence of the variant mRNA in the cell.

166. The method of claim 164 or 165, wherein the labeled detection probe comprises different fluorescent labels, enzymatic labels, chromogenic labels, radioactive labels, luminescent labels, magnetic labels, or electron-density labels.

167. The method of any one of claims 128 to 166, wherein multiple different RNAs are detected using multiple different padlock probes.

168. The method of any of claims 128 to 162, wherein subjecting the circularized padlock probe(s) to rolling circle amplification comprises adding labeled nucleotides to generate labeled, amplified padlock(s).

169. The method of claim 168, wherein at least two different padlock probes are used.

170. The method of any of claims 128 to 169, wherein RNA is detected in a single cell.

171. The method of claims 128 to 170, wherein the sample comprises a fixed tissue section, touch imprint samples, a formalin-fixed paraffin-embedded tissue section or a cytological preparation comprising one or more cells.

172. The method of any of claims 128 to 170, wherein the sample comprises a fresh frozen tissue.

173. The method of claim 171, wherein the sample is a formalin-fixed paraffin-embedded tissue section.

174. The method of any one of claims 128 to 173, wherein the sample is derived from a tissue or organ of the body, or from a bodily fluid.

175. The method of claim 174, wherein the sample is a colon, lung, pancreas, prostate, skin, thyroid, liver, ovary, endometrium, kidney, brain, testis, lymphatic fluid, blood, plasma, urinary bladder, or breast sample.

176. The method of any one of claims 128 to 175, wherein the sample is suspected to comprise an RNA found in a cancer cell.

177. The method of claim 176, wherein the cancer is colorectal cancer, lung cancer, pancreas cancer, prostate cancer, skin cancer, thyroid cancer, liver cancer, ovary cancer, endometrium cancer, kidney cancer, cancer of the brain, testis cancer, acute non lymphocytic leukemia, myelodysplasia, urinary bladder cancer, head and neck cancer or breast cancer, or any early development stages thereof.

178. The method of claim any of claims 128 to 177, wherein the RNA is at least one mRNA selected from the group comprising or consisting of mRNAs that code for *KRAS, HER2*, *cMyc*, *TERT*, *APC, Braf, PTEN, PI3K, EGFR* and β-actin.

179. The method of claim 178, wherein the RNA is an mRNA that codes for *KRAS*.

180. The method of claim 179, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant *KRAS* mRNA sequences selected from the group consisting of 12AGT, 12CGT, 12TGT, 12GAT, 12GCT, 12GTT, and 13GAC, mutants

of *KRAS* codon 61, mutants of *KRAS* codon 146, and mutants of the 3' untranslated region of *KRAS*.

181. The method of claim 179 or 180, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more of the mutant *KRAS* mRNA sequences.

182. The method of claim 179, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more wild-type *KRAS* mRNA sequences selected from the group consisting of 12GGT and 13GGC, wild-type sequences of *KRAS* codon 61, *KRAS* codon 146, and of the 3' untranslated region of *KRAS* .

183. The method of claim 182, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more of the wild-type *KRAS* mRNA sequences.

184. The method of any one of claims 179 to 183, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant *KRAS* mRNA sequences selected from the group consisting of 12AGT, 12CGT, 12TGT, 12GAT, 12GCT, 12GTT, and 13GAC mutants of *KRAS* codon 61, mutants of *KRAS* codon 146, and mutants of the 3' untranslated region of *KRAS;* and to one or more wild-type *KRAS* mRNA sequences selected from the group consisting of 12GGT and 13GGC, wild-type sequences of *KRAS* codon 61, *KRAS* codon 146, and of the 3' untranslated region of *KRAS*.

185. The method of claim 184, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more of the mutant and wild-type *KRAS* mRNA sequences.

186. The method of claim 178, wherein the RNA is an mRNA that codes for EGFR.

187. The method of claim 186, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more variant *EGFR* mRNA sequences.

188. The method of claim 187, wherein one or more of the variant *EGFR* mRNA sequences comprises an insertion mutation or a deletion mutation.

189. The method of claim 178, wherein the RNA is an mRNA that codes for *Braf, APC, PTEN or PI3K*.

190. The method of claim 189, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

191. The method of claim 190, comprising detecting the presence or absence of a rolling circle amplification product corresponding to one or more mutant *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

192. The method of claim 189, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more wild-type *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

193. The method of claim 192, comprising detecting the presence or absence of a rolling circle amplification product corresponding to the wild-type *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

194. The method of any one of claims 189 to 193, wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to one or more mutant *Braf, PTEN* and/or *PI3K* mRNA sequences, and to one or more wild-type *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

195. The method of claim 194, comprising sequencing to identify the presence or absence of a rolling circle amplification product corresponding to one or more of the mutant and wild-type *Braf, APC, PTEN* and/or *PI3K* mRNA sequences.

196. The method of claim 178, wherein the RNA is mRNA that codes for *Braf* and for *KRAS,* or mRNA that codes for *KRAS* and *APC,* or mRNA that codes for *KRAS* and *PTEN*, or mRNA that codes for *KRAS* and *PI3K*.

197. The method of claim 196 wherein the padlock probe(s) are configured to hybridize to cDNA(s) corresponding to wild-type *KRAS* and *Braf* mRNA sequences; or corresponding to wild-type *KRAS* and *APC* mRNA sequences*;* or corresponding to wild-type *KRAS* and *PTEN* mRNA sequences; or corresponding to wild-type *KRAS* and *PI3K* mRNA sequences.

198. The method of any one of claims 1 to 197, wherein the sample comprises a tissue section, with the proviso that the tissue section is not, or does not include a preparation of a fresh frozen tissue.

199. The method of any one of claims 1 to 198, wherein the sample comprises a tissue section, with the proviso that the tissue section is not or does not include a preparation including seeding on Superfrost Plus slides.

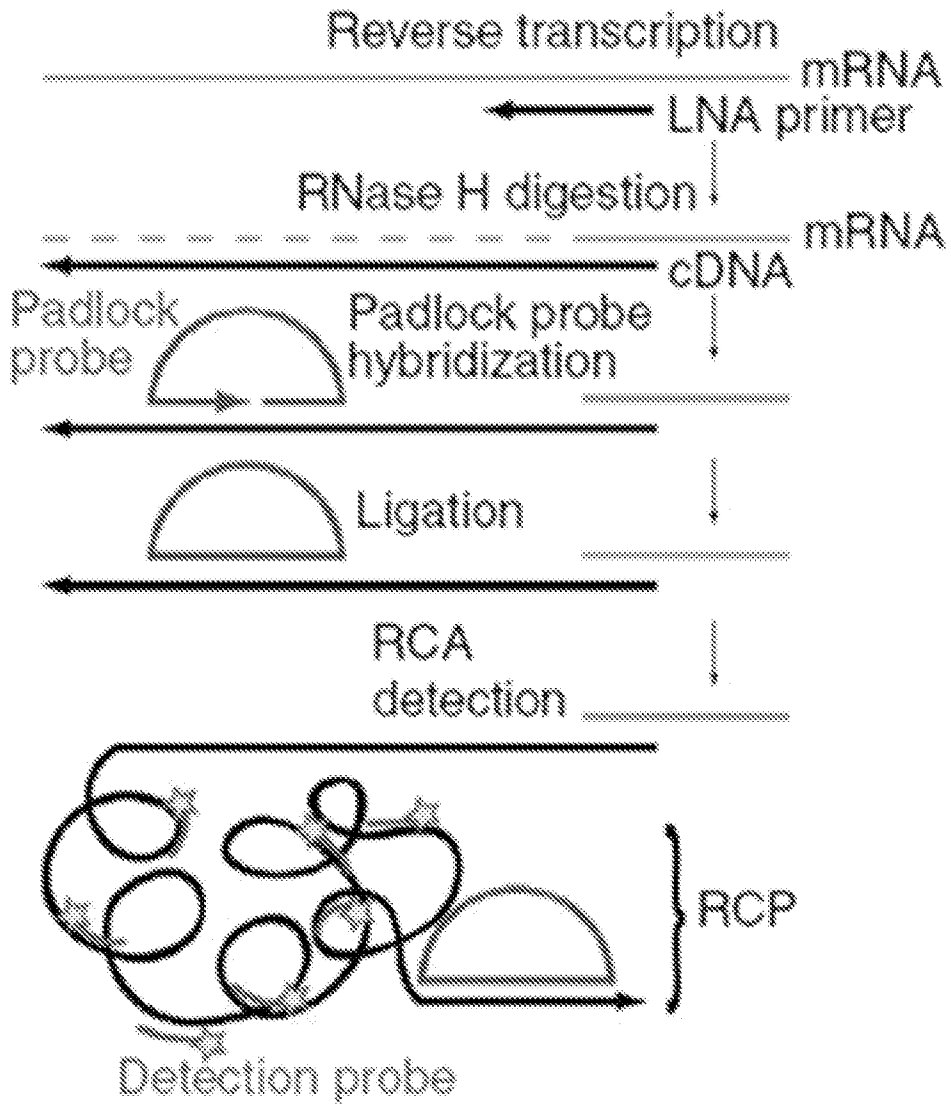
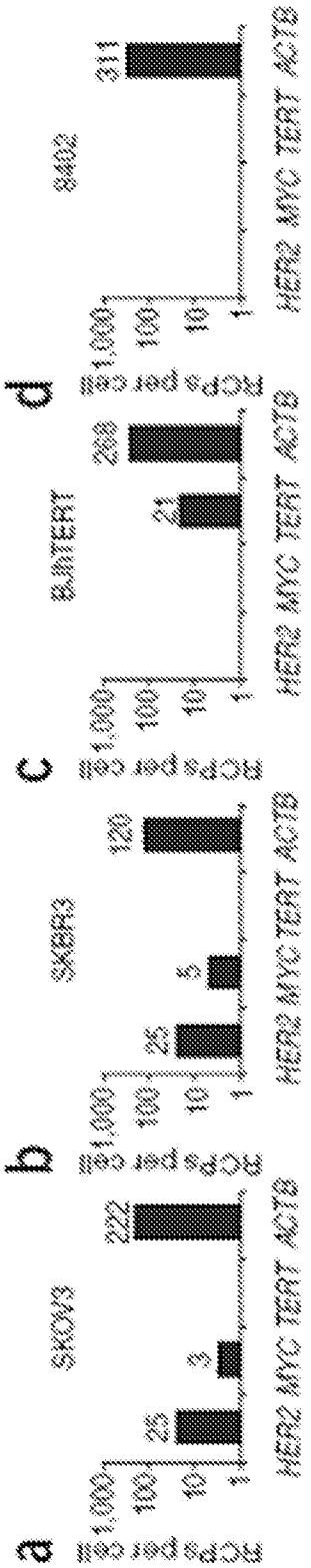

Reverse transcription
mRNA
LNA primer
RNase H digestion
mRNA
cDNA
Padlock probe
Padlock probe hybridization
Ligation
RCA detection
RCP
Detection probe

FIG. 1


a
SKOV3
RCPs per cell
1,000
100
10
1
25
3
222
HER2 MYC TERT ACTB
b
SKBR3
RCPs per cell
1,000
100
10
1
25
5
120
HER2 MYC TERT ACTB
c
BJNTERT
RCPs per cell
1,000
100
10
1
21
208
HER2 MYC TERT ACTB
d
8402
RCPs per cell
1,000
100
10
1
311
HER2 MYC TERT ACTB

FIGs. 2a-d

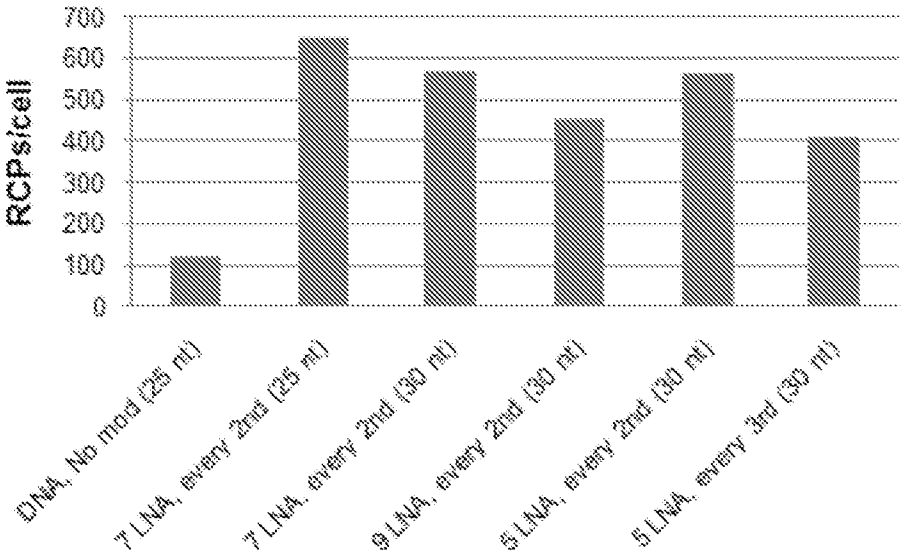

RCPs/cell
700
600
500
400
300
200
100
0
DNA, No mod (25 nt)
7 LNA, every 2nd (25 nt)
7 LNA, every 2nd (30 nt)
9 LNA, every 2nd (30 nt)
6 LNA, every 2nd (30 nt)
6 LNA, every 3rd (30 nt)

FIG. 3

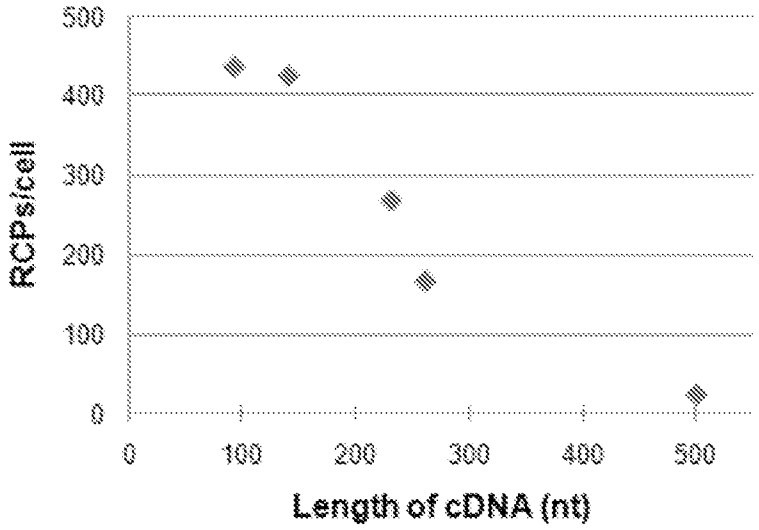

500
400
300
200
100
0
RCPs/cell
0
100
200
300
400
500
Length of cDNA (nt)

FIG. 4

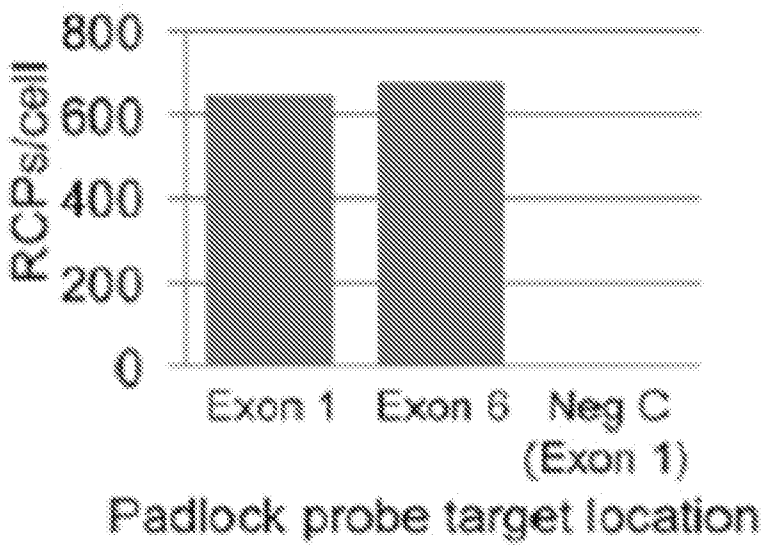
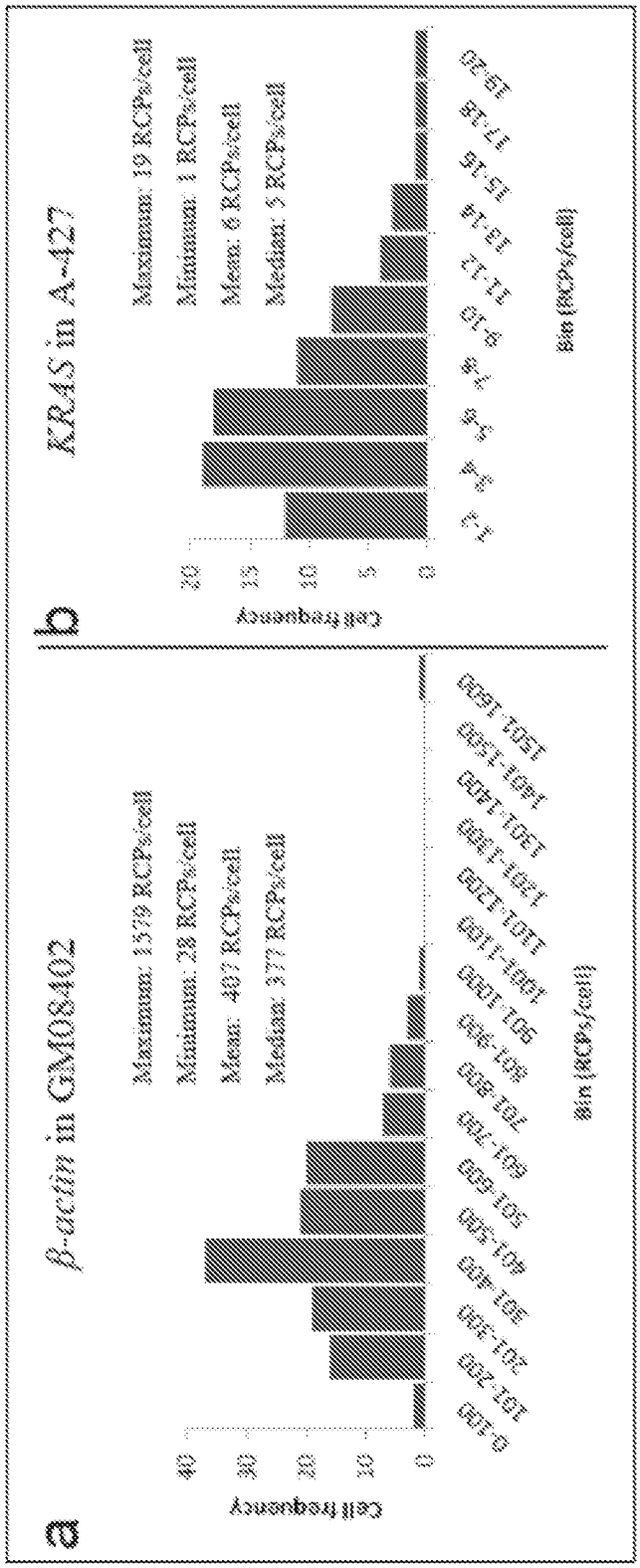

800
600
400
200
0
RCPs/cell
Exon 1
Exon 6
Neg C
(Exon 1)
Padlock probe target location

FIG. 5


a
β-actin in GM08402
Maximum: 1579 RCPs/cell
Minimum: 28 RCPs/cell
Mean: 407 RCPs/cell
Median: 377 RCPs/cell
Cell frequency
Bin (RCPs/cell)
b
KRAS in A-427
Maximum: 19 RCPs/cell
Minimum: 1 RCPs/cell
Mean: 6 RCPs/cell
Median: 5 RCPs/cell
Cell frequency
Bin (RCPs/cell)

FIGs. 6a-b

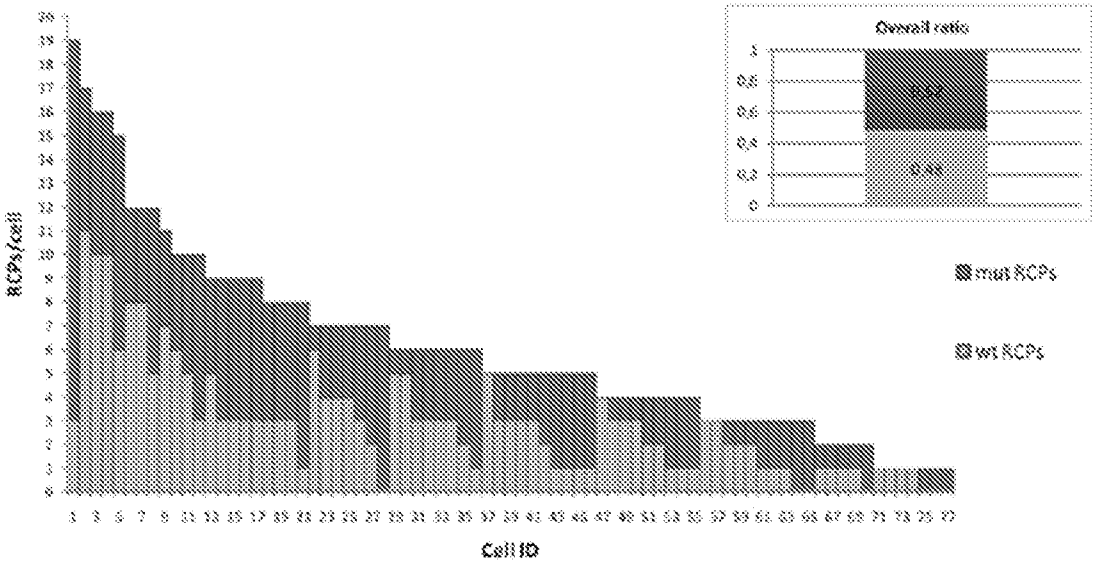


FIG. 7

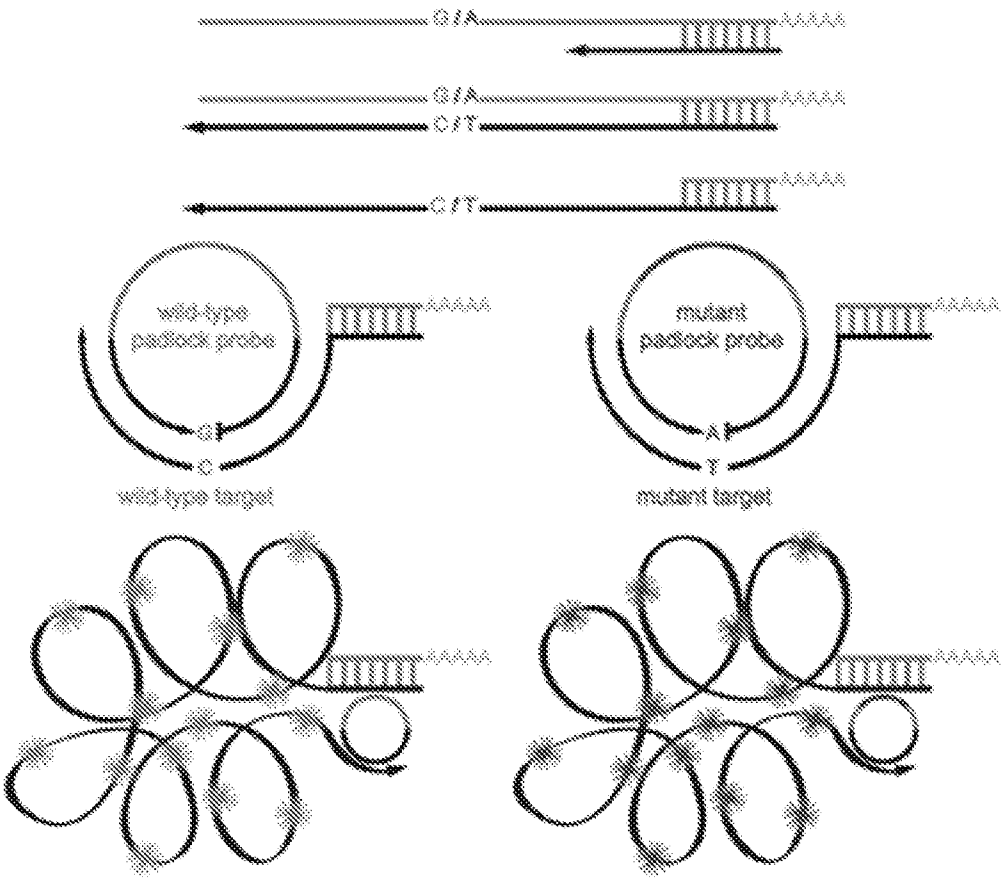
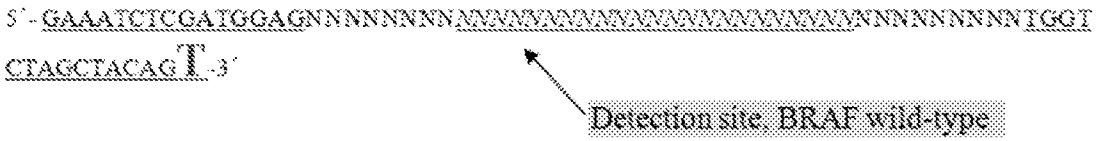

G / A
G / A
C / T
C / T
wild-type padlock probe
mutant padlock probe
G
C
A
T
wild-type target
mutant target

FIG. 8

Example of padlock probe (70 nt long) targeting BRAF wild-type:



Example of padlock probe (70 nt long) targeting BRAF mutant:

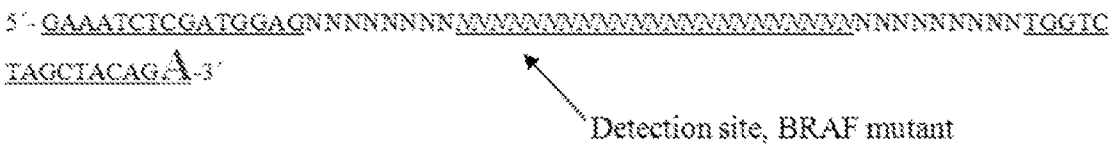


FIG. 9

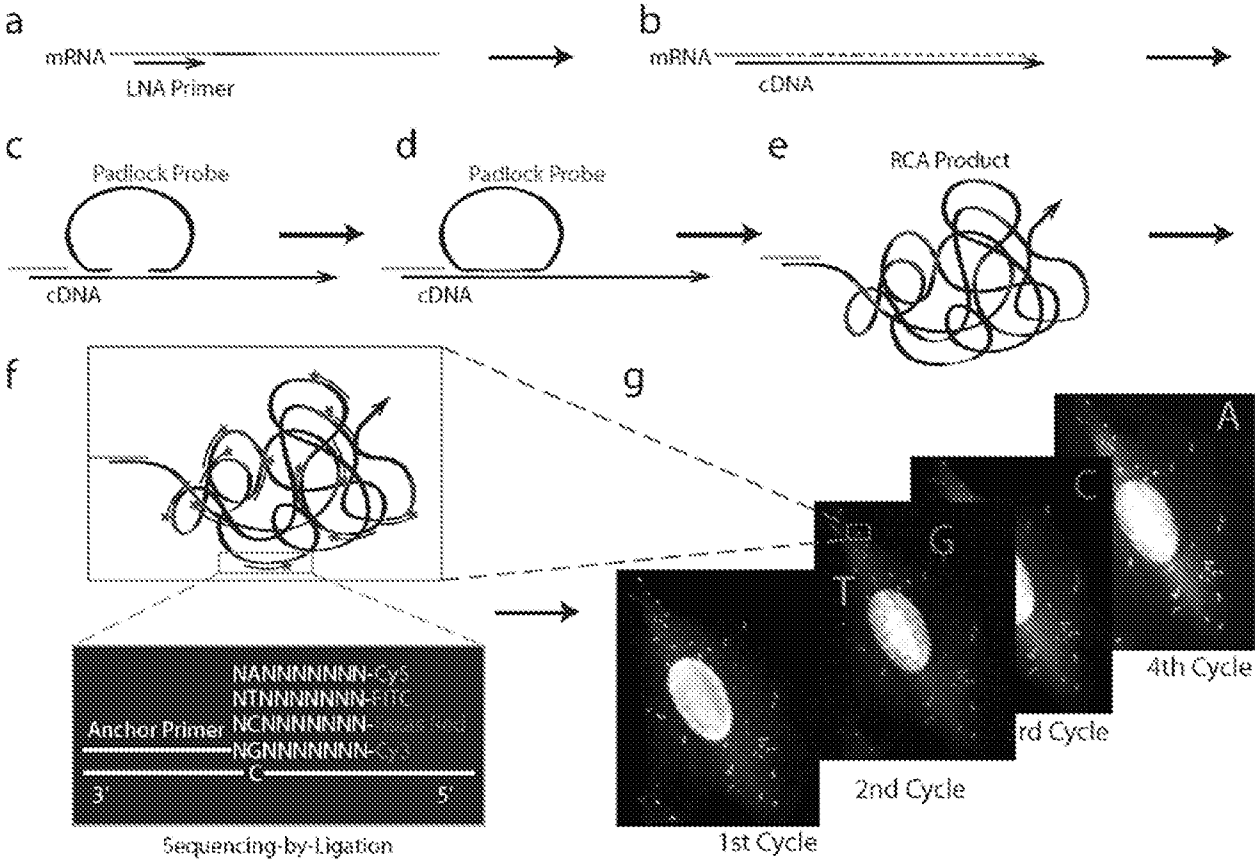

a
mRNA
LNA Primer
b
mRNA
cDNA
c
Padlock Probe
cDNA
d
Padlock Probe
cDNA
e
RCA Product
f
g
NANNNNNNN-Cy5
NTNNNNNNN-FITC
Anchor Primer NCNNNNNNN-
NGNNNNNNN-
3'
5'
Sequencing-by-Ligation
A
C
G
T
4th Cycle
rd Cycle
2nd Cycle
1st Cycle

FIG. 10a-g

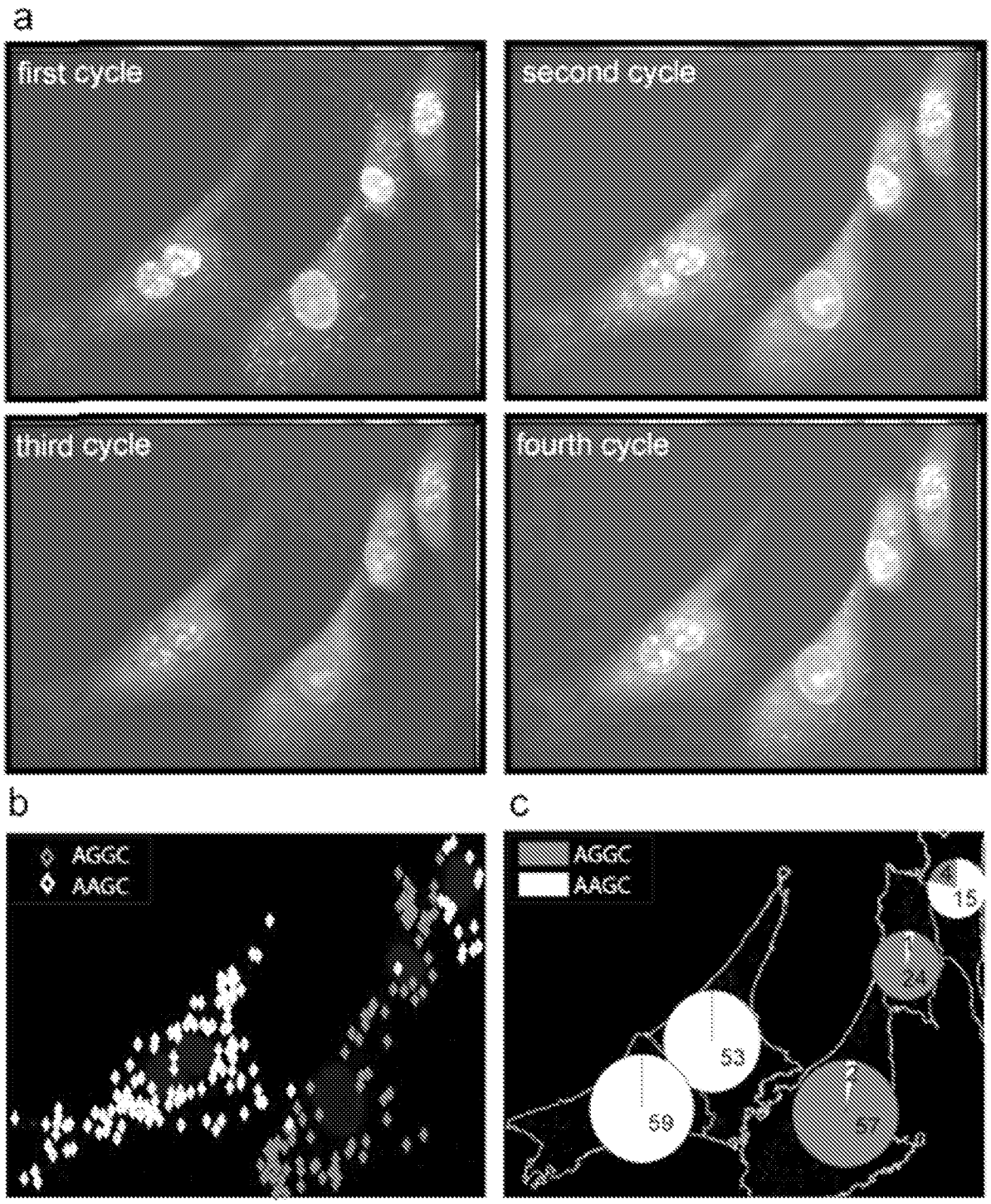

a
first cycle
second cycle
third cycle
fourth cycle
b
AGGC
AAGC
c
AGGC
AAGC
15
24
53
59
57

FIG. 11a-c

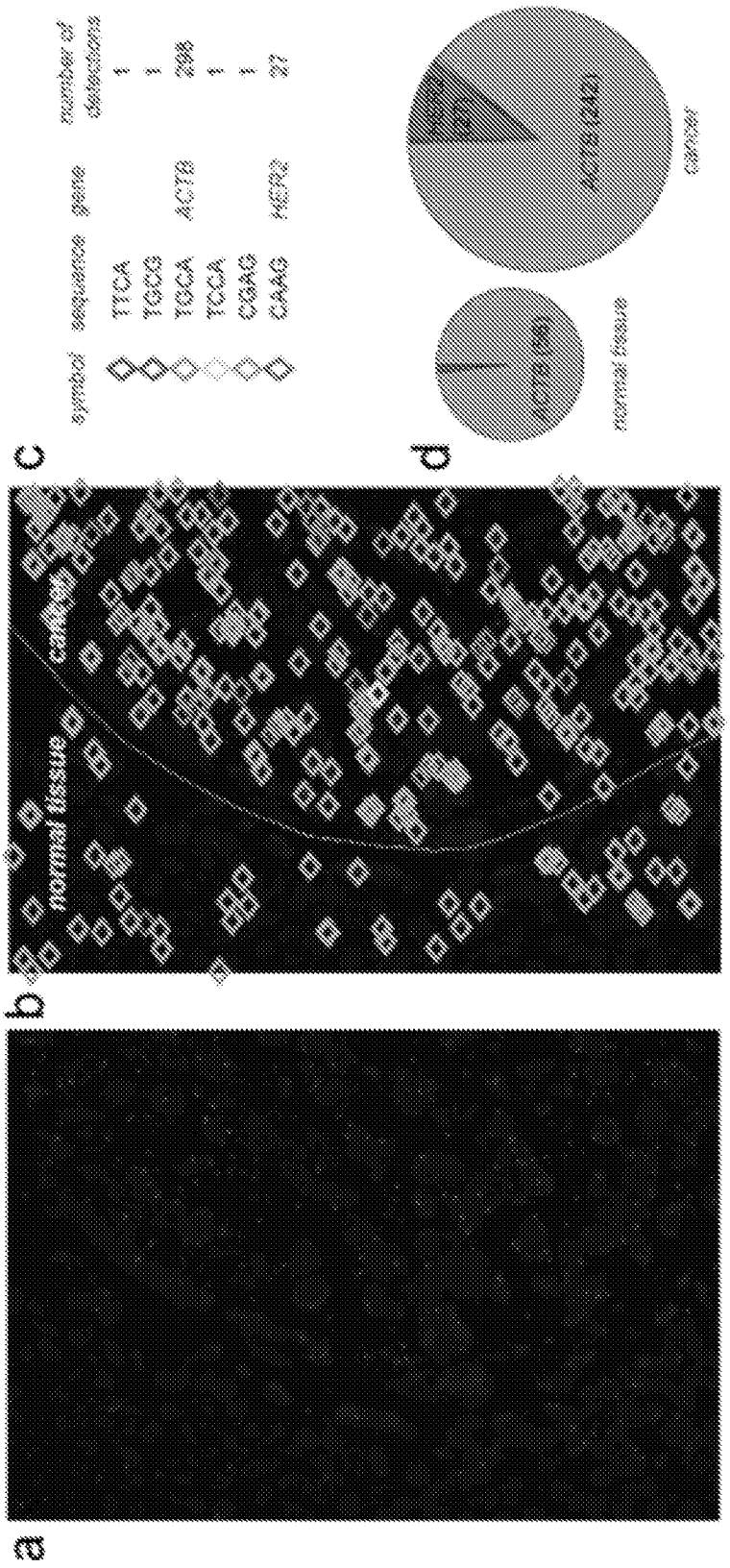

a
b
normal tissue
cancer
c
symbol
sequence
gene
number of deletions
TTCA
1
TGCG
1
TGCA
ACTB
286
TCCA
1
CGAG
1
CAAG
HER2
27
d
normal tissue
cancer

FIG. 12a-d

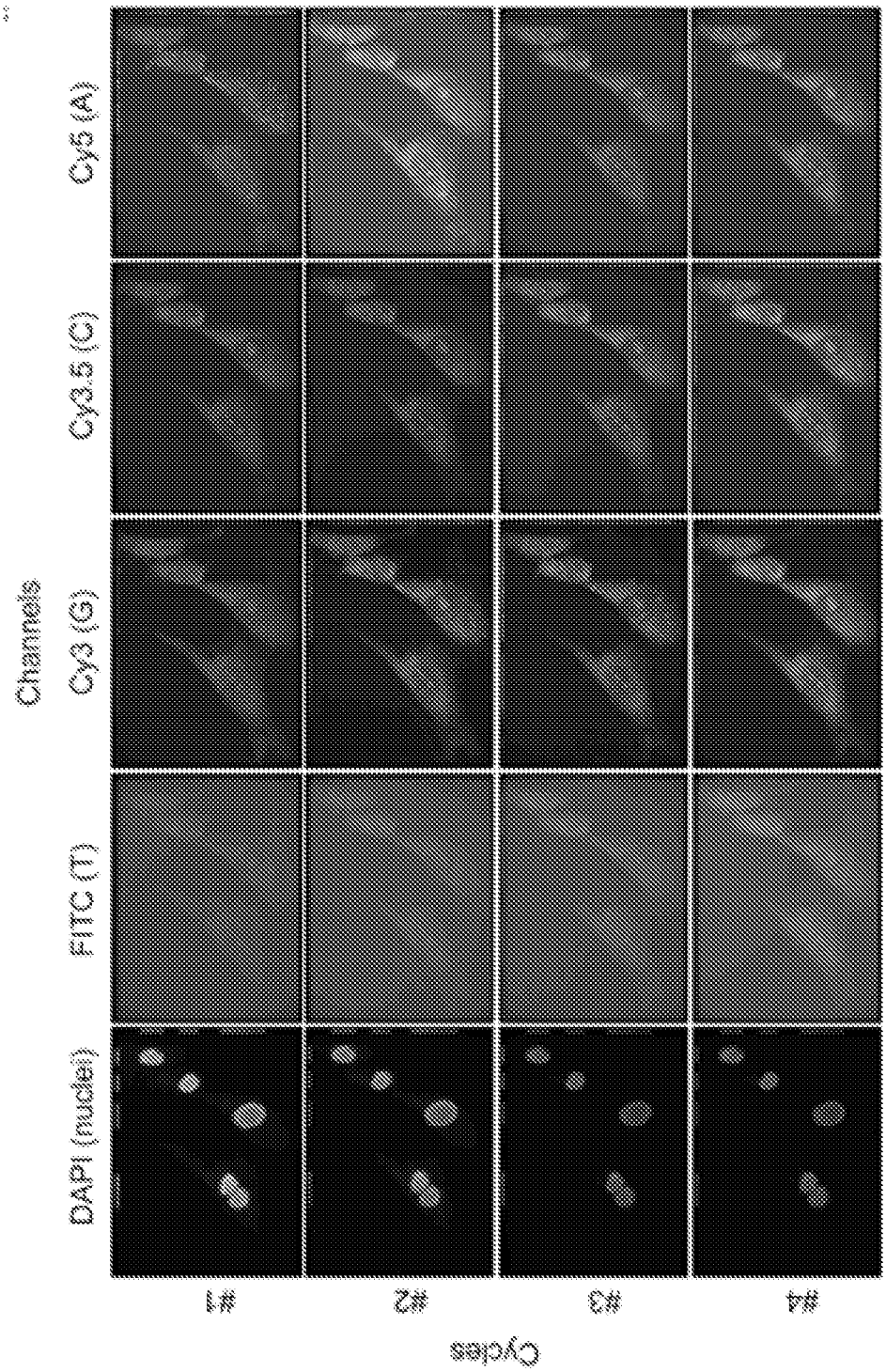


FIG. 13

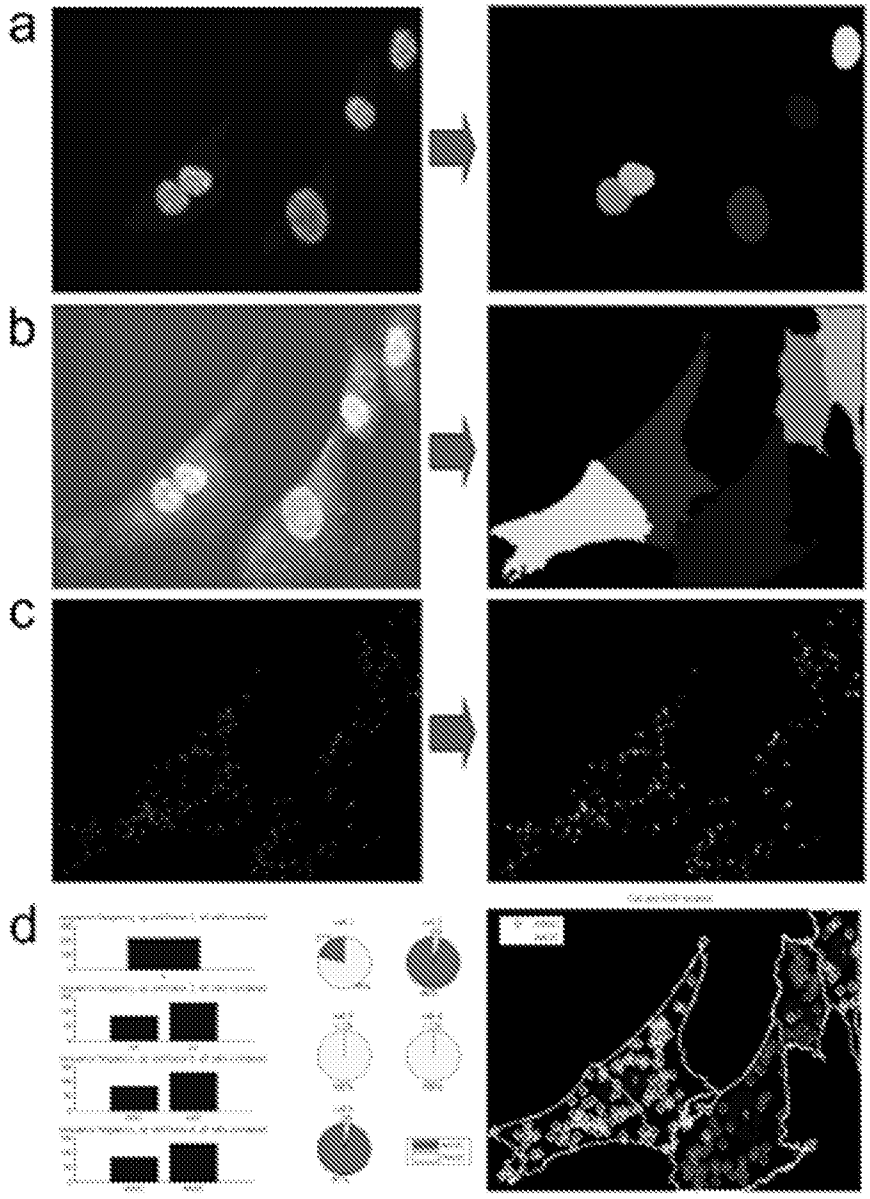


FIG. 14a-d

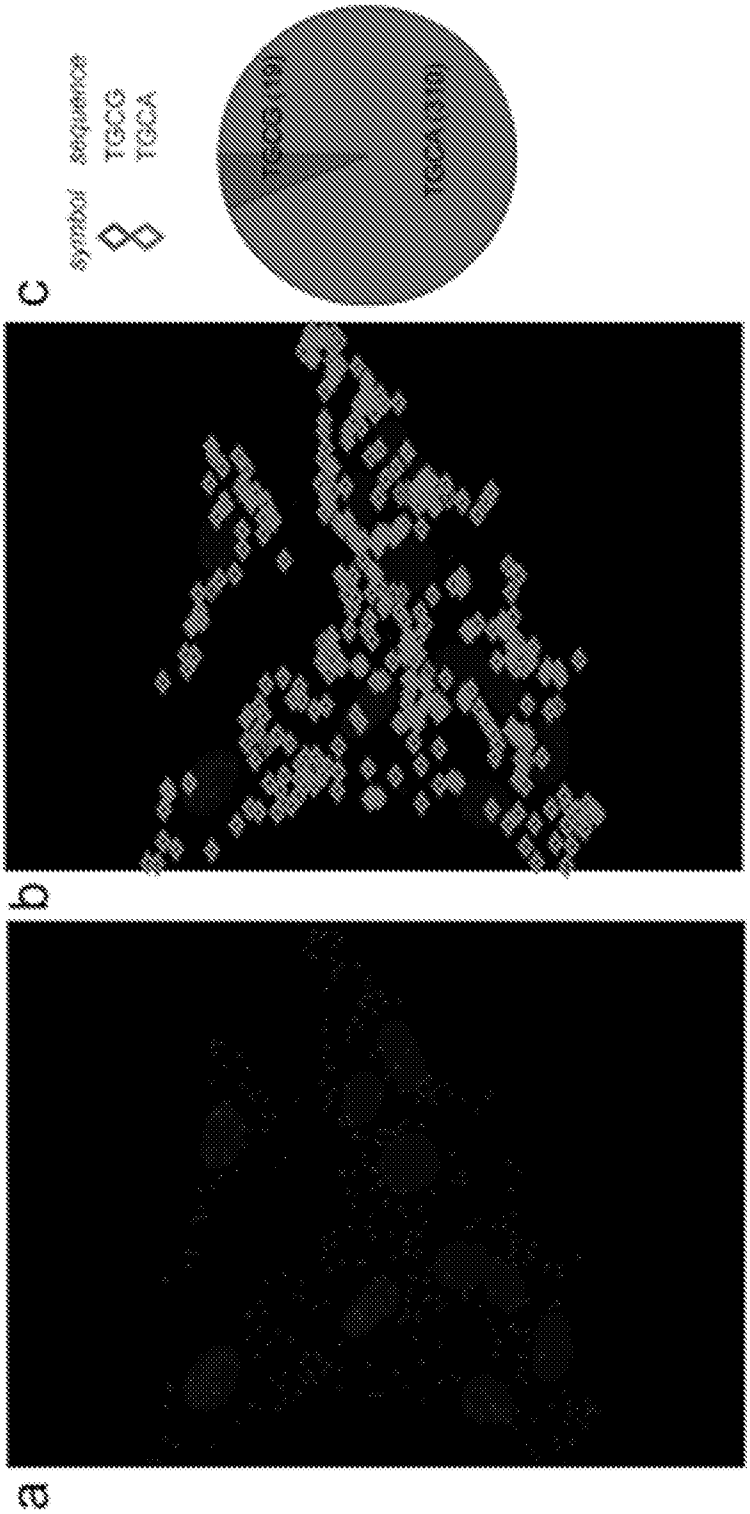


FIG. 15a-c

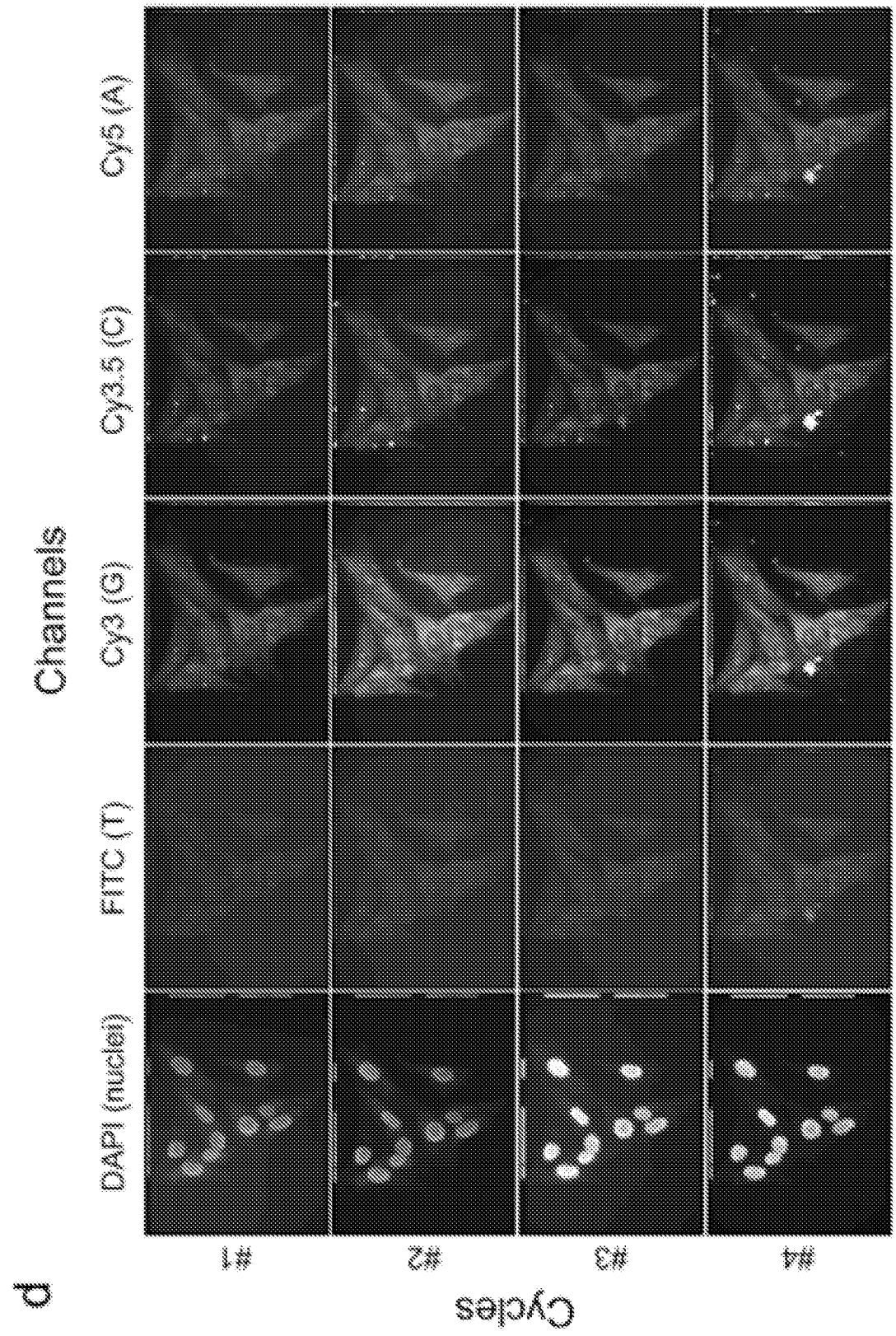


FIG. 15d

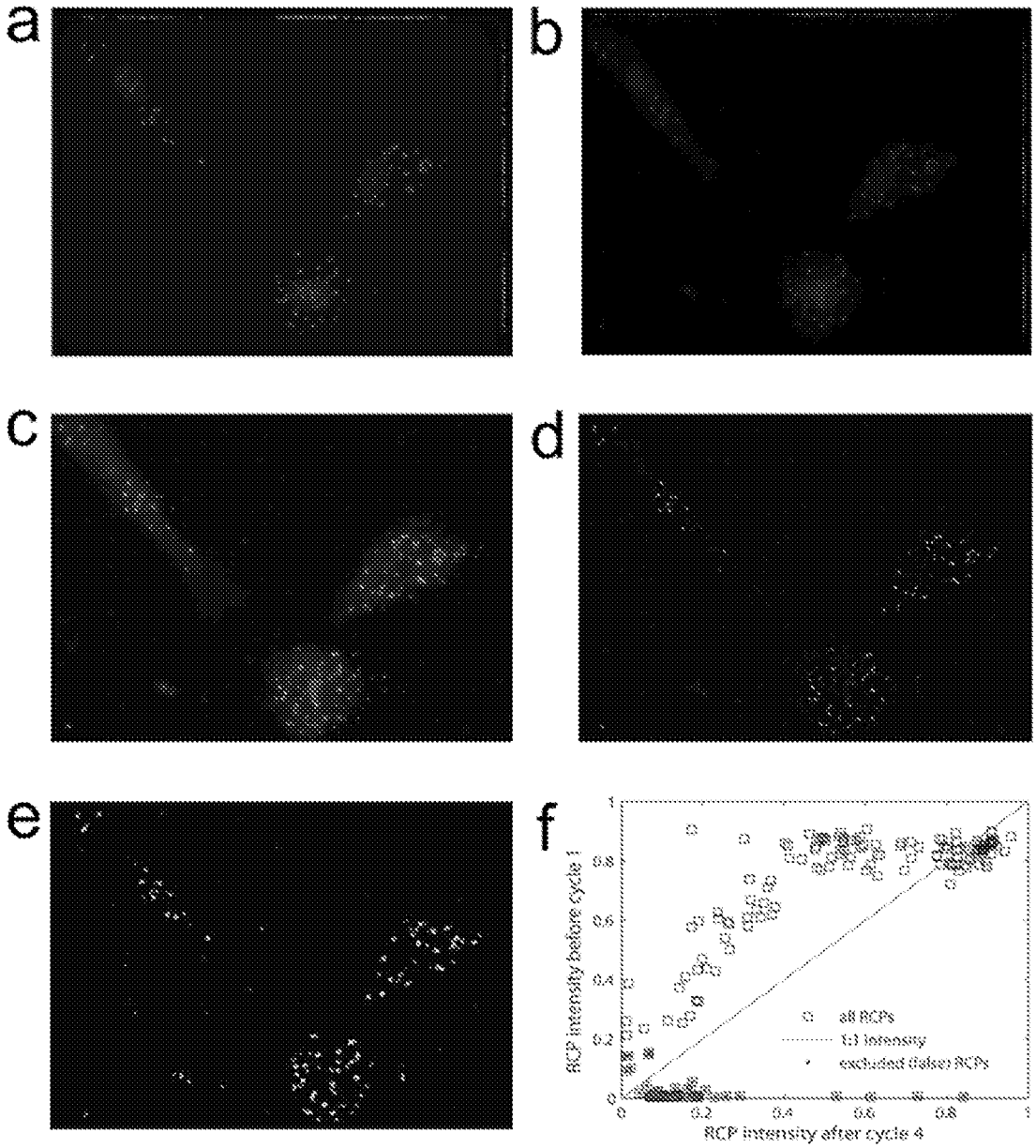

a
b
c
d
e
f
RCP intensity before cycle 1
RCP intensity after cycle 4
all RCPs
1:1 intensity
excluded (fake) RCPs
0
0.2
0.4
0.6
0.8
1

FIG. 16a-f

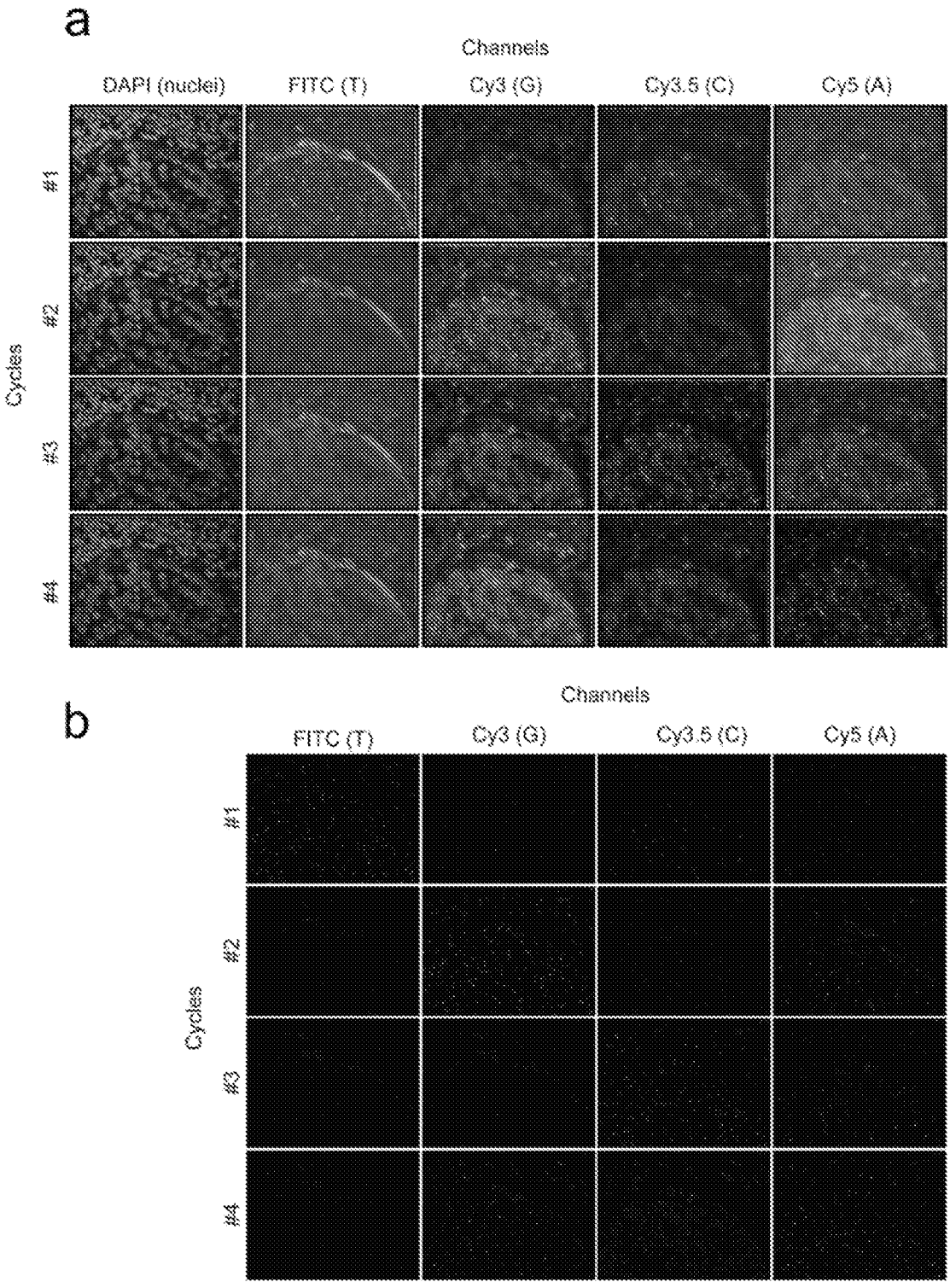

a
Channels
DAPI (nuclei)
FITC (T)
Cy3 (G)
Cy3.5 (C)
Cy5 (A)
Cycles
#1
#2
#3
#4
b
Channels
FITC (T)
Cy3 (G)
Cy3.5 (C)
Cy5 (A)
Cycles
#1
#2
#3
#4

FIG. 17a-b

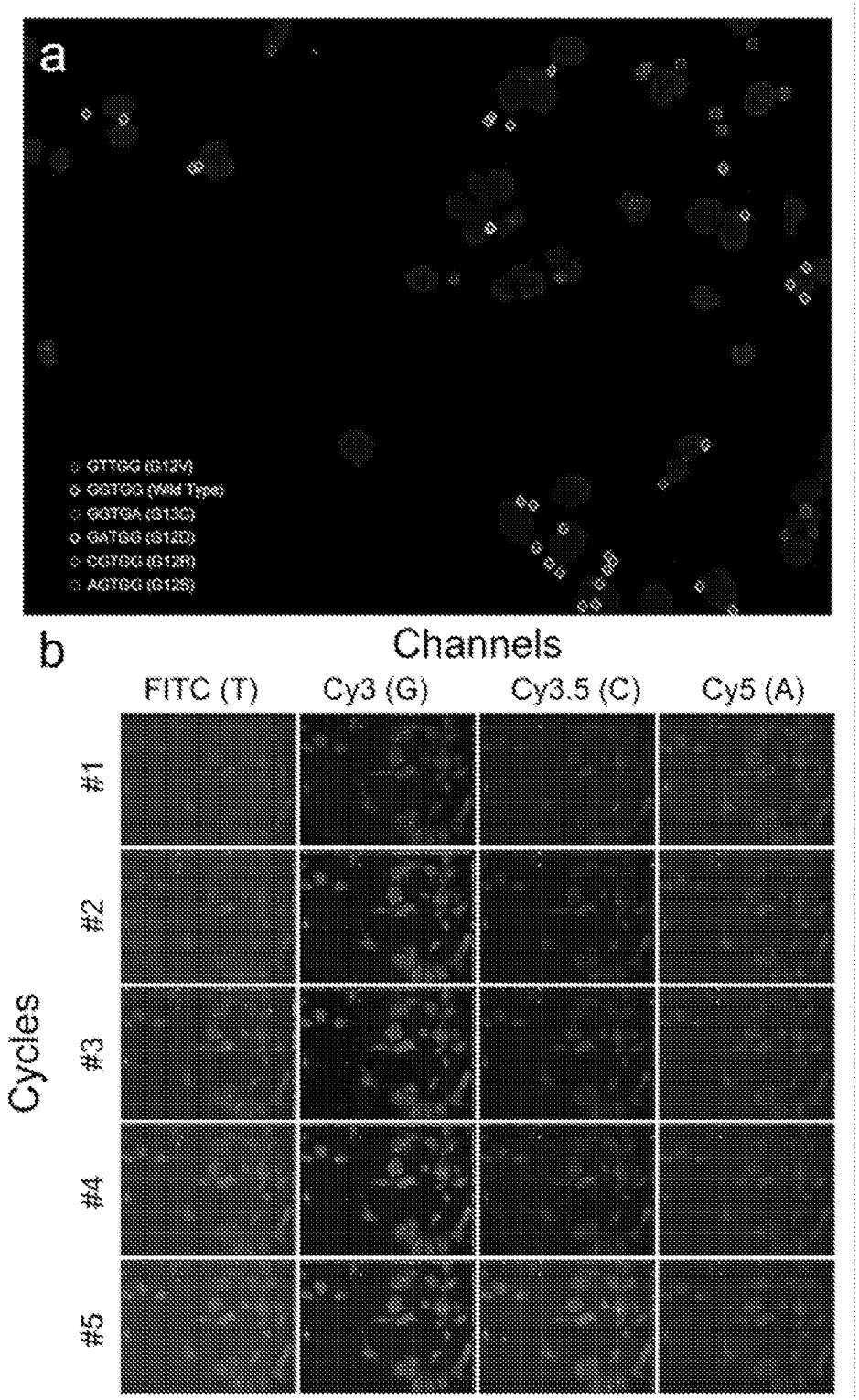

a
GTTGG (G12V)
GGTGG (Wild Type)
GGTGA (G13C)
GATGG (G12D)
CGTGG (G12R)
AGTGG (G12S)
b
Channels
FITC (T)
Cy3 (G)
Cy3.5 (C)
Cy5 (A)
Cycles
#1
#2
#3
#4
#5

FIG. 18a-b