tube may comprise an optical device capable of transmitting a light force, e.g., UV light, to the at least one cell. The pipette or a microarray pin may be attached to an array comprising a plurality of pipettes or microarray pins. The proximal solution may comprise an anionic polymer, e.g., dextran sulfate, and/or salmon sperm DNA and/or the collected signal oligonucleotide may be added to a solution comprising an anionic polymer, e.g., dextran sulfate, and/or salmon sperm DNA. Other non-specific blocking agents known in the art in addition to or instead of salmon sperm DNA may be used.

[0015] In embodiments, the method provides simultaneous spatially-resolved DNA, RNA, and/or protein detection of a tissue sample.

[0016] In embodiments, digital readout comprises a linear dynamic range of greater than or equal to 5 logs.

[0017] In embodiments, probes are provided to a sample at concentrations typically less than that used for immunohistochemistry (IHC) or for in situ hybridization (ISH). Alternately, the concentration may be significantly less than that used for IHC or ISH. For example, the probe concentration may be 2 fold less, 5 fold less, 10 fold less, 20 fold less, 25 fold less, 30 fold less, 50 fold less, 60 fold less, 70 fold less, 80 fold less, 90 fold less, 100 fold less, 200 fold less, 300 fold less, 400 fold less, 500 fold less, 600 fold less, 700 fold less, 800 fold less, 900 fold less, 1000 fold less, 2000 fold less, or less and any number in between. In embodiments, probes are provided at a concentration of 100 nM, 70 nM, 60 nM, 50 nM, 40 nM, 30 nM, 20 nM, 10 nM, 9 nM, 8 nM, 7 nM, 6 nM, 5 nM, 4 nM, 3 nM, 2 nM, 1 nM, 0.9 nM, 0.8 nM, 0.7 nM, 0.6 nM, 0.5 nM, 0.4 nM, 0.3 nM, 0.2 nM, 0.1 nM, 0.09 nM, 0.08 nM, 0.07 nM, 0.06 nM, 0.05 nM, 0.04 nM, 0.03 nM, 0.02 nM, 0.01 nM, and less and any concentration in between.

[0018] In embodiments, a tissue sample is attached to a slide and is first imaged using fluorescence (e.g., fluorescently-labeled antibodies and fluorescent stains (e.g., DAPI)) and then expression of proteins and/or nucleic acids is digitally counted from the sample.

[0019] In embodiments, a negative purification, e.g., comprising an affinity purification method comprising contacting intact probe molecules with an immobilized oligonucleotide that is complementary to a portion of the intact probe or an immobilized antibody or protein-binding motif that recognizes and binds to a portion of the intact probe, is used to remove intact probe molecules from the released signal oligonucleotides. In embodiments, the intact probe's target binding domain comprises a universal purification tag or sequence that is partially

complementary to the immobilized oligonucleotide or is capable of being recognized or bound by the immobilized antibody or protein-binding motif. Any such tag or sequence well-known in the art may be used in these embodiments.

[0020] Any aspect or embodiment described herein can be combined with any other aspect or embodiment as disclosed herein. While the disclosure has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the disclosure, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

[0021] The patent and scientific literature referred to herein establishes the knowledge that is available to those with skill in the art. All United States patents and published or unpublished United States patent applications cited herein are incorporated by reference. All published foreign patents and patent applications cited herein are hereby incorporated by reference. All other published references, documents, manuscripts and scientific literature cited herein are hereby incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

[0023] **Figure 1**: Shows two exemplary probes. Nucleic acid backbones (either single-stranded DNA or single-stranded RNA) are shown as a straight, black line. The probes each include a target-binding domain, shown in red. The top probe includes labeled RNA segments hybridized to the nucleic acid backbone whereas the bottom probe includes labeled DNA oligonucleotides hybridized to the nucleic acid backbone. A cleavable motif (*e.g.*, a cleavable linker, not shown) may be located between the backbone and a target-binding domain or within the backbone. The cleavable motif allows release of a signal oligonucleotide from a bound target nucleic acid or protein; then, the signal oligonucleotide is collected and detected.

[0024] **Figure 2**: Shows a first type of a probe that can bind directly to a target nucleic acid (top). In the below image, the probe has bound to the target nucleic acid, shown as a blue curvilinear line. In this figure and in later figures, a reporter probe includes six positions

hybridized to labeled oligonucleotides (identified by a colored circle). Since the probe comprises positions that can be hybridized to labeled oligonucleotides, the probe can also be referred to as a reporter probe.

[0025] Figure 3: Shows a first type of dual-probe composition of the present invention. Here, the first type of probe binds directly to a target nucleic acid and a first type capture probe binds directly to the target nucleic acid. The capture probe may include at least one affinity reagent, which is shown as an asterisk. The target nucleic acid in a sample is shown as a blue curvilinear line.

[0026] Figure 4: Shows a second type of a probe (or reporter probe) that can bind indirectly to a target nucleic acid in a sample (top). Here, the probe's target is an intermediary oligonucleotide, shown in green, which in turn binds to the target nucleic acid in a sample, shown as a blue curvilinear line in the bottom image. It could be said that the intermediary oligonucleotide is a probe, as defined herein, since it comprises a nucleic acid backbone and is capable of binding a target nucleic acid.

[0027] Figure 5: Shows a second type of dual-probe composition of the present invention. Here, the second type of probe binds indirectly to a target nucleic acid in a sample (*via* an intermediary oligonucleotide, shown in green) and a second type capture probe binds indirectly to the target nucleic acid in the sample (*via* another intermediary oligonucleotide, shown in orange). The capture probe may include at least one affinity reagent, which is shown as an asterisk.

[0028] Figure 6: Shows release of signal oligonucleotides from second type probes (illustrated in Figure 4) that are bound indirectly to a target nucleic acid in a sample. The location of a cleavable motif within a probe (or in a reporter probe) affects which material is included with a released signal oligonucleotide.

[0029] Figure 7: Shows three types of probes used for detecting proteins. In the top configuration, a probe comprises a nucleic acid attached to a protein-binding domain; in this configuration, a cleavable motif (*e.g.*, a cleavable linker, not shown) may be included between the nucleic acid and protein-binding domain or within the nucleic acid itself. In the middle configuration, a protein-binding domain is attached to a nucleic acid and a probe hybridizes to the nucleic acid. The probe (comprising the target-binding domain and the nucleic acid attached to the protein-binding domain (shown in green)) can be bound by a probe before or after the target binding domain binds a protein target (As shown in Figure 8). A cleavable motif may be

included in either or both of the backbone or the nucleic acid attached to the protein-binding domain. The first or second type probe shown in Figures 2 and 4 may be used in this configuration for detecting a protein. In the bottom configuration, a protein-binding domain is attached to a nucleic acid and an intermediary oligonucleotide (shown in red) hybridizes to both a probe and to the nucleic acid attached to the protein-binding domain. The first or second type probe shown in Figures 2 and 4 may be used in this configuration for detecting a protein.

[0030] Figure 8: Shows the middle and bottom probes of Figure 7. The top two images show the probe before and after it has bound a protein. The next image shows the probe after its cleavable motif has been cleaved; in this image the cleavable motif is between the nucleic acid and the target binding domain. Once the nucleic acid has been released, it can be considered a signal oligonucleotide. In the bottom image, the signal oligonucleotide (released nucleic acid of the probe) is bound by a reporter probe (*e.g.*, as shown in Figures 2 and 4).

[0031] Figure 9 Shows release of signal oligonucleotides from a probe of the middle configuration shown in Figure 7 and the probes of Figure 8. The location of a cleavable motif within a probe (or in a reporter probe) affects which material is included with a released signal oligonucleotide.

[0032] Figure 10: Shows steps in a method of the present invention in which signal oligonucleotides from one region-of-interest (ROI) are detected.

[0033] Figure 11: Shows steps in a method of the present invention in which regions-of-interest are located on a first serial section of a tissue sample and probes are applied to a second serial section of the tissue sample. Signal oligonucleotides are released and collected from probes bound to targets in a first region-of-interest of the second serial section. Then, signal oligonucleotides are released and collected from probes bound to targets in a second (up to the nth) region-of-interest of the second serial section.

[0034] Figure 12: Shows multiplexed detection of a plurality of target nucleic acids and/or proteins from a first region-of-interest followed by multiplexed detection of the plurality of target nucleic acids and/or proteins from a second region-of-interest.

[0035] Figure 13: Illustrates steps in methods of the present invention. The method shown may be referred herein as "nCounter® Digital Multiplexed Immunohistochemistry (IHC)".

[0036] **Figure 14**: Is a flow chart demonstrating the simplified workflow and higher multiplexing capable with nCounter® Digital Multiplexed IHC (top) when compared to standard TSA-based multiplexed IHC (bottom).

[0037] **Figure 15**: Are photographs showing a digital mirror device (DMD) attached to a Ti-E microscope (top) and a brightfield image of a FFPE tissue section (bottom). Light illumination (white spots) on the FFPE tissue (bright field image) shows multiple ROIs of about ~10-20 μm in size, i.e., the size of a single cell.

[0038] **Figure 16**: Illustrates components and light paths involved with the present invention when the method includes use of a digital mirror device (DMD). Wide-field illumination with the DMD focused onto sample. LED provides sufficient illumination to excite whole field of view at once and with single cell illumination such that ~80-600 DMD pixels illuminate a 10 μm diameter cell. A normal-grade DMD will provide sufficient single-cell resolution. DS: Dichroic mirror, FW: Filter wheel, and DMD: Digital mirror device.

[0039] **Figure 17**: Illustrates components and light paths involved with the present invention when the method includes use of a laser scanning device (e.g., confocal scanning device). In a confocal scanning configuration, galvo-mirrors direct light. This method requires an inexpensive 405 nm laser. DS: Dichroic mirror, FW: Filter wheel, and MM: Motorized mirror.

[0040] **Figure 18**: Shows a photomicrograph establishing overall tissue morphology of a tonsil sample that was initially imaged using two-color fluorescence of Ki-67 (cell proliferation marker; in green) and CD3 (immune cell marker; in red). Twelve regions (including the four regions magnified in **Figure 19**) are identified with white boxes.

[0041] **Figure 19**: Is a graph showing nCounter® data counts of Ki-67 and CD3 for four regions shown in **Figure 18**. Images were obtained from serial sections (to allow various additional controls to be examined). In general, samples can be imaged with fluorescent antibodies and then digitally counted (via *uv*-exposure) using the same slide. Multiple targets analyzed across twelve regions (including the four regions shown here) show distinct profiles of localization of Ki-67 and CD3. Below the graph are magnifications of the four regions.

[0042] **Figure 20**: Shows exemplary counts from a 30-plex oligo-antibody cocktail on the twelve regions of interest (ROI) from the tonsil sample shown in **Figure 18**. Data was obtained from serial sections (to allow various additional controls to be examined).

[0043] **Figure 21**: **S**hows a photomicrograph establishing overall tissue morphology of T cells in a melanoma sample from a lymph node that was initially imaged using three-color fluorescence of CD3 (in red), CD8 (in green), and DAPI (in blue). The white circle is 25 μm in diameter and surrounds three cells.

[0044] **Figure 22:** Shows nCounter® data of CD3 conjugate release from FFPE lymph node tissue sections (5 μm thickness) as a function of UV illumination area (100 μm to 1 mm in diameter). The field diaphragm size is shown below the figure.

[0045] **Figure 23**: **S**hows nCounter® data for CD45 conjugate release from FFPE lymph node tissue sections (5 μm thickness) as a function of UV illumination area (100 μm to 1 mm in diameter) and from the same experiment as shown in **Figures 21** and **22**.

[0046] **Figure 24**: **S**hows nCounter® data for PD1 conjugate release from FFPE lymph node tissue sections (5 μm thickness) as a function of UV illumination area (100 μm to 1 mm in diameter) and from the same experiment as shown in **Figures 21** to **23**.

[0047] **Figure 25**: Shows a tissue microarray (TMA; left panel) of breast tumor tissue containing variable levels of Her2 protein as shown in the photomicrograph (center panel) which identifies Her2 fluorescence by IHC staining. The right panel shows a magnification of a single region of the central panel.

[0048] **Figure 26**: Shows nCounter® count data for forty-eight representative regions versus Her2 status (ASCO-CAP guidelines).

[0049] **Figure 27**: Plots nCounter® Counts versus Sum Pixel Intensities ($x10^3$) for the forty-eight regions mentioned with respect to **Figure 26**.

[0050] **Figure 28**: Is a photomicrograph establishing overall tissue morphology of a melanoma sample that was initially imaged using two-color fluorescence of CD3 (in red) and DAPI (in blue). Ten exemplary regions are identified with white boxes.

[0051] **Figure 29**: Shows exemplary counts from a 30-plex oligo-antibody cocktail on the ten regions of interest (ROI) from the sample shown in **Figure 28**.

[0052] **Figure 30A** and **Figure 30B**: Are photomicrographs showing UV illumination using a digital mirror device (DMD) of single cells (in blue) in a tonsil tissue sample (in green).

[0053] **Figures 31A** to **31D:** Are photomicrographs showing UV illumination using a digital mirror device (DMD) of single cells (in bright white) in a tonsil tissue sample. **Figure 31B**

highlights the single cells noted in **Figure 31A**; **Figure 31D** highlights the single cell noted in **Figure 31C.**

[0054] **Figure 32**: Shows steps in a spatially-resolved FFPE Tissue Protein Assay. The steps are similar to those of a nucleic acid-detecting assay except, in the nucleic acid-detecting assay, the sample is bound with a probe comprising a nucleic acid target-binding domain rather than an antibody.

[0055] **Figure 33**: Shows steps in a spatially-resolved FFPE Tissue Protein Assay.

[0056] **Figure 34**: Shows data from an embodiment in which a whole tissue or whole sample is illuminated, e.g., with a standard UV gel box, to release signal oligonucleotides previously attached to a probe.

[0057] **Figure 35**: Shows an embodiment in which a portion of a tissue or sample is illuminated, e.g., with a microscope, i.e., UV cleavage under Microscope (Time titration experiment).

[0058] **Figure 36:** Shows an embodiment in which a portion of a tissue or sample is illuminated, e.g., with a microscope, i.e., UV cleavage under Microscope (Illumination area titration experiment).

[0059] **Figure 37:** Shows an embodiment in which a portion of a tissue or sample is illuminated, e.g., with a microscope, i.e., UV cleavage under microscope (Illumination area titration experiment – multiple targets).

[0060] **Figure 38:** Shows an embodiment in which a region of interest in a tissue (e.g., a breast cancer sample) is first identified for expression of a marker and this region of interest is then illuminated (e.g., with UV) to release signal oligonucleotides from a probe.

[0061] **Figure 39**: Shows an embodiment in which a tissue is embedded in flow cell. Data for multiple fractions is shown. As with the data of **Figure 38**, here a region of interest is pre-identified for expression of a fluorescently-labeled marker. Also shown are photographs and a schematic showing configuration of the apparatus.

[0062] **Figure 40**: Shows an embodiment in which a tissue is embedded in flow cell with small holes. Also shown are photographs and a schematic showing configuration of the apparatus.

[0063] **Figures 41A** to **41C**: Shows embodiments using a flow cell with small holes have significant signal to noise improvement rather than collection of eluate from entire surface of tissue. Also shown are photographs and a schematic showing configuration of the apparatus.

[0064] **Figures 42A** to **42C**: Shows data in the embodiments using a flow cell with small holes (12 or 96 hole formats).

[0065] **Figures 43A** and **B**: Shows data in comparing background signal from flow cells in which whole tissue elution was performed (**Figure 43A**) and background signal from flow cells in which elution occurred directly above a region of interest (**Figure 43B**).

[0066] **Figure 44**: Is a schematic showing eluent collection with an open surface for a multi-region of interest aspiration embodiment. Here is shown a multi-tube array for aspiration/dispensing eluents with rotary valve selection.

[0067] **Figure 45**: Includes photographs and a schematic showing an embodiment in which eluent collection is through a capillary (micro-aspirator).

[0068] **Figures 46A** and **B**: Shows data from the embodiment of **Figure 45** in which eluent collection is through a capillary (micro-aspirator).

[0069] **Figure 47**: Is a schematic showing eluent collection with an open surface for a multi-region of interest aspiration embodiment or for a single region of interest. Here is shown a multi-tube array using pipetting vs capillary action for aspiration/dispensing and a single tube/pipet with fixed position.

[0070] **Figure 48**: Is a schematic showing illumination and fluid collection through a combined capillary and lens.

[0071] **Figure 49**: Is a schematic showing steps in an embodiment of a spatially-resolved FFPE tissue assay comprising a 96 well grid.

[0072] **Figure 50:** Shows protein expression data obtained from a single cell or two cells using the herein described methods and apparatuses.

[0073] **Figure 51:** Identifies regions of interests located on serial sections from a single tumor sample.

[0074] **Figure 52:** Shows counts obtained for six of the nine RNA probes included in the assay of **Example 16**.

[0075] **Figure 53**: Shows the averages and standard deviations of counts shown in **Figure 52**.

[0076] **Figure 54**: Shows RNA expression data and protein data for probes that were simultaneously hybridized to nCounter® Molecular Barcodes, and digitally counted by an nCounter® system from NanoString Technologies®.

[0077] **Figure 55**: Shows RNA expression data obtained from single-stranded DNA probes and partially double-stranded DNA probes.

[0078] **Figure 56**: Shows RNA expression data obtained from probes hybridized in the present of salmon sperm DNA.

[0079] **Figure 57**: Shows RNA expression data from a probe specific to PSA (Prostate-Specific Antigen).

[0080] **Figure 58**: Shows specificity of probes increase at non-standard, sub-nM concentrations.

DETAILED DESCRIPTION OF THE INVENTION

[0081] The present invention is based in part on probes, compositions, methods, and kits for simultaneous, multiplexed detection and quantification of protein and/or nucleic acid expression in a user-defined region of a tissue, user-defined cell, and/or user-defined subcellular structure within a cell.

[0082] The present invention provides a comparison of the identity and abundance of target proteins and/or target nucleic acids present in a first region of interest (*e.g.*, tissue type, a cell (including normal and abnormal cells), and a subcellular structure within a cell) and the identity and abundance of target proteins and/or target nucleic acids present in a second region of interest. There is no pre-defined upper limit to the number of regions of interest and comparisons that can be made; the upper limit relates to the size of the region of interest relative the size of the sample. As examples, when a single cell represent a region of interest, then a section may have hundreds to thousands of regions of interest; however, if a tissue section includes only two cell types, then the section may have only two regions of interest (each including only one cell type).

[0083] The present invention provides a higher degree of multiplexing than is possible with standard immunohistochemical or *in situ* hybridization methods. Standard immunohistochemical methods allow for maximal simultaneous detection of six to ten protein targets, with three to four protein targets being more typical. Similarly, *in situ* hybridization methods are limited to simultaneous detection of fewer than ten nucleic acid targets. The present invention provides detection of large combinations of nucleic acid targets and/or protein targets from a defined region of a sample. The present invention provides an increase in objective measurements by

digital quantification and increased reliability and consistency, thereby enabling comparison of results among multiple centers.

[0084] The probes of the present invention may have nucleic acid backbones (single-stranded DNA or RNA) having defined positions capable of being hybridized (non-covalently bound) with at least one labeled oligonucleotide. *See*, Figure 1. Such probes (which have defined positions capable of being hybridized with at least one labeled oligonucleotide are also referred herein as reporter probes. The number of positions on a reporter probe's backbone ranges from 1 to 100 or more. In embodiments, the number of positions ranges from 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 to 15, 20, 30, 40, or 50, or any range in between. Indeed, the number of positions (for detecting a target nucleic acid and/or for detecting a target protein) on a backbone is without limit since engineering such a backbone is well-within the ability of a skilled artisan. The number of target nucleic acids and/or proteins detectable by a set of probes depends on the number of positions included in the probes' backbones.

[0085] As used herein a labeled oligonucleotide relates to an RNA segment including a detectable label or a DNA oligonucleotide including a detectable label.

[0086] A position of a nucleic acid backbone may be hybridized (non-covalently bound) with at least one labeled oligonucleotide. Alternately, a position may be hybridized with at least one oligonucleotide lacking a detectable label. Each position can hybridize to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 21 to 100 labeled (or unlabeled) oligonucleotides or more. The number of labeled oligonucleotides hybridized to each position depends on the length of the position and the size of the oligonucleotides. A position may be between about 300 to about 1500 nucleotides in length. The lengths of the labeled (or unlabeled) oligonucleotides vary from about 20 to about 1500 nucleotides in length. In embodiments, the lengths of labeled (or unlabeled) oligonucleotides vary from about 800 to about 1300 ribonucleotides. In other embodiments, the lengths of labeled (or unlabeled) oligonucleotides vary from about 20 to about 55 deoxyribonucleotides; such oligonucleotides are designed to have melting/hybridization temperatures of between about 65 and about 85 °C, e.g., about 80 °C. For example, a position of about 1100 nucleotides in length may hybridize to between about 25 and about 45 oligonucleotides comprising, each oligonucleotide about 45 to about 25 deoxyribonucleotides in length. In embodiments, each position is hybridized to about 34 labeled oligonucleotides of

about 33 deoxyribonucleotides in length. The labeled oligonucleotides are preferably single-stranded DNA.

[0087] Each labeled oligonucleotide may be labeled with one or more detectable label monomers. The label may be at a terminus of an oligonucleotide, at a point within an oligonucleotide, or a combination thereof. Oligonucleotides may comprise nucleotides with amine-modifications, which allow coupling of a detectable label to the nucleotide.

[0088] Labeled oligonucleotides of the present invention can be labeled with any of a variety of label monomers, such as a fluorochrome, quantum dot, dye, enzyme, nanoparticle, chemiluminescent marker, biotin, or other monomer known in the art that can be detected directly (*e.g.,* by light emission) or indirectly (*e.g.,* by binding of a fluorescently-labeled antibody). Preferred examples of a label that can be utilized by the invention are fluorophores. Several fluorophores can be used as label monomers for labeling nucleotides including, but not limited to, GFP-related proteins, cyanine dyes, fluorescein, rhodamine, ALEXA Flour™, Texas Red, FAM, JOE, TAMRA, and ROX. Several different fluorophores are known, and more continue to be produced, that span the entire spectrum.

[0089] Labels associated with each position (*via* hybridization of a position with a labeled oligonucleotide) are spatially-separable and spectrally-resolvable from the labels of a preceding position or a subsequent position. An ordered series of spatially-separable and spectrally-resolvable labels of a probe is herein referred to as barcode or as a label code. The barcode or label code allows identification of a target nucleic acid or target protein that has been bound by a particular probe.

[0090] The labeled oligonucleotides hybridize to their positions under a standard hybridization reaction, *e.g.,* 65 $^{\circ}$C, 5xSSPE; this allows for self-assembling reporter probes or probes. Probes using longer RNA molecules as labeled oligonucleotide (*e.g.,* as described in US2003/0013091) must be pre-assembled at a manufacturing site rather than by an end user and at higher temperatures to avoid cross-linking of multiple backbones *via* the longer RNA molecules; the pre-assembly steps are followed by purification to remove excess un-hybridized RNA molecules, which increase background. Use of the short single-stranded labeled oligonucleotide (*e.g.*, comprising deoxyribonucleotides) greatly simplifies the manufacturing of the probes and reduces the costs associated with their manufacture.

[0091] In embodiments, probes are provided to a sample at concentrations typically less than that used for immunohistochemistry (IHC) or for in situ hybridization (ISH). Alternately, the concentration may be significantly less than that used for IHC or ISH. For example, the probe concentration may be 2 fold less, 5 fold less, 10 fold less, 20 fold less, 25 fold less, 30 fold less, 50 fold less, 60 fold less, 70 fold less, 80 fold less, 90 fold less, 100 fold less, 200 fold less, 300 fold less, 400 fold less, 500 fold less, 600 fold less, 700 fold less, 800 fold less, 900 fold less, 1000 fold less, 2000 fold less, or less and any number in between. In embodiments, probes are provided at a concentration of 100 nM, 70 nM, 60 nM, 50 nM, 40 nM, 30 nM, 20 nM, 10 nM, 9 nM, 8 nM, 7 nM, 6 nM, 5 nM, 4 nM, 3 nM, 2 nM, 1 nM, 0.9 nM, 0.8 nM, 0.7 nM, 0.6 nM, 0.5 nM, 0.4 nM, 0.3 nM, 0.2 nM, 0.1 nM, 0.09 nM, 0.08 nM, 0.07 nM, 0.06 nM, 0.05 nM, 0.04 nM, 0.03 nM, 0.02 nM, 0.01 nM, and less and any concentration in between.

[0092] Probes can be detected and quantified using commercially-available cartridges, software, systems, *e.g.,* the nCounter® System using the nCounter® Cartridge.

[0093] Background noise, during protein detection, can be reduced by performing a negative purification of the intact probe molecule. This can be done by conducting an affinity purification of the antibody or photo-cleavable linker after collection of eluate from a region of interest. Normally, released signal oligonucleotides will not be pulled out of solution. A protein-G or -O mechanism in a pipet tip, tube, or plate can be employed for this step. Such devices and reagents commercially available.

[0094] Background noise, during nucleic acid detection, can be reduced by performing a negative purification of the intact probe molecule. This can be done by conducting an affinity purification of the target binding domain or photo-cleavable linker after collection of eluate from a region of interest. Normally, released signal oligonucleotides will not be pulled out of solution. To assist in the negative purification, a universal purification sequence may included in a probe, e.g., in the target binding domain.

[0095] Figure 1 shows two exemplary probes including a single-stranded nucleic acid backbone and a target-binding domain, shown in red. The top probe includes labeled RNA segments hybridized to positions in the backbone whereas the bottom probe includes labeled DNA oligonucleotides hybridized to positions in the nucleic acid backbone. The colors shown in Figure 1, and elsewhere in this disclosure, are non-limiting; other colored labels and other detectable labels known in the art can be used in the probes of the present invention.

[0096] Probes of the present invention can be used for detecting a target nucleic acid. Figures 2 and 4 illustrate this aspect. Such a probe includes at least a backbone and a target nucleic acid-binding region. The target nucleic acid-binding region is preferably at least 15 nucleotides in length, and more preferably is at least 20 nucleotides in length. In specific embodiments, the target nucleic acid-binding region is approximately 10 to 500, 20 to 400, 25, 30 to 300, 35, 40 to 200, or 50 to 100 nucleotides in length. Probes and methods for binding and identifying a target nucleic acid have been described in, *e.g.*, US2003/0013091, US2007/0166708, US2010/0015607, US2010/0261026, US2010/0262374, US2010/0112710, US2010/0047924, and US2014/0371088, each of which is incorporated herein by reference in its entirety.
A target nucleic acid may be DNA or RNA and preferably cRNA, messenger RNA (mRNA) or miRNA; preferably the DNA is cDNA.

[0097] The probes of the present invention can be used to directly hybridize to a target nucleic acid. Figure 2 illustrates a probe (or composition) of this embodiment. The probes include a target nucleic-acid binding domain, shown in red. The target nucleic acid is shown as a blue curvilinear line. Figure 3 illustrates a dual probe composition including the probe of Figure 2 and a capture probe. The capture probe comprises at least one affinity reagent, shown as an asterisk. The at least one affinity moiety may be attached to the capture probe by covalent or non-covalent means. Various affinity moieties appropriate for purification and/or for immobilization are known in the art. Preferably, the affinity moiety is biotin, avidin, or streptavidin. Other affinity tags are recognized by specific binding partners and thus facilitate isolation and immobilization by affinity binding to the binding partner, which can be immobilized onto a solid support. In these figures, each probe includes six positions hybridized to labeled oligonucleotides, each positions is identified by a colored circle.

[0098] Any probe of the present invention may comprise an affinity moiety.

[0099] The probes of the present invention can be used to indirectly hybridize to a target nucleic acid present in a sample (*via* an intermediary oligonucleotide). Figure 4 illustrates a probe (or composition) of this embodiment. The probes include a target nucleic-acid binding domain, shown in red, which binds to a synthetic oligonucleotide (the intermediary oligonucleotide; shown in green) that in turn binds to a target nucleic acid in a biological sample. It could be said that the intermediary oligonucleotide is a probe, as defined herein, since it comprises a nucleic acid backbone and is capable of binding a target nucleic acid. The target nucleic acid present in

a biological sample is shown as a blue curvilinear line. Figure 5 illustrates a dual-probe composition including the probe of Figure 4 and a capture probe. In these embodiments, a probe's target nucleic acid-binding region hybridizes to a region of an intermediary oligonucleotide (*i.e.*, a synthetic oligonucleotide) which is different from the target nucleic acid present in a sample. Thus, the probe's target binding region is independent of the ultimate target nucleic acid in the sample. This allows economical and rapid flexibility in an assay design, as the target (present in a sample)-specific components of the assay are included in inexpensive and widely-available synthetic DNA oligonucleotides rather than the more expensive probes. Such synthetic oligonucleotides are simply designed by including a region that hybridizes to the target nucleic acid present in a sample and a region that hybridizes to a probe. Therefore, a single set of indirectly-binding probes can be used to detect an infinite variety of target nucleic acids (present in a sample) in different experiments simply by replacing the target-specific (synthetic) oligonucleotide portion of the assay.

[00100] A probe or probe of the present invention can include a region which permits the release of a signal oligonucleotide following the application of a suitable force. In one non-limited example, the region is a cleavable motif (*e.g.*, a restriction enzyme site or cleavable linker). The cleavable motif allows release of a signal oligonucleotide from a bound target nucleic acid or protein and the signal oligonucleotide is then collected and detected. As used herein a signal oligonucleotide is a region of a probe that presently has positions hybridized with at least one labeled oligonucleotide or is a region of a probe (*e.g.*, a nucleic acid molecule) that can be released from the target-binding domain of the probe. A signal oligonucleotide is said to be releasable when it can be separated (*i.e.*, cleaved and released) from the remainder of the probe. Examples of cleavable motives include but are not limited to photo-cleavable linkers.

[00101] In a probe of the present invention (as described herein), the cleavable motif may be located between a nucleic acid and a target binding domain, the backbone and a target-binding domain, or within the backbone. In Figure 6, non-limiting options for a cleavable motif's position can be inferred from gaps within a probe or a gap within an intermediary oligonucleotide.

[00102] Probes of the present invention can be used for detecting a target protein. Figure 7 illustrates probes (or compositions) of this embodiment. Such probes include at least a backbone and a target protein-binding region. In protein-targeting probes of the present invention, a signal

oligonucleotide may the nucleic acid attached to the protein-binding domain. In these probes, the signal oligonucleotide is targeted and bound by a probe that comprises positions for hybridizing to labeled oligonucleotides. Such a probe is shown in Figure 7, middle image. There, the signal oligonucleotide is seen as a green line. The probe may be bound by a probe before the probe (*via* its protein-binding domain) binds a protein or afterward it binds the protein. The signal oligonucleotide need not be bound by the probe until it has already been released from the target-binding domain (this embodiment is not shown).

[00103] A probe's region capable of binding to a target protein include molecules or assemblies that are designed to bind with at least one target protein, at least one target protein surrogate, or both and can, under appropriate conditions, form a molecular complex comprising the protein probe and the target protein. The region capable of binding to a target protein includes an antibody, a peptide, an aptamer, or a peptoid. The antibody can be obtained from a variety of sources, including but not limited to polyclonal antibody, monoclonal antibody, monospecific antibody, recombinantly expressed antibody, humanized antibody, plantibodies, and the like. The terms protein, polypeptide, peptide, and amino acid sequence are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids or synthetic amino acids. The terms also encompass an amino acid polymer that has been modified, for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term amino acid refers to either natural and/or unnatural or synthetic amino acids, including but not limited to glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. Probes and methods for binding and identifying a target protein have been described, *e.g.*, in US2011/0086774, the contents of which is incorporated herein by reference in its entirety.

[00104] In embodiments, a probe is prepared by a cysteine bioconjugation method that is stable, site-specific to, preferably, the antibody's hinge-region heavy-chain. This preparation method provides relatively controllable labeled oligonucleotides to antibody stoichiometric ratios. A probe can comprise a plurality (i.e., more than one, e.g., 2, 3, 4, 5, or more) labeled oligonucleotides per antibody. Generally, "heavier" probes, which comprise 3 or 4 labeled

oligonucleotides per antibody, are significantly less sensitive than antibodies lacking a labeled oligonucleotide or "lighter" probes, which comprise 1 or 2 labeled oligonucleotides per antibody.

[00105] Protein-targeting probes and nucleic acid-targeting probes may be applied simultaneously as long as conditions allow for binding of both a protein target and a nucleic acid target. Alternately, protein-targeting probes and nucleic acid-targeting probes may be applied sequentially when conditions allowing for binding of both a protein target and a nucleic acid target are not possible.

[00106] A set of probes is synonymous with a composition of probes. A set of probes includes at least one species of probes, *i.e.*, directed to one target. A set of probes preferably includes at least two, e.g., 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more species of probes. A probe set may include one or multiple copies of each species of probe.

[00107] A first set of probes only may be applied to a sample. Alternately, a second set (or higher number) of probes may be later applied to the sample. The first set and second (or higher number) may target only nucleic acids, only proteins, or a combination thereof.

[00108] In the present invention, two or more targets (*i.e.*, proteins, nucleic acids, or a combination thereof) are detected; 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more targets, and any number there between, are detected.

[00109] A set of probes may be pre-defined based upon the cell type or tissue type to be targeted. For example, if the tissue is a breast cancer, then the set of probes will include probes directed to proteins relevant to breast cancer cells (e.g., Her2, EGFR, and PR) and/or probes directed to proteins relevant to normal breast tissues. Additionally, the set of probes may be pre-defined based upon developmental status of a cell or tissue to be targeted. Alternately, the set of probes may be pre-defined based upon subcellular localizations of interest, e.g., nucleus, cytoplasm, and membrane. For example, antibodies directed to Foxp3, Histone H3, or P-S6 label the nucleus, antibodies directed to CD3, CD4, PD-1, or CD45RO label the cytoplasm, and antibodies directed to PD-L1 label membranes.

[00110] A probe may be chemically synthesized or may be produced biologically using a vector into which a nucleic acid encoding the probe has been cloned.

[00111] Any probe or set of probes described herein may be used in methods and kits of the present invention.

[00112] For the herein-described probes, association of label code to target nucleic acid or target protein is not fixed.

[00113] Probes of the present invention can be used to detect a target nucleic acid or protein present in any sample, *e.g.*, a biological sample. As will be appreciated by those in the art, the sample may comprise any number of things, including, but not limited to: cells (including both primary cells and cultured cell lines) and tissues (including cultured or explanted). In embodiments, a tissue sample (fixed or unfixed) is embedded, serially sectioned, and immobilized onto a microscope slide. As is well known, a pair of serial sections will include at least one cell that is present in both serial sections. Structures and cell types, located on a first serial section will have a similar location on an adjacent serial section. The sample can be cultured cells or dissociated cells (fixed or unfixed) that have been immobilized onto a slide.

[00114] In embodiments, a tissue sample is a biopsied tumor or a portion thereof, i.e., a clinically-relevant tissue sample. For example, the tumor may be from a breast cancer. The sample may be an excised lymph node.

[00115] The sample can be obtained from virtually any organism including multicellular organisms, *e.g.*, of the plant, fungus, and animal kingdoms; preferably, the sample is obtained from an animal, *e.g.*, a mammal. Human samples are particularly preferred.

[00116] In some embodiments, the probes, compositions, methods, and kits described herein are used in the diagnosis of a condition. As used herein the term diagnose or diagnosis of a condition includes predicting or diagnosing the condition, determining predisposition to the condition, monitoring treatment of the condition, diagnosing a therapeutic response of the disease, and prognosis of the condition, condition progression, and response to particular treatment of the condition. For example, a tissue sample can be assayed according to any of the probes, methods, or kits described herein to determine the presence and/or quantity of markers of a disease or malignant cell type in the sample (relative to the non-diseased condition), thereby diagnosing or staging a disease or a cancer.

[00117] In general, samples attached to a slide can be first imaged using fluorescence (e.g., fluorescent antibodies or fluorescent stains (e.g., DAPI)) to identify morphology, regions of

interest, cell types of interest, and single cells and then expression of proteins and/or nucleic acids can be digitally counted from the sample on the same slide.

[00118] Compositions and kits of the present invention can include probes and other reagents, for example, buffers and other reagents known in the art to facilitate binding of a protein and/or a nucleic acid in a sample, *i.e.*, for performing hybridization reactions.

[00119] A kit also will include instructions for using the components of the kit, including, but not limited to, information necessary to hybridize labeled oligonucleotides to a probe, to hybridize a probe to a target-specific oligonucleotide, to hybridize a target-specific oligonucleotide to a target nucleic acid and/or to hybridize a probe to target protein.

[00120] An exemplary protocol for detecting target nucleic acids and/or target proteins is described as follows and as shown in **Figures 10** to **14** (top).

[00121] Cells (live or fixed) or tissue sections (*e.g.,* formalin-fixed paraffin embedded (FFPE)) that are prepared consistent with multiplexed immunohistochemistry methods and/or nucleic acid *in situ* hybridization methods are prepared and immobilize onto a glass slide or suitable solid support. Access to the surface of cells or tissue-section is preserved, allowing for fluidic exchange; this can be achieved by using a fluidic chamber reagent exchange system (*e.g.*, Grace™ Bio-Labs, Bend OR). Regions-of-interest (ROIs) are identified on the serial section to be provided probes or on an adjacent serial section. In the first instance, full "macroscopic-features" imaging methodology to cell/tissues of interest is performed, *e.g.,* DAPI staining, membrane staining, mitochondrial staining, specific epitope staining, and specific transcript staining, to determine overall macroscopic features of cell/tissue of interest. Alternately, regions-of-interest (ROIs) are identified on a serial section adjacent to the serial section to be provided the probes; here, full "macroscopic-features" imaging (as described above) is performed on a first serial section (section #1 in Figures 11 and 12). This imaging will generally identify regions-of-interest on the adjacent serial section (red line in panel B in Figure 10 and green oval and green triangle of section #2 in Figures 11 and 12) where signal oligonucleotides will be released from the probes upon application of a suitable and directed force. Serial sections may be approximately 5μm to 15μm from each other.

[00122] **Figure 13** and **Figure 14** (top) further illustrate steps of the present invention. Steps shown in **Figure 13** include the following. (1) Process: FFPE slide mounted tissue is incubated with a cocktail of primary antibodies conjugated to DNA oligos via a photo-cleavable linker,

together with a limited number of visible-wavelength imaging reagents. (2) View: Regions of interest (ROI) are identified with visible-light based imaging reagents at low-plex to establish overall "architecture" of tumor slice (e.g., image nuclei and/or using one or two key tumor biomarkers). (3) Profile: Select ROIs are chosen for high-resolution multiplex profiling and oligos from the selected region are released following exposure to UV light. (4) Plating: Free photocleaved oligos are then collected, e.g., via a microcapillary-based "sipper", and stored in a microplate well for subsequent quantitation. (5) Digitally Count: During the digital counting step, photocleaved oligos from the spatially resolved ROIs in the microplate are hybridized to 4-color, 6-spot optical barcodes, enabling up to ~ 1 million digital counts of the protein targets (distributed over up to 800-plex markers) in a single ROI using standard NanoString nCounter® read-out instrument (e.g., SPRINT, Flex, and MAX).

[00123] A region of interest may be a tissue type present in a sample, a cell type, a cell, or a subcellular structure within a cell.

[00124] A composition comprising a set of probes, each probe comprising a releasable signal oligonucleotide, is applied to the serial section. The set of probes or may include probes that target proteins, target nucleic acids, or both. The composition may include capture probes. When probes indirectly bind to a target (protein and/or nucleic acid), the applied composition includes intermediary oligonucleotides. The composition will include other reagents known in the art to facilitate binding of a protein and/or a nucleic acid in a sample.

[00125] Blocking steps are performed before and/or after the composition is applied.

[00126] For probes including photo-cleavable linkers, the solid support (*e.g.*, microscope slide) is placed in a microscope that is capable of providing excitation light at a wavelength capable of cleaving the photo-cleavable linker. A first region-of-interest (red line in panel B in Figure 10 and ROI_i in Figures 11 and 12) is excited with the light, thereby cleaving the photo-cleavable linker and releasing the signal oligonucleotides. As illustrated in Figures 6 and 9, a signal oligonucleotide includes at least a region of a probe that presently has positions bound with at least one labeled oligonucleotide or the nucleic acid from a probe that is bound or can be bound by a reporter probe. By directing excitation light only to ROIi, signal oligonucleotides are only released from probes within ROIi and not from probes located outside of ROIi, which retain their signal oligonucleotides. Thus, signal oligonucleotides are collected only for probes that are

bound to targets within ROIi, thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located within ROIi.

[00127] The surface of the section is washed with small amount of buffer (~5 to 30 µl) and the eluate (containing the released signal oligonucleotides) is collected into a first sample container (shown as Sample "i" in Figure 12). The surface of the section is further rinsed to remove any released signal oligonucleotides that were omitted from the eluate.

[00128] A second region-of-interest (ROI_j in Figures 11 and 12) is excited with light, thereby cleaving the photo-cleavable linker and releasing the signal oligonucleotides from the second region-of-interest. Again, by directing excitation light only to ROIj, signal oligonucleotides are only released from probes within ROIj and not from probes located outside of ROIj, which retain their signal oligonucleotides. Thus, signal oligonucleotides are collected only for probes that are bound to targets within ROIj, thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located within ROIj.

[00129] The surface of the section is washed with small amount of buffer (~5 to 30 µl) and the eluate (containing the released signal oligonucleotides) is collected into a first sample tube (shown as Sample "j" in Figure 12). The surface of the section is further rinsed to remove any released signal oligonucleotides that were omitted from the eluate.

[00130] The excitation step, washing step, and rinsing step are repeated until signal oligonucleotides from all regions-of-interest (up to ROI_n) have been collected.

[00131] Additional advantages, features, and embodiments of the present invention are illustrated in the **Appendix** filed herewith. As examples, various methods and devices for collecting a signal oligonucleotide and various ways of providing a force are shown. Moreover, the **Appendix** provides unexpectedly improved results obtained from certain embodiments of the present invention over other embodiments. Data demonstrating about 7-fold to about 200-fold signal-to-noise improvements are shown.

[00132] Detection can use any microscope-type device or system known in the art. A device or system may include wide field illumination along with a digital mirror device (DMD; see **Figures 15** and **16**); advantages of this include reduced costs since the DMD and controller can also drive the LED (which photocleaves probes) and adds essentially no additional cost, provides ease of implementation, allows small feature size of ~1 mm which will include 10-40 mm cells, and leverages available consumer electronics (like projectors). A device or system may include

a laser scanning device, e.g., confocal, see **Figure 16**. An advantage of this is smaller morphological features can be illuminated and imaged; however, additional costs are involved with these devices.

[00133] The plurality of target proteins and/or target nucleic acids present in each region of interest in a sample are identified in each eluate sample using a polymerase reaction, a reverse transcriptase reaction, hybridization to an oligonucleotide microarray, mass spectrometry, hybridization to a fluorescent molecular beacon, a sequencing reaction, or nCounter® Molecular Barcodes. nCounter® systems and methods from NanoString Technologies®, as described in US2003/0013091, US2007/0166708, US2010/0015607, US2010/0261026, US2010/0262374, US2010/0112710, US2010/0047924, US2014/0371088, and US2011/0086774), are a preferred means for identifying target proteins and/or target nucleic acids. nCounter® systems and methods from NanoString Technologies® allow simultaneous multiplexed identification a plurality (800 or more) distinct target proteins and/or target nucleic acids.

[00134] Together, a comparison of the identity and abundance of the target proteins and/or target nucleic acids present in first region of interest (*e.g.*, tissue type, a cell type (including normal and abnormal cells), and a subcellular structure within a cell) and the identity and abundance of the target proteins and/or target nucleic acids present in second region of interest or more regions of interest can be made.

[00135] The present invention provides multiplexed detection and comparison of up to 800 proteins of interest from discrete regions within a tumor (for example) and its adjacent normal tissue; thus, enabling systematic interrogation of the tumor and its microenvironment.

[00136] The present invention can be used in ongoing clinical studies to elucidate novel responses to immunotherapies and other targeted therapies.

[00137] The present invention also enables the discovery of immune biomarkers in tumors (for example) which can be used in the development of companion diagnostics.

[00138] Immunohistochemistry is a powerful technique for analyzing protein expression and localization in FFPE tissue sections. However, it suffers from a number of challenges, including a lack of dynamic range, difficult quantitation, and labor intensive workflow for very limited multiplexing. Here is disclosed a novel platform based on the nCounter® barcoding technology which enables spatially-resolved, digital characterization of proteins in a highly multiplexed (up

to 800-plex) assay, i.e., the nCounter® Digital Multiplexed Immunohistochemistry (IHC) assay. The assay relies upon antibodies coupled to photo-cleavable oligonucleotide tags which are released from discrete regions of a tissue using focused through-objective UV (e.g., ~ 365nm) exposure. Cleaved tags are quantitated in an nCounter® assay and counts mapped back to tissue location, yielding a spatially-resolved digital profile of protein abundance. The protein-detection may be performed along with or separate from a nucleic acid-detection assay which uses nucleic acid probes comprising photo-cleavable oligonucleotide tags. Thus, the present invention can provide spatially-resolved digital profile of protein abundance, spatially-resolved digital profile of protein and nucleic acid abundance, or spatially-resolved digital profile of nucleic acid abundance.

[00139] Advantages of the assay include, but are not limited to: high sensitivity (e.g., ~ 1 to 4 cells), all digital counting, with large dynamic range (> 10^5), highly multiplexed (e.g., 30 targets and scalable, with no change in instrumentation, to 800 targets), simple workflow, compatibility with FFPE, no secondary antibodies (for protein detection) or amplification reagents, and potential for clinical assays.

[00140] As used in this Specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

[00141] Unless specifically stated or obvious from context, as used herein, the term "or" is understood to be inclusive and covers both "or" and "and".

[00142] Unless specifically stated or obvious from context, as used herein, the term "about" is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from the context, all numerical values provided herein are modified by the term "about."

[00143] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although other probes, compositions, methods, and kits similar, or equivalent, to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein. It is to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

EXAMPLES

[00144] Example 1: The present invention provides a "barcoding-potential" to quantify multiplexed targets in a FFPE tissue section.

[00145] Intratumoral heterogeneity has emerged as a critical challenge to the implementation of targeted therapeutics. Historically, immunohistochemistry (IHC) has been used to assess spatial heterogeneity of proteins; however, it has been difficult to quantify protein abundance at high multiplex and wide dynamic range.

[00146] In this example, proteins in a formalin-fixed paraffin embedded (FFPE) tissue section were labeled with antibody-comprising probes that included photo-cleavable linkers and fluorescent barcodes. The probes—in a user-defined ROI of the FFPE tissue section—were subsequently exposed to focused UV light, thereby releasing the signal oligonucleotides (comprising the fluorescent barcodes) from the ROI. The released signal oligonucleotides were washed away from the FFPE sample and collected. The fluorescent barcodes from the released signal oligonucleotides were then recognized and digitally counted by an nCounter® system from NanoString Technologies®, thereby quantifying the abundance of each targeted protein in the user-defined spatial region of a tissue section. After the signal oligonucleotides from a first ROI were released and collected, the focused UV light was exposed to a second user-defined ROI of the FFPE tissue section, thereby releasing the signal oligonucleotides from the second ROI. In this non-limiting example, a high degree of linearity ($0.97 < R^2 < 0.99$) for the number of observed counts versus area of UV illumination was observed and with a detection spatial resolution of about 100 μm x 100 μm, or approximately 100 cells. Unexpectedly, the present invention provides a "barcoding-potential" to quantify up to 800 targets with 5.5 logarithms (base 10) of dynamic range in a single FFPE tissue section.

[00147] Example 2: The present invention provides a practical and feasible approach for quantifying protein expression without signal amplification and for achieving higher-order target antigen multiplexing in a FFPE tissue section.

[00148] Quantitative, multiplexed immunohistochemistry has emerged as an area of great interest within oncology since it has the unique capability of identifying spatiotemporal organization and

interdependencies that further define how checkpoint blockade impacts tumor microenvironment. This example describes a one-step, amplification-free staining method using a photo-cleavable oligo-tagged primary antibody which interacts with the target antigen within an FFPE tissue section. Illumination with ultraviolet (UV) light is applied which releases the oligo from the antibody and is followed by eluent collection, quantification, and digital counting that corresponds to antigen abundance.

[00149] First was investigated a variety of conjugation methods; this established a cysteine bioconjugation method that is stable, site-specific to predominantly the hinge-region heavy-chain, and relatively controllable in terms of oligonucleotide to antibody stoichiometric ratios.

[00150] Next was performed a linear regression analysis to determine the relationship between UV-induced cleavage area and measured digital protein counts; from this was observed a high degree of linearity ($0.97 < R^2 < 0.99$), confirming the basic mechanism/premise associated with this multiplexed protein counting method on FFPE tissue.

[00151] To determine the impact of the presence of a conjugated oligonucleotide on antibody-antigen interaction, the performance of a labeled oligonucleotides-conjugated antibody to the unmodified antibody under identical conditions in FFPE tissue sections was compared in terms of sensitivity, specificity and signal intensity. Antibodies were selected that targeted antigens localized to the nucleus, cytoplasm, or membrane to determine the relationship between antibody performance and subcellular location of target antigens. Selected antibodies targeted Foxp3, Histone H3, P-S6 (nuclear antigens), CD3, CD4, PD-1, CD45RO (cytoplasmic antigens), and PD-L1 (membranous antigen). In terms of sensitivity, generally, "heavier" oligonucleotide-conjugated antibodies (having 3 or 4 labeled oligonucleotide per antibody) were found to be significantly less sensitive when compared to unconjugated antibodies or "lighter" oligonucleotide-conjugated antibodies (having 1 or 2 labeled oligonucleotide per antibody). No significant difference was observed between unconjugated or "lighter" oligonucleotide-conjugated antibodies in terms of sensitivity, specificity, or intensity across nuclear, cytoplasmic and membranous target antigens.

[00152] The present invention provides highly multiplexed protein profiling that measures absolute protein expression levels using practical and feasible methods to comprehensively define the immune landscape in tumors before and during immunotherapeutic intervention.

[00153] Example 3: The present invention provides spatially-resolved, multiplexed protein detection from FFPE tissue

[00154] Methods

[00155] Antibodies – Antibodies used in this Example and Examples 4 to 6 may include: "target (clone ID, vendor))": H3 (D1H2, CST), CD8 (OTI3H6, Origene), CD4 (SP35, Spring Bio), FOXP3 (D2W8E, CST), B7-H3 (D9M2L, CST), S6 (54D2, CST), B7-H4 (D1M8I, CST), Granzyme B (OTI4E4, Origene), Ki67 (8D5, CST), PD-1 (Nat105, Cell Marque), CD3 (MRQ-39, Cell Marque), Vista (D1L2G, CST), Her2 (29D8, CST), PR (D8Q2J, CST), ER (SP1, Spring Bio), EGFR (D38B1, CST), CD56 (MRQ-42, Cell Marque), PD-L1 (E1L3N, CST), CD45 (2B11&PD7/26, Cell Marque), TIM-3 (D5D5R, CST), and Pan Keratin (C11, CST), CD45RO (UCHL1, Cell Marque).

[00156] Tonsil Microscopy – 5 μm sections of a tonsil FFPE block (Amsbio) were mounted on slides. IHC was performed using standard protocols. Antigen retrieval was performed with a pressure cooker. Staining of the tonsil section was performed with CD3 primary antibody MRQ-39 (Rabbit mAb, Cell Marque) and Ki-67 primary antibody 8D5 (Mouse mAb, CST). Secondary incubations were performed with Alexa594 labeled Goat α Rabbit (Life Tech.) and Alexa488 labeled Goat α Mouse (Life Tech.)

[00157] Here, samples attached to a slide were first imaged using fluorescent antibodies and then expression of proteins was digitally counted from the sample.

[00158] Steps similar to those illustrated in **Figure 10 to Figure 14** (top) were used. UV-cleavage of selected ROIs allowed full 30-plex digital profiling (nCounter® counts).

[00159] Results

[00160] Figure 18 shows a photomicrograph establishing overall tissue morphology of a tonsil sample that was initially imaged using 2-color fluorescence of Ki-67 (cell proliferation marker; in green) and CD3 (immune cell marker; in red). Multiple targets analyzed across twelve regions (including the four regions magnified in **Figure 19**) show three distinct profiles of Ki-67 and CD3 localization. **Figure 19** shows nCounter® counts for Ki-67 and CD3 for four regions shown in **Figure 18**. **Figure 20** shows exemplary counts from a 30-plex oligo-antibody cocktail on the twelve regions of interest (ROI) from the tonsil sample shown in **Figure 18**. Data was obtained from serial sections (to allow various additional controls to be examined). As shown, regions of the tissue sample can be classified based on the intensity and identity of the markers expressed.

Exemplary classifications shown: "CD3-enriched", "Ki67-enriched", "Mixed", and "Connective tissue".

[00161] These data show that the present invention provides spatially-resolved detection of a plurality (here, at least 30) of protein markers. By scaling up the number of protein probes (antibodies) used, up to 800 different protein markers can be detected and with similar resolution.

[00162] Example 4: The present invention provides multiplexed protein detection from FFPE tissue and approaching single-cell resolution

[00163] Methods

[00164] Melanoma Microscopy – 5 μm sections of a melanoma (lymph node derived) FFPE block (Asterand) were mounted on slides. IHC was performed using standard protocols. Antigen retrieval was performed with a pressure cooker.

[00165] Here, samples were first imaged using fluorescence and then expression of proteins was digitally counted from the sample.

[00166] Steps similar to those illustrated in **Figure 10** to **Figure 14** (top) were used.

[00167] Results

[00168] **Figure 21** shows a photomicrograph establishing overall tissue morphology of T cells in a melanoma sample of lymph node that was initially imaged using three-color fluorescence of CD3 (in red), CD8 (in green), and DAPI (in blue). The white circle is 25 μm in diameter and surrounds three cells.

[00169] **Figure 22** shows nCounter® data of CD3 conjugate release from FFPE lymph node tissue section (5 μm thickness) as a function of UV illumination area (100 μm to 1 mm in diameter). The limit of detection counts (LOD = background counts + 2x standard deviation) corresponds to spatial resolution of 26 μm in diameter. The field diaphragm size is shown below the figure. **Figure 23** and **Figure 24** show data for CD45 and PD1 (respectively, from the same experiment).

[00170] The data shows a spatial detection ability of the present invention corresponding to about one to four cells.

[00171] Example 5: The present invention provides quantitative performance in a clinically-relevant assay

[00172] Method

[00173] Steps similar to those illustrated in **Figure 10** to **Figure 14** (top) were used.

[00174] Breast Cancer tissue microarray (TMA): TMA BR1504a obtained from US Biomax, Inc., H&E staining image obtained from US Biomax website (World Wide Web (*www*) biomax.us/tissue-arrays/Breast/BR1504a). Section from the same block as the section shown on in the left panel of **Figure 25** were stained with Her2 primary antibody 29D8 (Rabbit mAb, CST), and Alexa594 labeled Goat α Rabbit (Life Tech.). Counts were also obtained for Histone H3, Ribsomal Protein S6, Estrogen Receptor, Progesterone Receptor, Mouse IgG isotype control, and Rabbit IgG isotype control (data not shown). Her2 pathologist scores for TMA BR1504a were provided by US Biomax, Inc. (World Wide Web (*www*) biomax.us/tissue-arrays/Breast/BR1504a). Staining was performed with Her2 primary antibody 29D8 (Rabbit mAb, CST), and Alexa594 labeled Goat α Rabbit (Life Tech.). Although other Rabbit primary antibodies were used in the primary cocktail, fluorescence from these antibodies was negligible compared to Her2 fluorescence. Sum Pixel Intensities (at $\lambda = 594$) were obtained using ImageJ software. For this, the background value was set to intensity = 0 and the highest intensity was set to intensity = 255. The summation of all pixel intensities per ROI is shown.

[00175] Here, samples were first imaged using fluorescence and then expression of proteins was digitally counted from the sample.

[00176] Results

[00177] Figure 25 (left panel) shows a tissue microarray (TMA) of breast tumor tissue containing variable levels of Her2 protein as shown in the photomicrograph (center panel) which identifies Her2 fluorescence by IHC staining. The right panel shows a magnification of a single region of the central panel; such regions were stained with a multiplexed antibody cocktail.

[00178] Figure 26 shows nCounter® count data for forty-eight representative regions versus Her2 status (ASCO-CAP guidelines). **Figure 27** plots nCounter® Counts versus Sum Pixel Intensities ($x10^3$) for the forty-eight regions mentioned above.

[00179] These digital count data show a high correlation with fluorescence intensities ($R2 = 0.92$, **Figure 27**) compared to visual Her2 status scoring via ASCO-CAP guidelines ($R2 = 0.51$, **Figure 26**).

[00180] Example 6: The present invention reveals abundances of specific cell types in a tissue sample

[00181] Steps similar to those illustrated in **Figure 10** to **Figure 14** (top) were used; a melanoma sample attached to a slide was first imaged using fluorescence and then expression of proteins was digitally counted from the sample.

[00182] Figure 28 shows a photomicrograph establishing overall tissue morphology of a melanoma sample using two-color fluorescence of CD3 (immune cell marker; in red) and DAPI (cell nuclei, in blue). Expression data using a 30 antibody cocktail was obtained from the ten regions identified with white boxes. **Figure 29** shows exemplary nCounter® counts from a 30-plex oligo-antibody cocktail on the ten regions of interest (ROI) from the melanoma sample shown in **Figure 28**. Counts for thirteen markers, each having expression counts above background, are shown. Regions 5, 6, 7, identified as "Immune infiltrate-enriched" have the highest expression of T-cell markers and T-cell regulatory markers.

[00183] These data show that the present invention provides spatially-resolved detection of a plurality (here, at least 30) of protein markers. By scaling up the number of protein probes (antibodies) used, up to 800 different protein markers can be detected and with similar resolution.

[00184] Example 7: A digital mirror device (DMD) is capable of illuminating single cells

[00185] Figures 30 and **31** are photomicrographs showing that UV illumination using a digital mirror device (DMD) is capable if illuminating single cells in a tonsil tissue sample.

[00186] These data show that the present invention is capable of single cell resolution when using a DMD.

[00187] Example 8: A gel box is capable of illuminating an entire sample and releasing signal oligonucleotides from probes bound to the entire sample

[00188] Figure 34: Shows an embodiment in which a whole tissue or sample is illuminated, e.g., with a standard laboratory UV gel box. Here, a FFPE tissue slide was placed on the light panel, a wax pen was used to hold buffer solution (TBS) covering the FFPE tissue, and UV light

exposure (276 - 362nm, e.g., 302nm; ~5mW/cm^2) was applied to the tissue through the glass slide (1 mm thickness). The data shows that within about one minute of UV exposure, most of signal oligonucleotides are released from FFPE bound antibodies. Counts are normalized to a positive control.

[00189] Example 9: Illumination from a microscope is capable of illuminating a region of interest in a sample and releasing signal oligonucleotides from probes bound to the region of interest

[00190] Figure 35 shows an embodiment in which a portion of a tissue or sample is illuminated, e.g., with a microscope, i.e., UV cleavage under Microscope (Time titration experiment). This is in contrast to the experiment of **Example 8**, in which a whole sample is illuminated. Here, UV LED (at 365nm) is applied at about ~150mW/cm^2 with a 20x objective. UV illumination scans the whole tissue area identified by previous fluorescence (~590nm excitation) bright field imaging. Within about one second of UV exposure per field of view (FOV), most signal oligonucleotides are released from FFPE bound probes. The gel box experiment of **Example 8** was utilized as a non-spatially resolved 100% release control. Counts are normalized ratio to positive control. **Blue**: microscope data with variable lengths of exposure time; **Red**: Gel box 2.5 minutes exposure data. Also shown are photographs and a schematic showing configuration of the microscope apparatus.

[00191] Figure 36 shows signal oligonucleotides are released from a uniformly distributed anti-Histone(H3) antibody bound to lung tissue sample. Tissue was exposed to one second of UV (365nm, ~150mW/cm^2 with 20x objective) per field-of-view (FOV) of about 450µm x 330µm = 0.15 mm^2. "Macro-Volume" used to collect effluent was about 70 µl. This decreases the limit of detection to (FOV / 5) ~ 99µm X 99µm with collection effluent of about 5µl. Hence, in this example, the limit of detection is approximately 10 cells X 10 cells niche. These data show that antibody signal is proportional to spatially resolved illumination area FOV and estimate "macro-fluidics" limit-of-detection (LOD).

[00192] Figure 37 shows an embodiment in which a portion of a tissue or sample is illuminated, e.g., with a microscope, i.e., UV cleavage under microscope (Illumination area titration experiment) and for multiple targets. Shown is UV cleavage of multiple targets in a tissue: two

positive targets (Histone H3 and Ribosomal S6) and eight 8-negative targets. Only one negative target (Ox40), showed high background. Data from zero, one, four, nine and sixteen fields of view are shown.

[00193] Example 10: A region of interest may be pre-identified by a labeling technique and then the region of interest is illuminated signal oligonucleotides are released from probes bound to the pre-identified region of interest

[00194] Figure 38 shows an embodiment in which a region of interest in a tissue (e.g., a breast cancer sample) is first identified for expression of a marker (here Her2) and this region of interest is then illuminated (e.g., with UV) to release signal oligonucleotides from a bound probe. Data shown compares the amount of signal oligonucleotides, for two targets (here, Her2 and Histone H3), released from two locations: one region of interest that was pre-identified as Her2+ and one that was pre-identified as Her2-.

[00195] Example 11: A sample embedded in a flow cell provides collection of elution from the entire sample and not only from a region of interest that is illuminated and from which signal oligonucleotides are released

[00196] Figure 39 shows an embodiment in which a tissue is embedded in flow cell. Here, FFPE Tissue embedded in microfluidic flow cell (a 9mm circular chamber with volume of 100 μm height with an approximate 25μl volume [when the flow cell has a 300μm height the approximate volume is 75μl]) controlled by a syringe pump. UV cleavage inside flow cell, showing elution profile illumination one area (9 FOVs) and elution, then illumination another area (9 FOVs) and elution. Data for multiple fractions is shown. As with the data of **Example 10,** here a region of interest was pre-identified for expression of a fluorescently-labeled marker.

[00197] Example 12: A sample embedded in a flow cell comprising small holes over a region of interest provides efficient collection of elution from the region of interest that is illuminated and where signal oligonucleotides are released and not from the entire sample

[00198] Figure 40 shows an embodiment in which a tissue is embedded in flow cell with small holes. Here, elution occurs directly above the region of interest. 0.4-1mm diameter holes above fluidic chamber allow collection of eluent (e.g., 5μl collection volume). Tested were 9-hole, 96-

hole format, and 12-hole format (for tissue microarray (TMA)). The fluorescence image was created by combining multiple fields of view. Also shown are photographs and a schematic showing configuration of the apparatus.

[00199] **Figures 41A** to **41C** shows that embodiments using a flow cell with small holes have significant signal to noise improvement rather than collection of eluate from entire surface of tissue. The data shows that collecting eluent through a hole above a region of interest increases signal-to-noise by about 7 fold. In this embodiment, 1 mm diameter holes above fluidic flow cell (25µl chamber) were used to collect eluent (5µl fractions). Data for multiple fractions is shown.

[00200] **Figures 42A** to **42C** shows data in the using a flow cell with small holes (12 or 96 hole formats). The data shows that collecting eluent through a hole above a region of interest increases signal-to-noise by about 7 fold. In this embodiment, field of view illumination was focused at the center of a hole; 5µl volume of elution per hole.

[00201] **Figures 43A** and **B** shows data in comparing background signal from flow cells in which whole tissue elution was performed (**Figure 43A**; as in **Example 11**) and background signal from flow cells in which elution occurred directly above a region of interest (**Figure 43B**). As seen in **Figure 43A**, there is higher background for the whole tissue elution relative to the background seen in the **Figure 43B**. Additionally, **Figure 43B** shows no difference between in-flow cell and non-flow cell incubation.

[00202] **Example 13: Released signal oligonucleotides can be elected via a single tube/pipet, a plurality of tubes/pipets, or a multi-tube/pipet array**

[00203] **Figure 44** is a schematic showing eluent collection with an open surface for a multi-region of interest aspiration embodiment. Here is shown a multi-tube array for aspiration/dispensing eluents with rotary valve selection. See also, **Figure 47**.

[00204] **Figure 45** includes photographs and a schematic showing an embodiment in which eluent collection is through a capillary (micro-aspirator). See also, **Figure 47**. **Figures 46A** and **B** shows data from the embodiment of **Figure 45** in which eluent collection is through a capillary (micro-aspirator). This embodiment has a dramatic improvement in signal to noise: signal to noise ratio increases about 10 fold, compared to flow cell through hole elution and signal to noise ratio increases about 200 fold, compared to whole tissue elution. Here, the LOD area is approximately 60 µm x 60 µm.

[00205] Example 14: A device comprising both illuminating and elution capabilities can efficiently and accurately obtain nucleic acid and/or protein expression data from a defined region of interest

[00206] Figure 48 is a schematic showing illumination and fluid collection through a combined capillary and lens.

[00207] Example 15: Protein expression can be detected and quantified from a single cell

[00208] Figure 50 shows protein expression data obtained from a single cell or two cells using the herein described methods and apparatuses. In the top panel, S6 protein is detected and quantified from at least one cell and in the bottom panel, CD45 protein is detected and quantified from at least one cell.

[00209] Example 16: The herein described methods and apparatuses provide an accurate and efficient detection and quantification of spatially-resolved, multiplexed RNA target and/or protein target expression

[00210] In situ hybridization (ISH) was performed to hybridize DNA oligo-based probes ("RNA probes"), each comprising a target-binding domain, a signal oligonucleotide, and a photo-cleavable linker, to an endogenous RNA. 5 μm FFPE HER2 3+ breast tissue sections were deparaffinized in xylene, partially rehydrated in graded ethanols, and incubated in 70% ethanol for 1 hour at room temperature. Then sections were incubated in 40μg/ml proteinase K for 25 minutes at 37°C. Tissues were then incubated in 50% formamide/2X SSC for 15 minutes at room temperature and hybridized overnight at 37C in a solution of 1nM probes, 40% formamide, 1mg/ml yeast tRNA, 10% dextran sulfate, and 0.2% BSA in 2X SSC. After hybridization, two stringent washes in 50% formamide/2X SSC were performed for 25 minutes each at 37°C. Sections were stained with TO-PRO®-3 (Thermo Fisher Scientific) fluorescent nucleic acid stain to visualize tissue morphology. Focused UV light, directed by a digital micromirror device, was then used to cleave DNA signal oligonucleotides from probes in a user-defined region of interest (ROI). For each tissue section, two ROIs comprised a tumorous tissue, two ROIs comprised normal tissue, and two ROIs comprised no tissue at all (histology slide itself). After cleavage, signal oligonucleotides were collected, hybridized to nCounter® Molecular Barcodes, and

digitally counted by an nCounter® system from NanoString Technologies®. H&E was performed on tissue sections to verify tumorous and normal tissue ROIs.

[00211] On serial sections, standard immunohistochemistry (IHC) was performed using "Protein probes," each comprising an antibody as target-binding domain, a DNA signal oligonucleotide, and a photo-cleavable linker. Sections were then stained with an anti-rabbit Alexa 594 secondary antibody and TO-PRO®-3 (Thermo Fisher Scientific) fluorescent nucleic acid stain to visualize tissue morphology. Focused UV light, directed by a digital micromirror device (DMD), was then used to cleave DNA signal oligonucleotides from probes in a user-defined ROI. For each tissue section, two ROIs comprised tumorous tissue, one ROI comprised normal tissue, and two ROIs comprised no tissue at all (histology slide itself). ROIs were matched to the ROIs selected for ISH probe cleavage. Following cleavage, the signal oligonucleotides from protein targets were mixed with the signal oligonucleotides from RNA targets and all were quantitated as described above. H&E was performed on tissue sections to verify tumorous and normal tissue ROIs and to verify ROIs were correctly matched between ISH and IHC tissues.

[00212] **Figure 51** shows ROIs sampled from serial sections of the same tumor sample. Regions 1-4 are not shown in this image and, instead, were taken from portions of the tissue that did not contain tissue (negative controls – "No Tissue"). Regions 5-8 contained low numbers of tumor cells ("Normal Tissue"). Regions 9-12 contained high numbers of tumor cells ("Tumor").

[00213] **Figure 52** shows counts obtained for six of the nine RNA probes included in this assay. For each ROI, a sample was collected prior to applying the UV illumination (the "-UV" set of data) and prior to collecting a plus UV sample from the same region (the "+UV" set of data). Background levels of counts are obtained when UV was not applied to the sample; thus showing the UV-dependence of an obtained signal. ROIs that were +UV, but not directed to tissue (i.e., ROIs 1-4 – "No Tissue") gave background counts. Regions that were primarily normal tissue (i.e., ROIs 5-8 – "Normal Tissue) gave low counts for the HER2 probe (orange bars on the graph). Regions that were primarily tumor tissue (i.e., ROIs 9-12 - "Tumor") gave higher counts for HER2. A similar, but less dramatic, increase was seen for the Ribosome S6 probe (green bars in graph). Additional control probes targeted RNAs not expected to be expressed highly in this tissue type gave consistent counts that did not show differential levels between Normal and Tumor Tissue. These control probes were designed to target CD45, PSA (Prostate-Specific

Antigen), and two unique ERCC sequences. For clarity, **Figure 53** shows the averages and standard deviations of data shown in **Figure 52**.

[00214] These RNA probe samples were also run simultaneously with Protein probes that analyzed the sample regions of the tumor sample. For this, RNA and Protein probes were simultaneously hybridized to nCounter® Molecular Barcodes, and digitally counted by an nCounter® system from NanoString Technologies®. Counts for this assay are shown in **Figure 54**. An increase in HER2 RNA probe counts (Red bars in top graph) and Protein probe counts (Red and orange bars in bottom graph) are seen in the Tumor regions compare to the normal regions. Only + UV samples are shown. The - UV control sample, as described above, are not shown in this graph because they gave background counts (similar to "No Tissue" counts). ROI 6 and ROI 8 were dropped from this analysis because matching Protein probe samples were not obtained. Thus, signal oligonucleotides from Protein probes and signal oligonucleotides from RNA probes can be detected and quantified together.

[00215] Example 17: Partially double-stranded probes have higher signal-to-noise ratios when compared to single-stranded probes

[00216] DNA probes (that recognize and bind to mRNA) were hybridized in situ, as described in **Example 16**, to RNA in 5 μm FFPE tissues. UV cleavage was performed on whole tissue sections, mounted on separate slides, for 3 minutes using a UV light box (gel box) in 2X SSC + 0.1% Tween 20. After cleavage and release of the signal oligonucleotides, the signal oligonucleotides were collected by a pipette and detected as in **Example 16**. Single-stranded DNA probes, partially double-stranded DNA probes, and no probe controls counts are shown for HER2 3+ breast tissue and tonsil tissue in **Figure 55** (top graph). Signal-to-noise ratios were determined by dividing counts by average background counts (average ERCC counts); see, **Figure 55,** bottom graph.

[00217] Example 18: Addition of salmon sperm DNA improves probe hybridization

[00218] DNA probes (that recognize and bind to mRNA) were hybridized in situ, as described above, to RNA in 5 μm FFPE tissues. 1 mg/ml sonicated, denatured salmon sperm DNA was used instead of yeast tRNA during hybridization. Slides were hybridized with a solution of 1nM probes, 40% formamide, 1mg/ml sonicated, denatured salmon sperm DNA, 10% dextran sulfate,

and 0.2% BSA in 2X SSC. UV cleavage and signal oligonucleotide collection and detection were performed as described in **Example 17**. Single stranded DNA probes are shown in HER2 3+ breast and tonsil (**Figure 56**). Signal-to-noise ratios were determined by dividing counts by average background counts (average ERCC counts).

[00219] Example 19: PSA (Prostate-Specific Antigen) RNA probe is highly specific

[00220] DNA probes (that recognize and bind to mRNA) were hybridized in situ, as described above, to RNA in 5 μm sections of FFPE prostate. A ten minute incubation in MES for at 97°C was used instead of a one hour ethanol incubation. UV cleavage, signal oligonucleotide collection and detection, and signal-to-noise ratio calculations were performed as described in **Example 17**. Counts and ratios are shown in **Figure 57**.

[00221] Example 20: Specificity of probes increase at non-standard, sub-nM concentrations

[00222] Typically, in situ hybridization (ISH) probes that are used to recognize RNA are hybridized at 5 to 200 nM. Surprisingly, nucleic acid recognizing-probes of the present invention performed best at, or below, 0.2 nM, which is 25 to 1000-fold lower than standard ISH probe concentrations.

[00223] DNA probes were hybridized to RNA in situ, as described above, in 5 μm sections of FFPE HER2 3+ breast samples. Probes were used at 5, 1, 0.2, and 0.4 nM. UV cleavage, signal oligonucleotide collection and detection, and fold change calculation were performed as described in **Example 17.**

[00224] Figure 58 shows that counts decreased with decreasing probe concentrations (top graph). However, unexpectedly, there was a significant gain in signal-to-noise when positive probe counts are compared to negative control probes, when probes are hybridized at sub-nM concentrations.

What is claimed is:

1. A method comprising:

(1) contacting at least one nucleic acid target in or from at least one cell in a sample with at least one probe comprising a target-binding domain and a signal oligonucleotide;

(2) providing a force to a location of the sample sufficient to release the signal oligonucleotide; and

(3) collecting and identifying the released signal oligonucleotide, thereby detecting the at least one nucleic acid target in or from a specific location of the sample that was provided the force.

2. The method of claim 1, wherein detecting comprises determining the identity and the abundance of the at least one nucleic acid target.

3. The method of claim 2, wherein the at least one nucleic acid target comprises at least two distinct nucleic acid targets or at least two copies of the same nucleic acid target.

4. The method of claim 3, wherein detecting comprises comparing the abundance of each distinct nucleic acid target.

5. The method of claim 1, further comprising repeating at least steps (2) and (3) on at least a second specific location of the sample, the second specific location comprising at least a second cell.

6. The method of claim 5, wherein detecting comprises comparing the abundance of the at least one nucleic acid target in or from the specific location and in or from the at least second specific location.

7. The method of claim 6, wherein the at least one cell and the at least second cell are the same cell type.

8. The method of claim 6, wherein the at least one cell and the at least second cell are distinct cell types.

9. The method of claim 8, wherein detecting comprises comparing the abundance of the at least one nucleic acid target in or from a first cell type and in or from the at least second cell type.

10. The method of claim 9, wherein the first cell type and the at least second cell type are independently selected from a normal cell and an abnormal cell.

11. The method of any one of claims 1 to 10, wherein the at least one cell is directly immobilized to a surface or is indirectly immobilized to the surface via at least one other cell.

12. The method of claim 11, wherein the sample is a 2 to 1000 μm thick tissue section.

13. The method of claim 12, wherein the tissue section is obtained from a formalin-fixed paraffin embedded (FFPE) sample.

14. The method of any one of claims 1 to 11, wherein the at least one cell is a cultured cell, a primary cell, or a dissociated cell from an explant.

15. The method of claim 12 or 14, wherein the at least one cell is fixed or unfixed.

16. The method of any one of claims 1 to 15, wherein the at least one cell is stained or labeled prior to step (2) thereby allowing visualization of a subcellular, cellular, or tissue-related structure in the stained or labeled cell.

17. The method of any one of claims 1 to 16, wherein the at least one probe further comprises a linker located between the target-binding domain and the signal oligonucleotide.

18. The method of any one of claims 1 to 17, wherein the signal oligonucleotide is a single-stranded nucleic acid or a partially double-stranded nucleic acid.

19. The method of any one of claims 1 to 18, wherein a negative purification is used to remove intact probe molecules from the released signal oligonucleotides.

20. The method of claim 19, wherein the negative purification comprises an affinity purification comprising contacting an intact probe with an immobilized oligonucleotide that is complementary to a portion of the intact probe or an immobilized antibody or protein-binding motif that recognizes and binds to a portion of the intact probe.

21. The method of claim 20, wherein the intact probe's target binding domain comprises a universal purification tag or sequence that is partially complementary to the immobilized oligonucleotide or is capable of being recognized or bound by the immobilized antibody or protein-binding motif.

22. The method of any one of claims 1 to 21, wherein the linker comprises a cleavable linker.

23. The method of claim 22, wherein the cleavable linker is photo-cleavable.

24. The method of claim 23, wherein the provided force is light.

25. The method of claim 24, wherein the light is provided by a light source selected from the group consisting of an arc-lamp, a laser, a focused UV light source, and light emitting diode (LED).

26. The method of claim 25, wherein the light irradiates at least one subcellular structure of the at least one cell.

27. The method of any one of claims 1 to 26, wherein detecting comprises determining the abundance of the at least one nucleic acid target in or from the at least one subcellular structure of the at least one cell.

28. The method of claim 5, wherein the force is light which is provided by a light source selected from an arc-lamp, a laser, a focused UV light source, and a light emitting diode (LED) and the light source irradiates at least one subcellular structure in the at least one cell and at least one subcellular structure in the at least second cell.

29. The method of claim 28, wherein detecting comprises comparing the abundance of the at least one nucleic acid target in or from the at least one subcellular structure in the at least one cell and the at least one subcellular structure in the at least second cell.

30. The method of any one of claims 1 to 29, wherein the target-binding domain comprises a single-stranded nucleic acid or a partially double-stranded nucleic acid.

31. The method of any one of claims 1 to 30, wherein detecting comprises a polymerase reaction, a reverse transcriptase reaction, hybridization to an oligonucleotide microarray, mass spectrometry, hybridization to a fluorescent molecular beacon, a sequencing reaction, or nCounter® Molecular Barcodes.

32. The method of any one of claims 1 to 31, wherein the nucleic acid target is DNA or RNA.

33. The method of claim 32, wherein the RNA is cRNA, mRNA or miRNA.

34. The method of claim 32, wherein the DNA is cDNA.

35. The method of any one of claims 1 to 34, wherein the signal oligonucleotide is collected from the sample *via* liquid laminar, turbulent, or transitional flow in a channel of 25 to 500 μm depth between the sample and a fluidic device or impermeable barrier placed over the sample.

36. The method of any one of claims 1 to 35, wherein the released signal oligonucleotide is collected from a solution proximal to the at least one cell.

37. The method of claim 36, wherein the proximal solution is at least immediately above the at least one cell.

38. The method of claim 37, wherein the proximal solution is collected by aspirating.

39. The method of claim 38, wherein the aspirating is via a pipette, a capillary tube, a microarray pin, and/or a flow cell comprising holes.

40. The method of claim 39, wherein the capillary tube comprises an optical device capable of transmitting a light force to the at least one cell.

41. The method of claim 40, wherein the light force is UV light.

42. The method of claim 41, wherein the pipette or microarray pin is attached to an array comprising a plurality of pipettes or microarray pins.

43. The method of any one of claims 1 to 42, wherein the proximal solution comprises an anionic polymer, preferably dextran sulfate, and/or salmon sperm DNA.

44. The method of any one of claims 1 to 43, wherein the collected signal oligonucleotide is added to a solution comprising an anionic polymer, preferably dextran sulfate, and/or salmon sperm DNA.

45. The method of any one of claims 1 to 44, wherein the method further comprises illuminating a region of interest using a laser scanning device.

46. The method of any one of claims 1 to 44, wherein the method further comprises illuminating a region of interest using a digital mirror device (DMD).

47. The method of any one of claims 1 to 46, wherein the method provides simultaneous spatially resolved DNA, RNA, and/or protein detection of a sample.

48. The method of any one of claims 1 to 47, wherein digital readout comprises a linear dynamic range of > 5 logs.

49. The method of any one of claims 1 to 48, wherein the sample is attached to a slide and is first imaged using fluorescence and then expression of proteins and/or nucleic acids is digitally counted from the sample.

50. The method of any one of claims 1 to 49, wherein the probe is provided at a concentration of 5 nM or less.

51. The method of claim 50, wherein the probe is provided at a concentration of 1 nM or less.

52. The method of claim 51, wherein the probe is provided at a concentration of 0.4 nM or less.

53. The method of claim 52, wherein the probe is provided at a concentration of 0.2 nM or less.

Figure 1

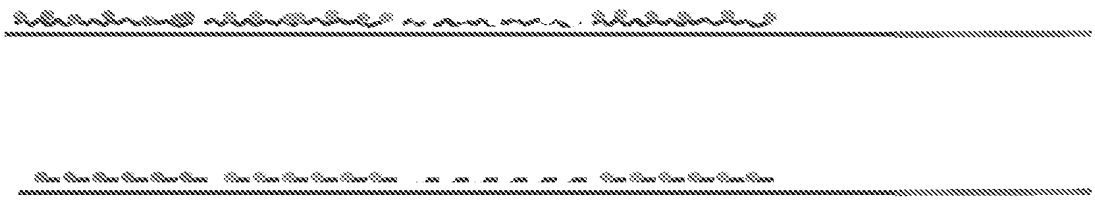


Figure 2

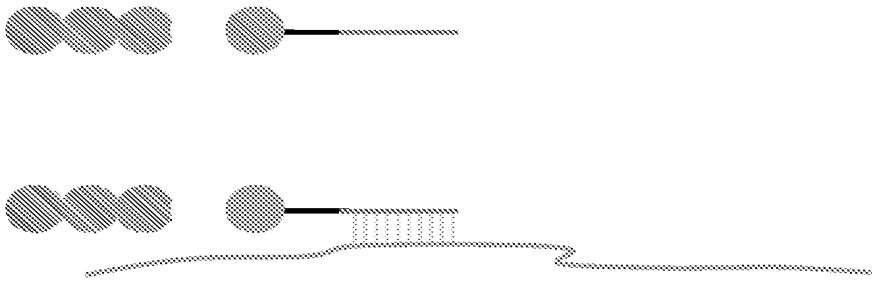


Figure 3

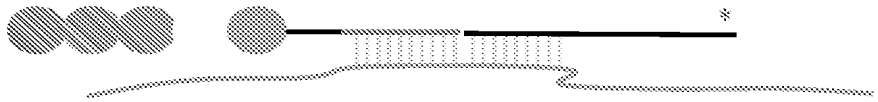


Figure 4

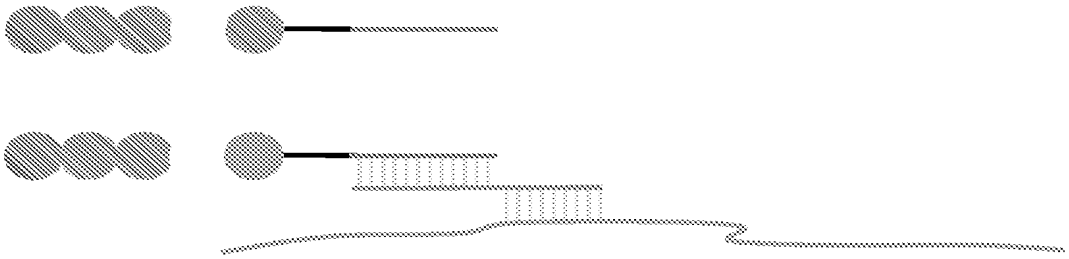


Figure 5

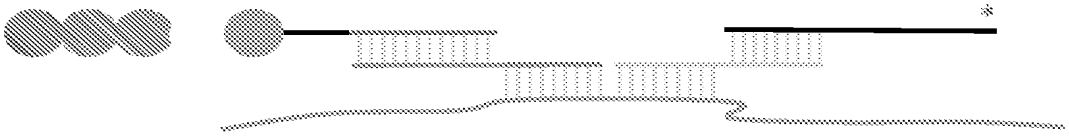


Figure 6

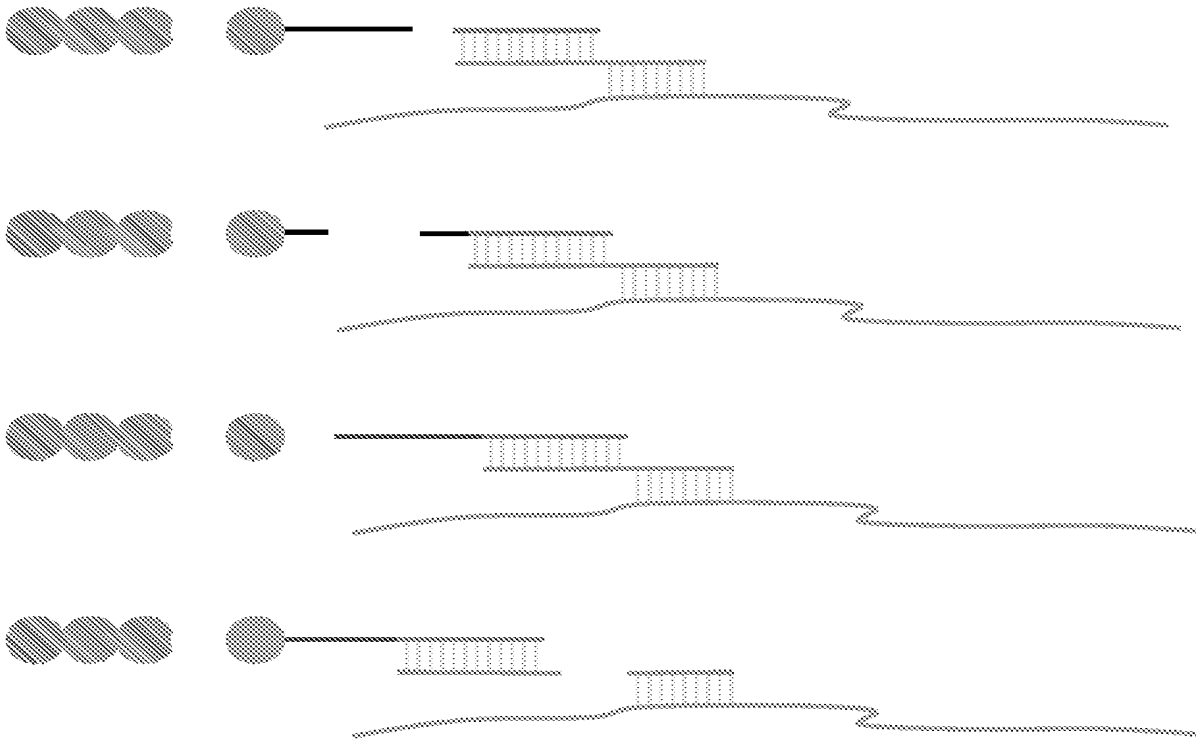


Figure 7

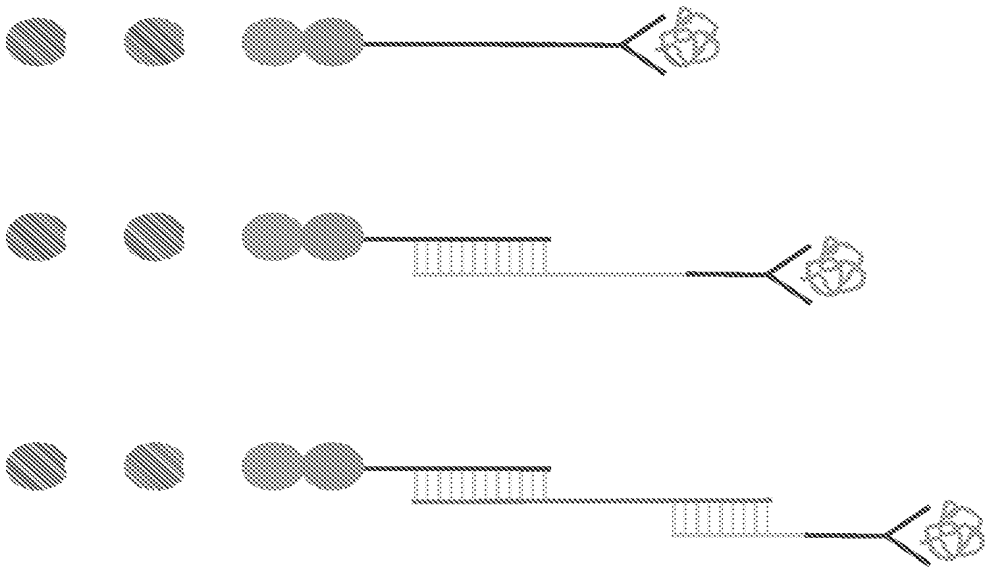


Figure 8

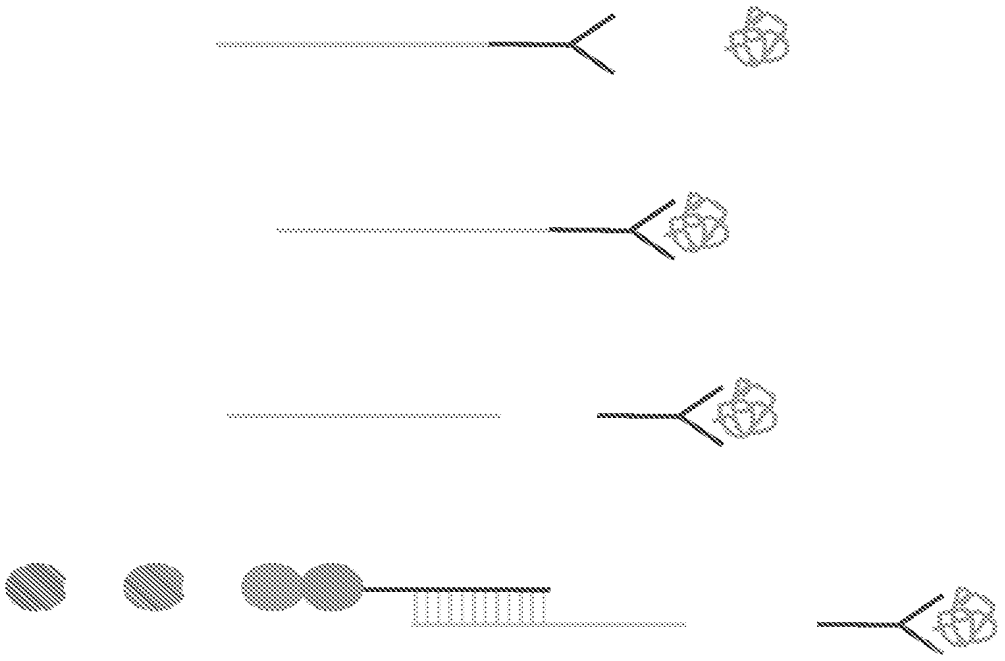


Figure 9

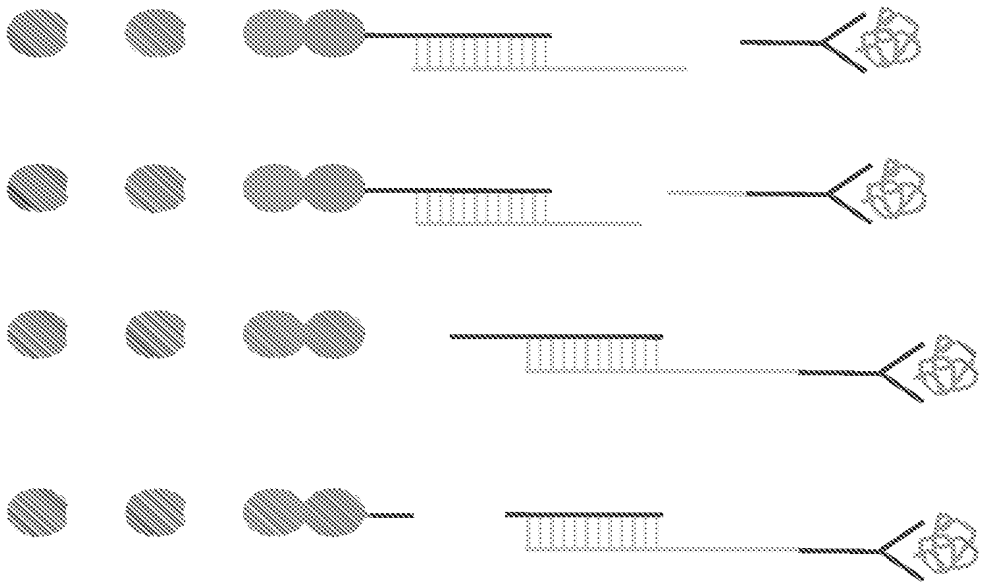


Figure 10

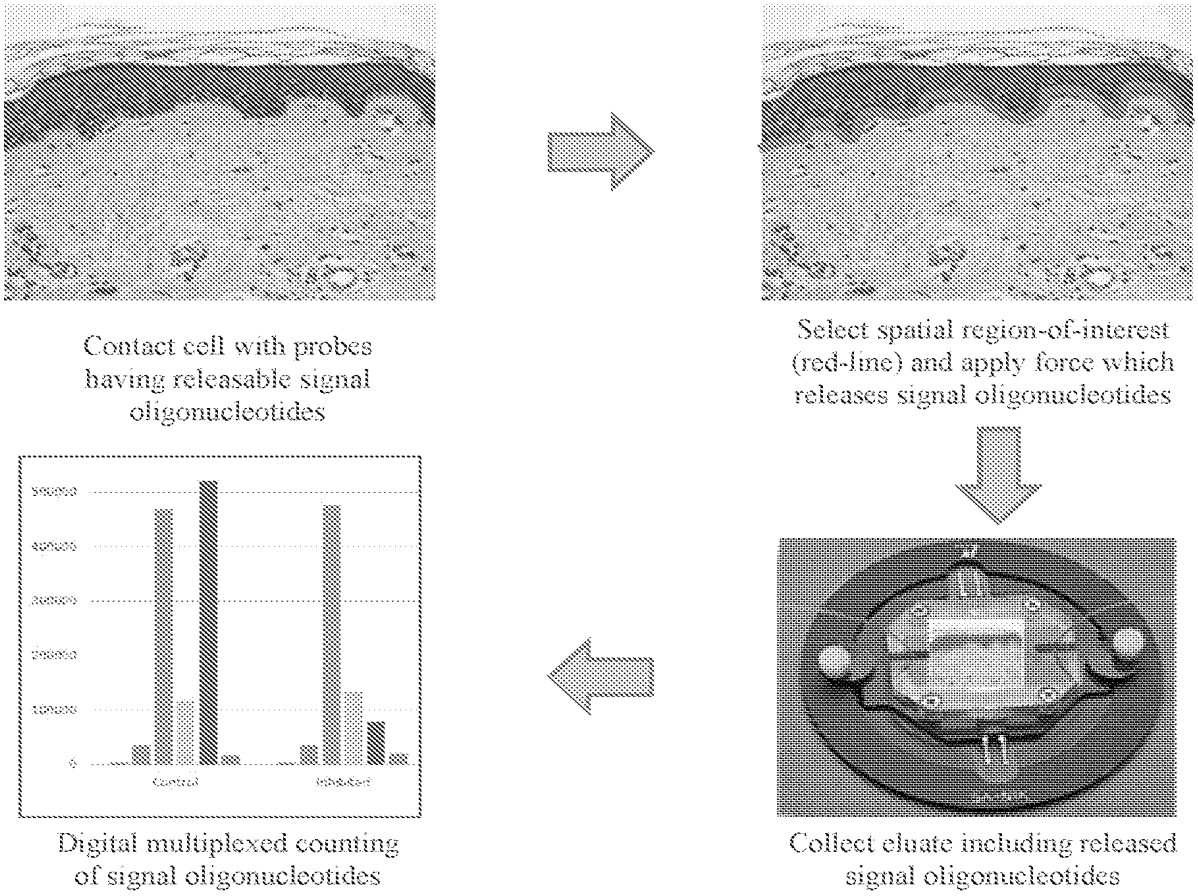

Contact cell with probes having releasable signal oligonucleotides
Select spatial region-of-interest (red-line) and apply force which releases signal oligonucleotides
Collect eluate including released signal oligonucleotides
Digital multiplexed counting of signal oligonucleotides

Figure 11

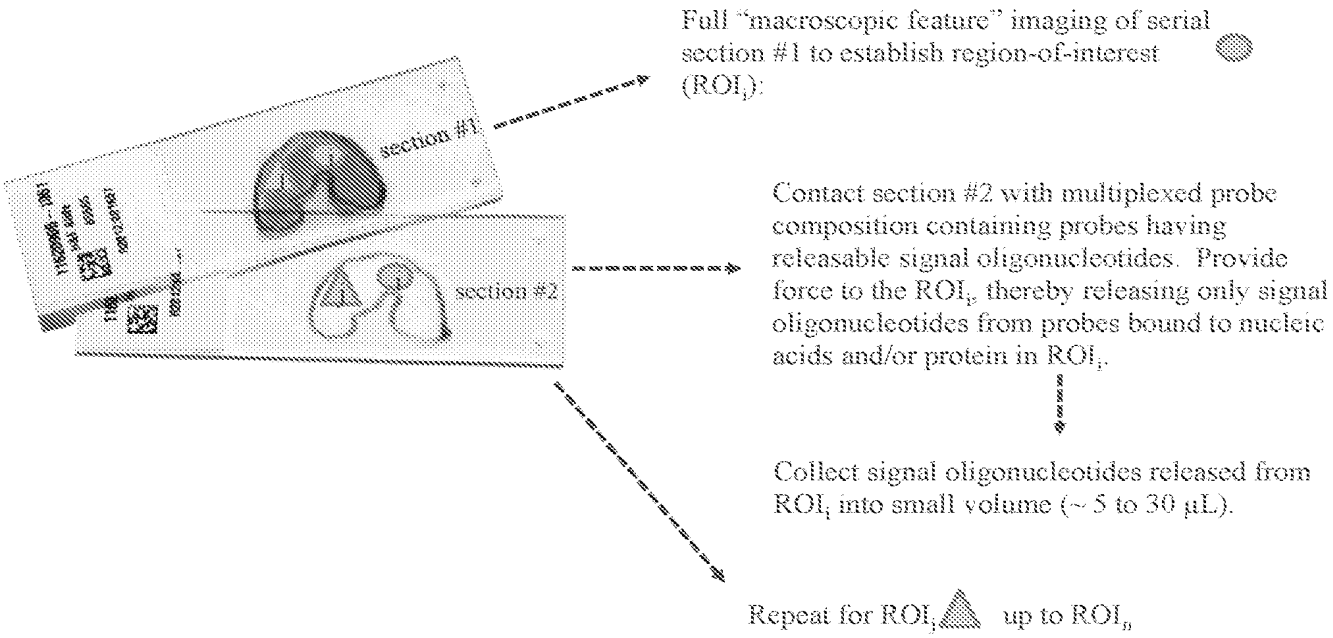
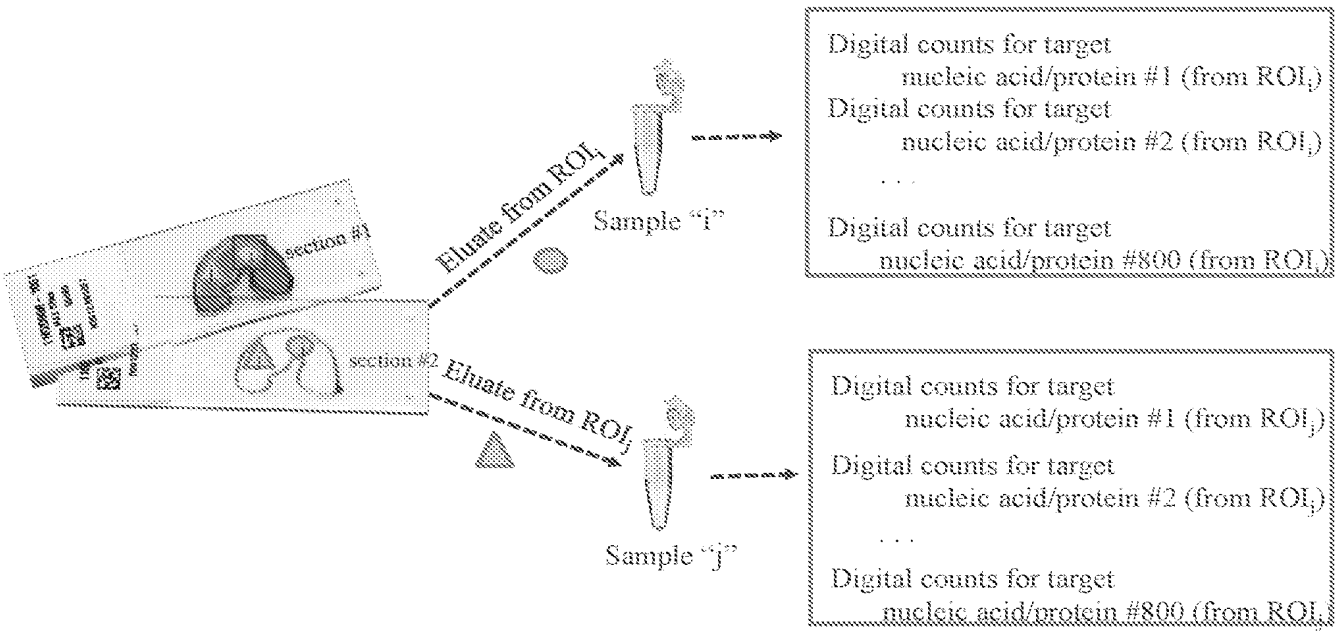

Full "macroscopic feature" imaging of serial section #1 to establish region-of-interest (ROI_i):
section #1
section #2
Contact section #2 with multiplexed probe composition containing probes having releasable signal oligonucleotides. Provide force to the ROI_i, thereby releasing only signal oligonucleotides from probes bound to nucleic acids and/or protein in ROI_i.
Collect signal oligonucleotides released from ROI_i into small volume (~ 5 to 30 μL).
Repeat for ROI_j up to ROI_n

Figure 12


section #1
section #2
Eluate from ROI$_i$
Eluate from ROI$_j$
Sample "i"
Sample "j"
Digital counts for target nucleic acid/protein #1 (from ROI$_i$)
Digital counts for target nucleic acid/protein #2 (from ROI$_i$)
. . .
Digital counts for target nucleic acid/protein #800 (from ROI$_i$)
Digital counts for target nucleic acid/protein #1 (from ROI$_j$)
Digital counts for target nucleic acid/protein #2 (from ROI$_j$)
. . .
Digital counts for target nucleic acid/protein #800 (from ROI$_j$)

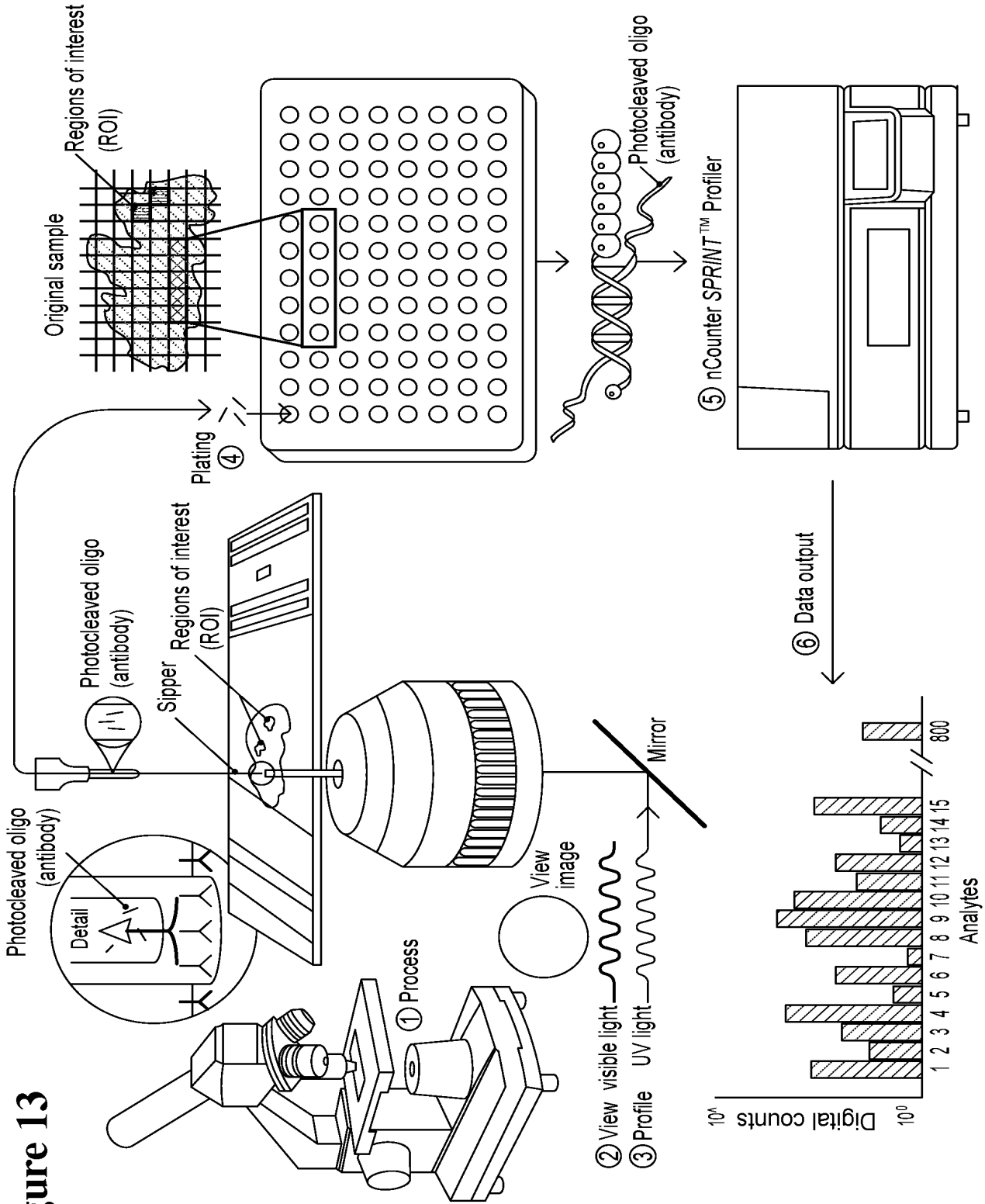

Original sample
Regions of interest (ROI)
Plating ④
Photocleaved oligo (antibody)
⑤ nCounter SPRINT™ Profiler
⑥ Data output
Digital counts
Analytes
Photocleaved oligo (antibody)
Detail
Sipper
Regions of interest (ROI)
Photocleaved oligo (antibody)
① Process
② View visible light
③ Profile UV light
View image
Mirror

Figure 13

Figure 14

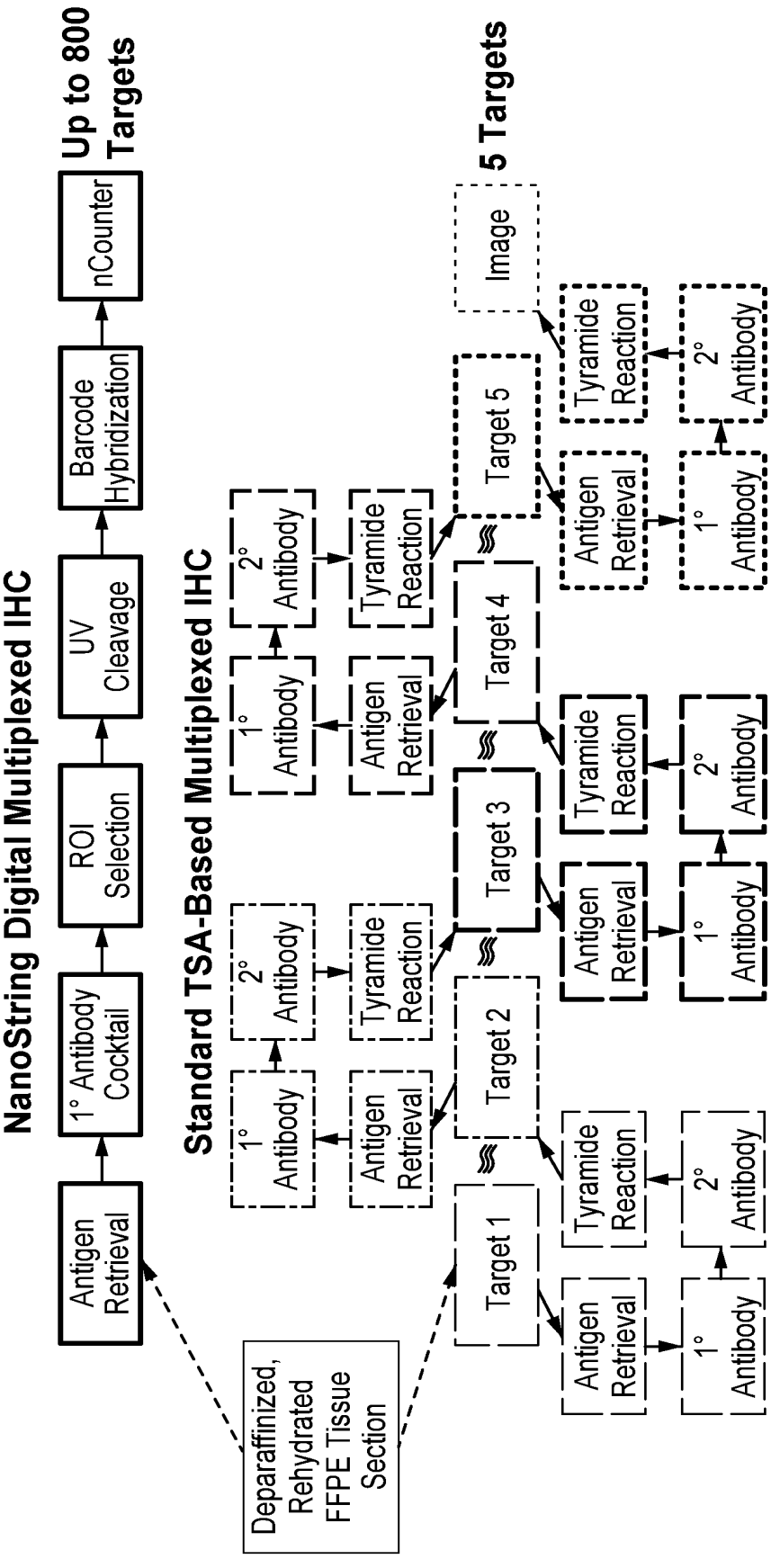

NanoString Digital Multiplexed IHC
Antigen Retrieval
1° Antibody Cocktail
ROI Selection
UV Cleavage
Barcode Hybridization
nCounter
Up to 800 Targets
Standard TSA-Based Multiplexed IHC
Deparaffinized, Rehydrated FFPE Tissue Section
Antigen Retrieval
1° Antibody
2° Antibody
Tyramide Reaction
Target 1
Target 2
Target 3
Target 4
Target 5
Image
5 Targets

Figure 15

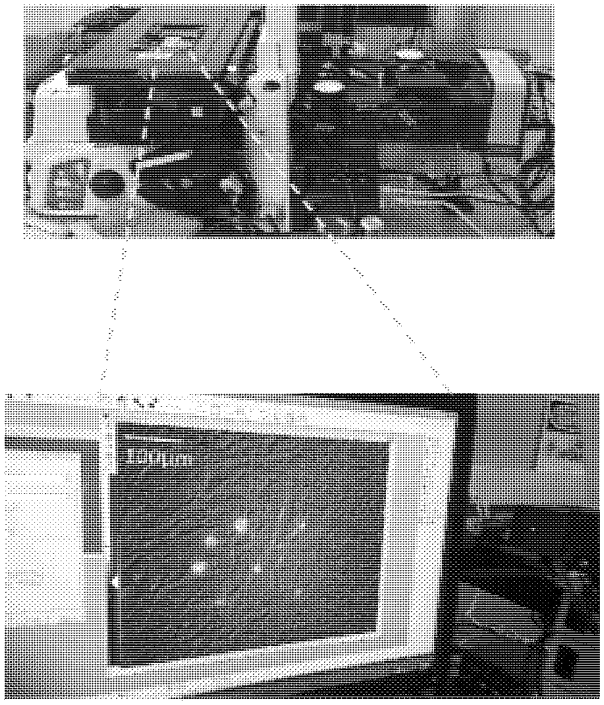
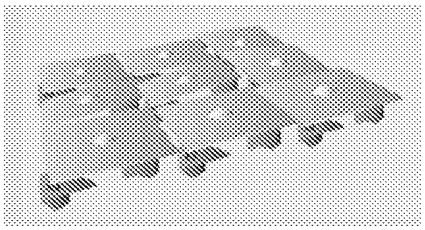


Figure 16



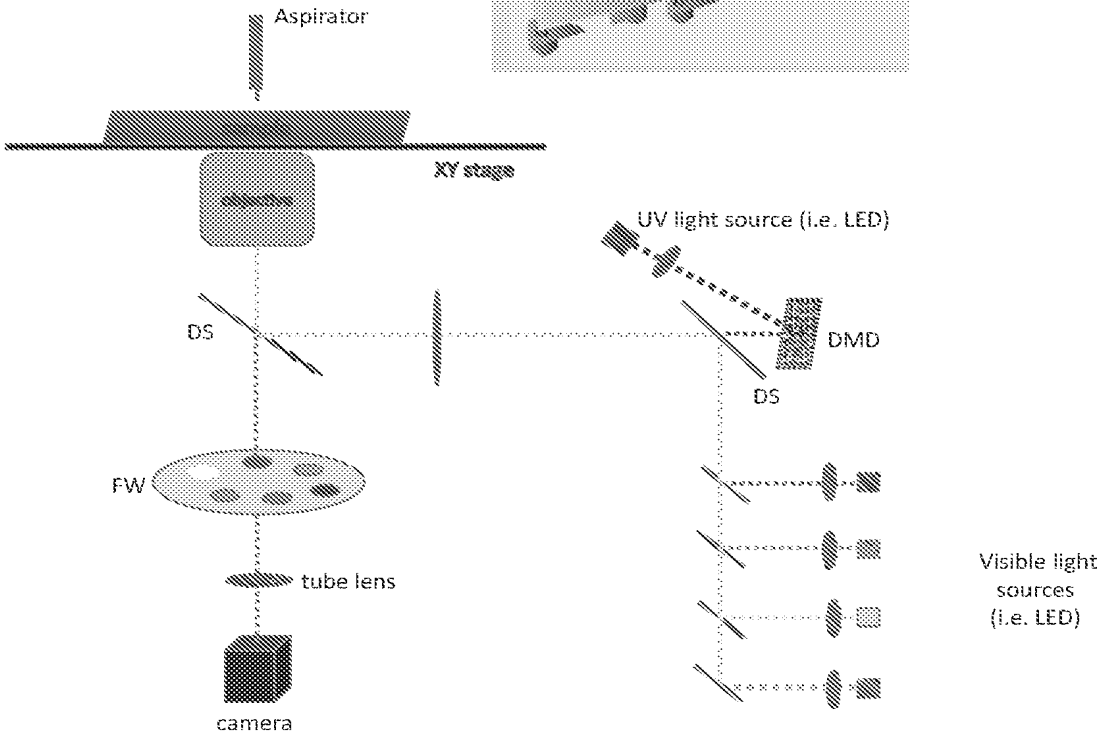

Aspirator
XY stage
Objective
UV light source (i.e. LED)
DS
DMD
DS
FW
tube lens
camera
Visible light sources (i.e. LED)

Figure 17

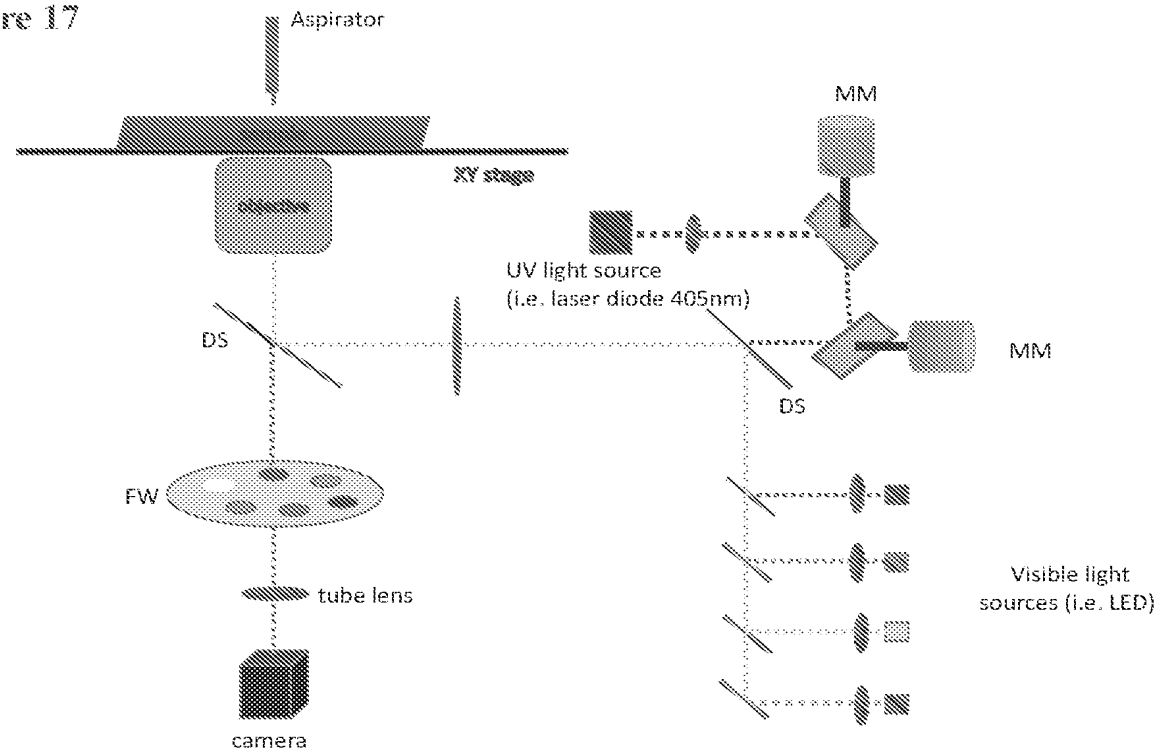

Aspirator
XY stage
MM
UV light source
(i.e. laser diode 405nm)
DS
MM
DS
FW
tube lens
camera
Visible light
sources (i.e. LED)

Figure 18

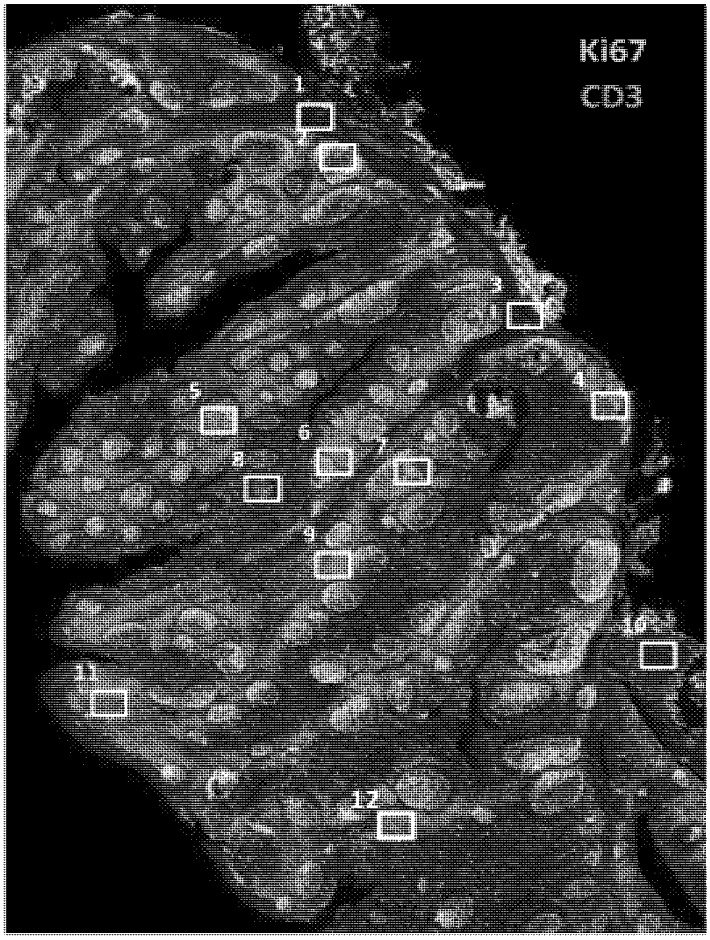


Figure 19

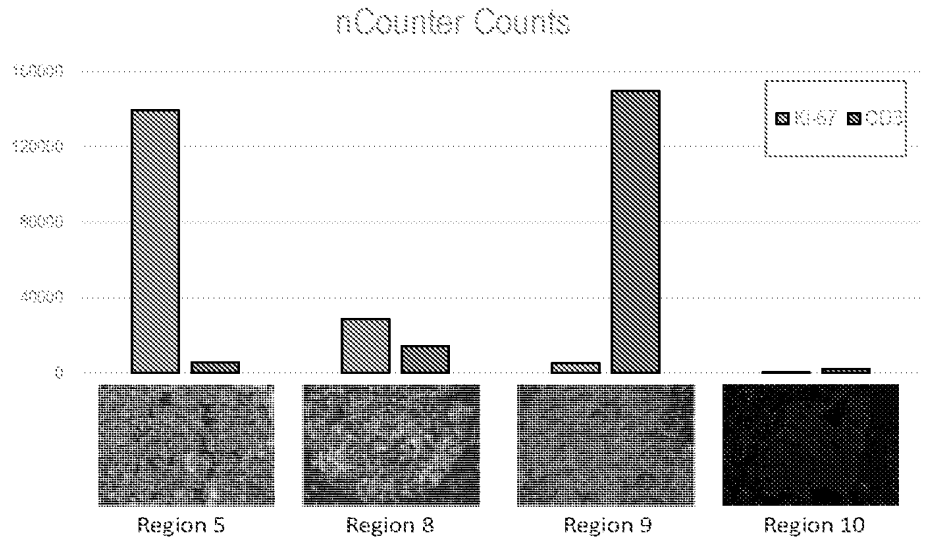


Figure 20

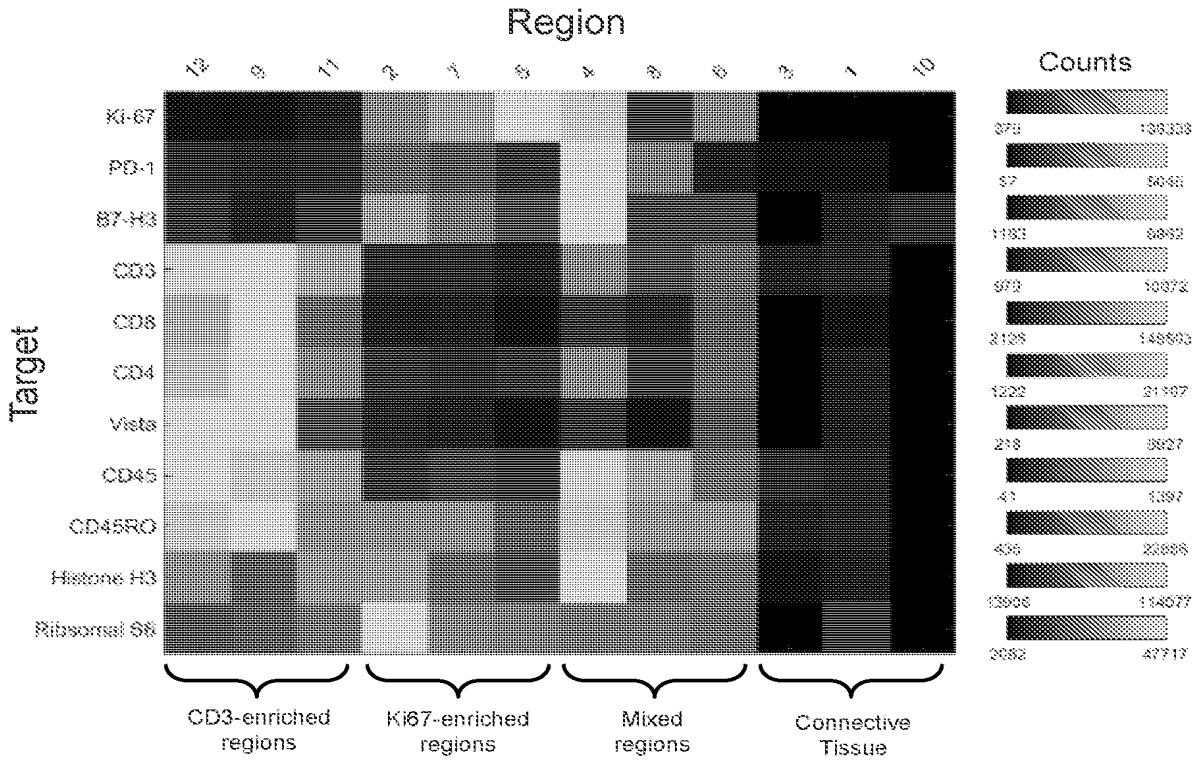

Region
12
8
11
2
7
5
4
3
6
9
1
10
Counts
Target
Ki-67
PD-1
B7-H3
CD3
CD8
CD4
Vista
CD45
CD45RO
Histone H3
Ribsomal S6
CD3-enriched regions
Ki67-enriched regions
Mixed regions
Connective Tissue

Figure 21

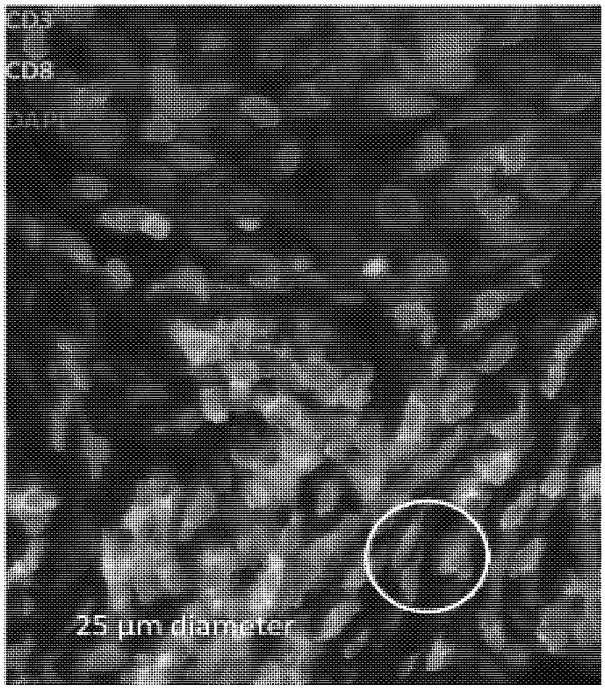


Figure 22

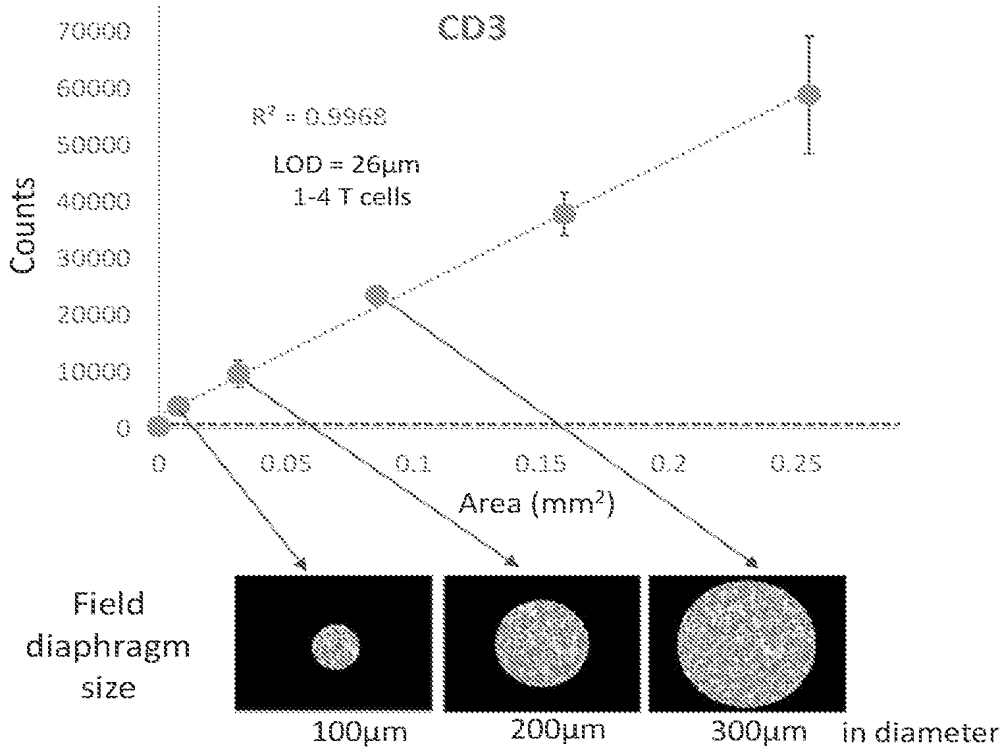

CD3
R^2 = 0.9968
LOD = 26µm
1-4 T cells
Counts
70000
60000
50000
40000
30000
20000
10000
0
0
0.05
0.1
0.15
0.2
0.25
Area (mm^2)
Field diaphragm size
100µm
200µm
300µm
in diameter

Figure 23

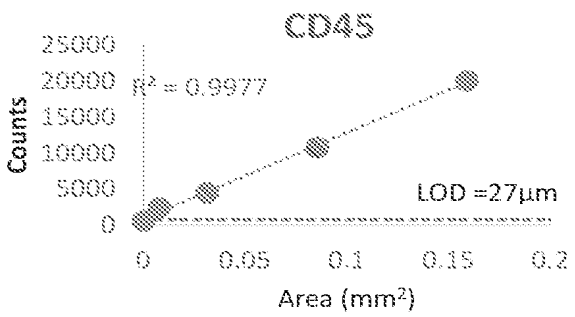


Figure 24

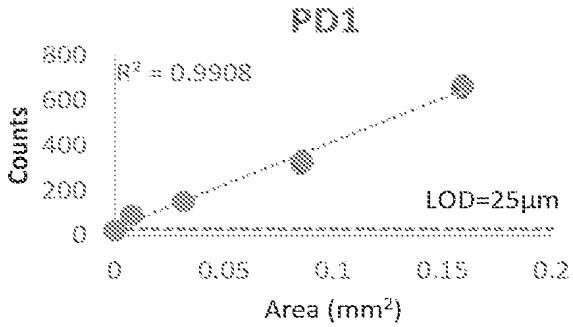


Figure 25

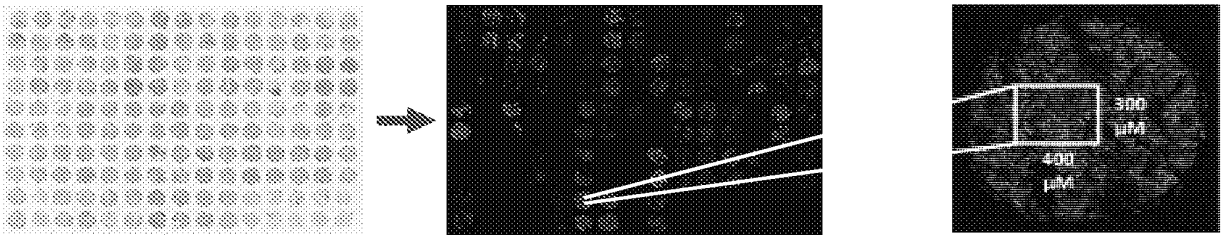


Figure 26

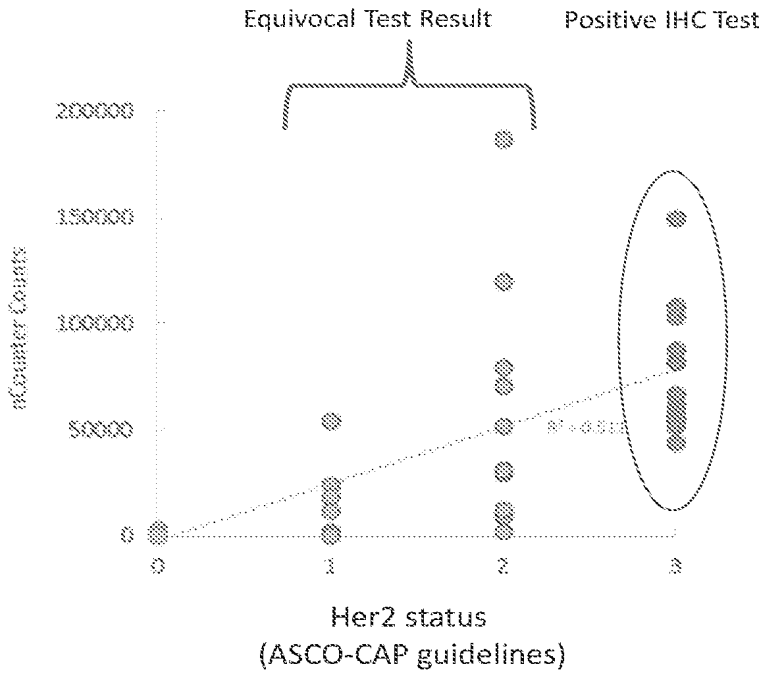

Equivocal Test Result
Positive IHC Test
200000
150000
100000
50000
0
Counter Counts
0
1
2
3
Her2 status
(ASCO-CAP guidelines)

Figure 27

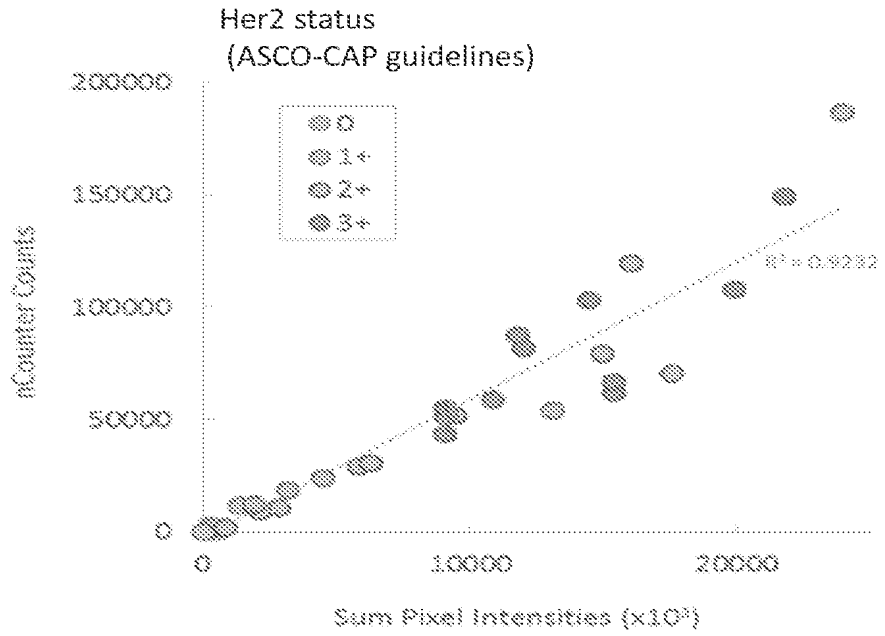

Her2 status
(ASCO-CAP guidelines)
200000
150000
100000
50000
0
nCounter Counts
0
1+
2+
3+
R² = 0.9232
0
10000
20000
Sum Pixel Intensities (x10³)

Figure 28

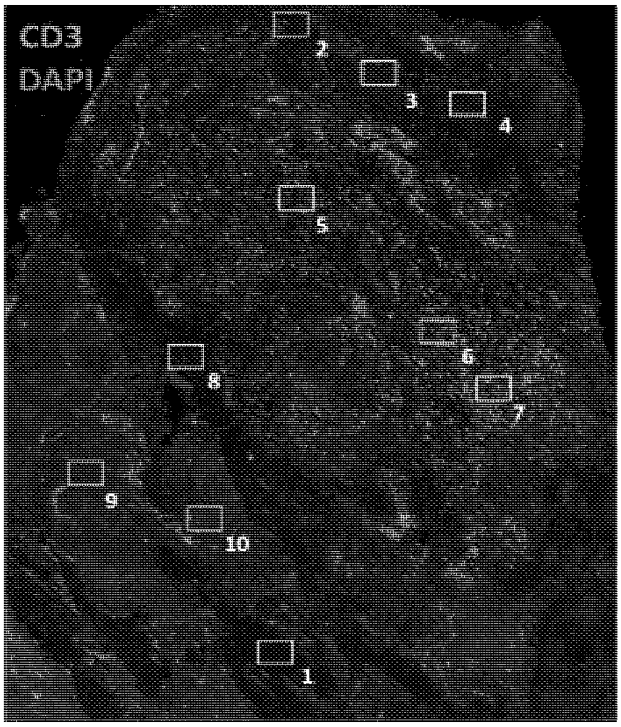


Figure 29

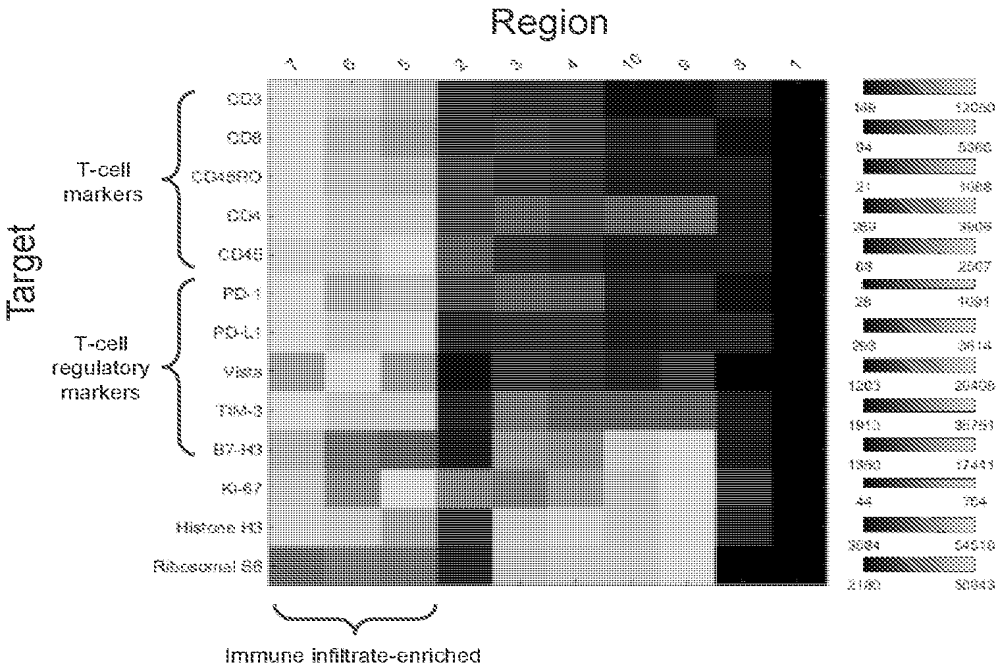

Region
Target
T-cell markers
T-cell regulatory markers
CD3
CD8
CD45RO
CD4
CD45
PD-1
PD-L1
Vista
TIM-3
B7-H3
KI-67
Histone H3
Ribosomal S6
Immune infiltrate-enriched

Figure 30A

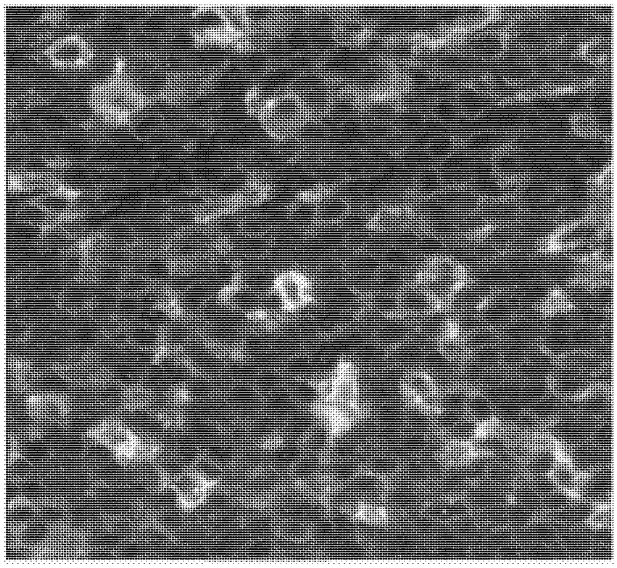


Figure 30B

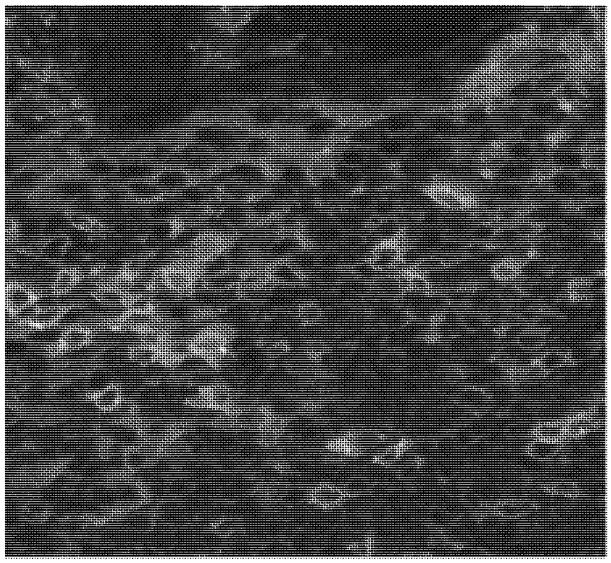


Figure 31B

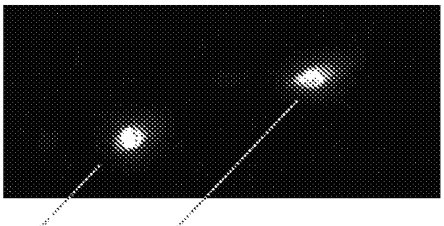
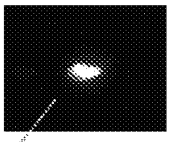


Figure 31D



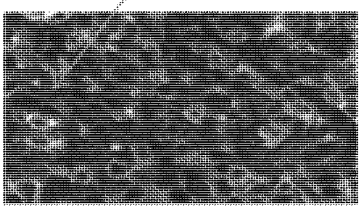


Figure 31C

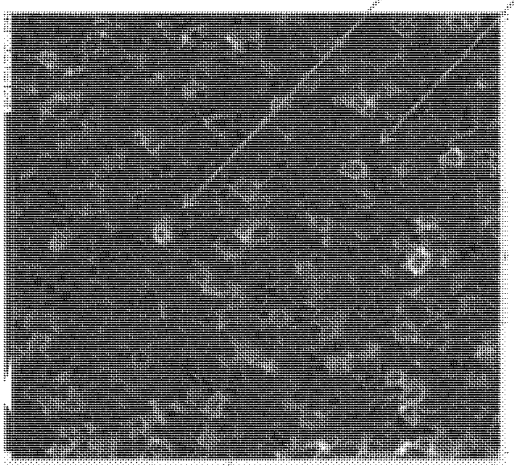


Figure 31A

Figure 32

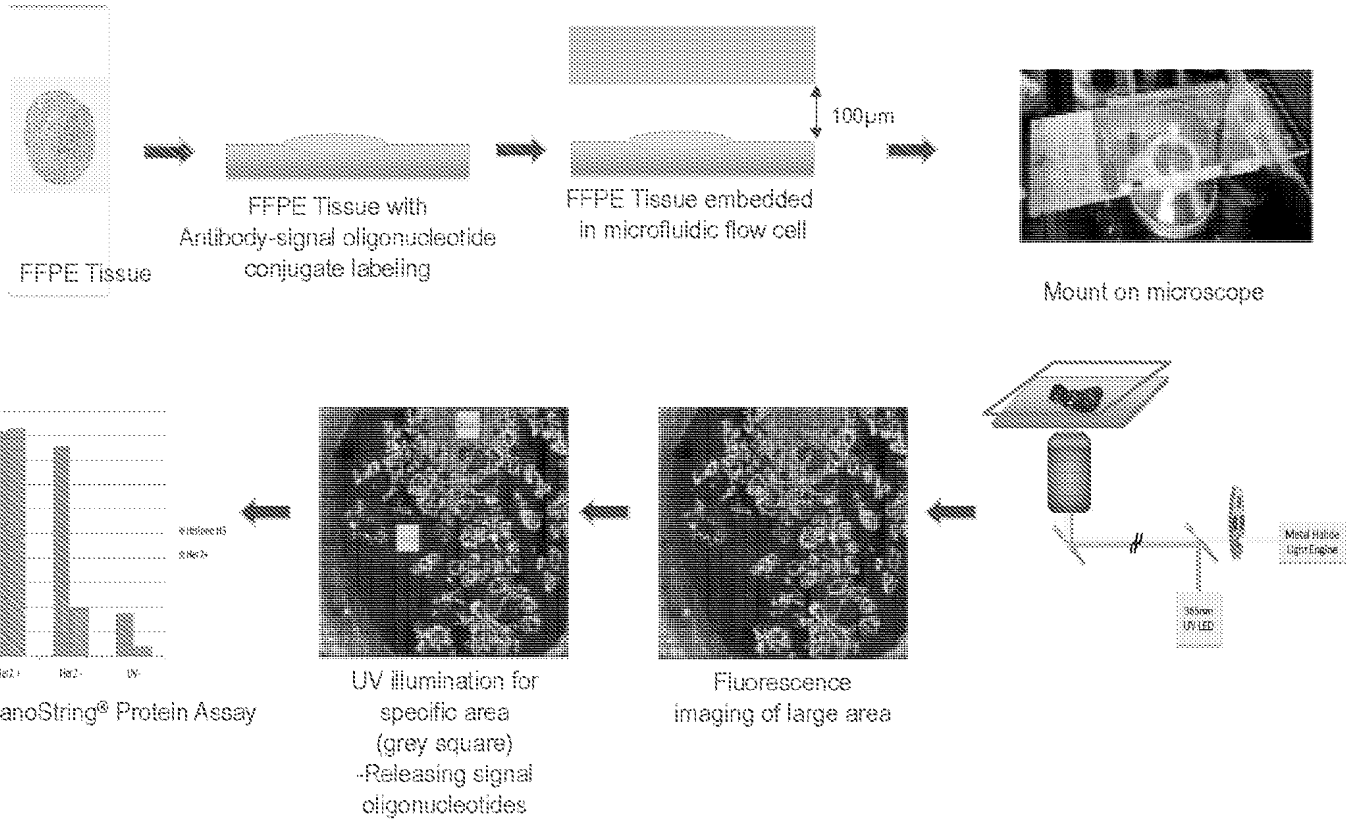

FFPE Tissue
FFPE Tissue with
Antibody-signal oligonucleotide
conjugate labeling
100μm
FFPE Tissue embedded
in microfluidic flow cell
Mount on microscope
Metal Halide
Light Engine
365nm
UV LED
Fluorescence
imaging of large area
UV illumination for
specific area
(grey square)
-Releasing signal
oligonucleotides
NanoString® Protein Assay

Figure 33

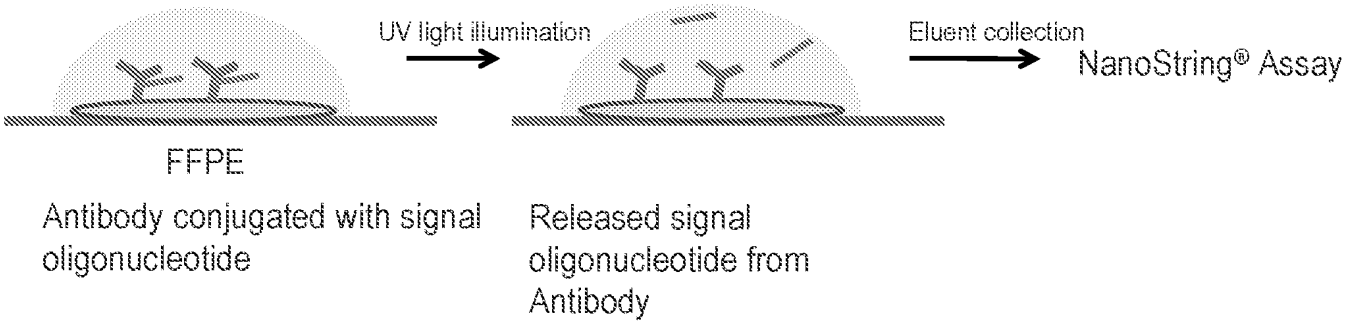


Figure 34

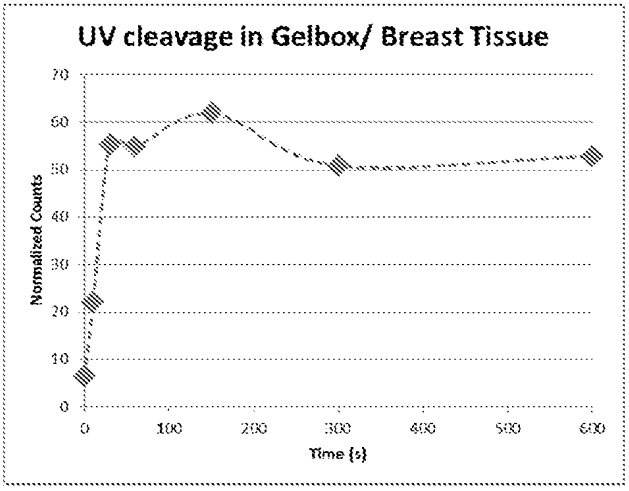
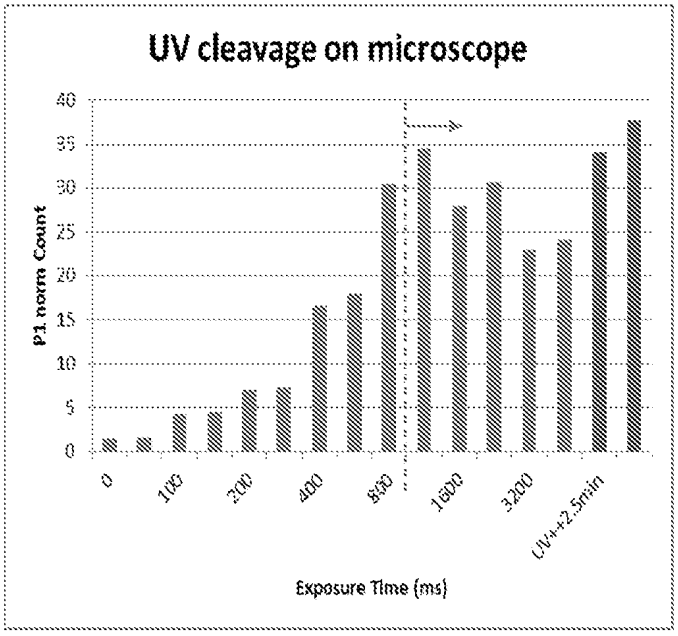
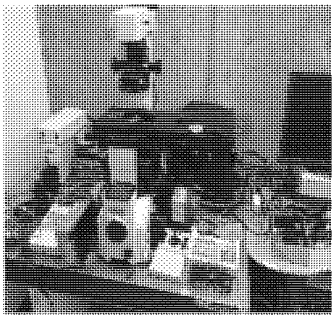
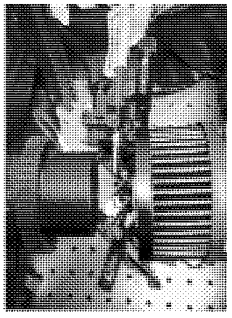


Figure 35


UV cleavage on microscope
P1 norm Count
40
35
30
25
20
15
10
5
0
0
100
200
400
800
1600
3200
UV+2.5min
Exposure Time (ms)





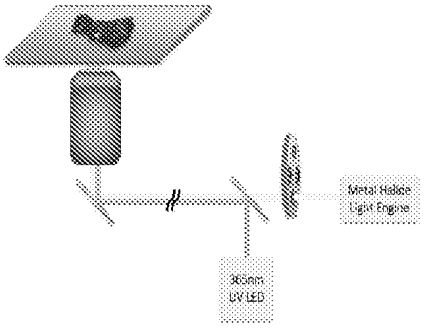

Metal Halide Light Engine
365nm UV LED

Figure 36

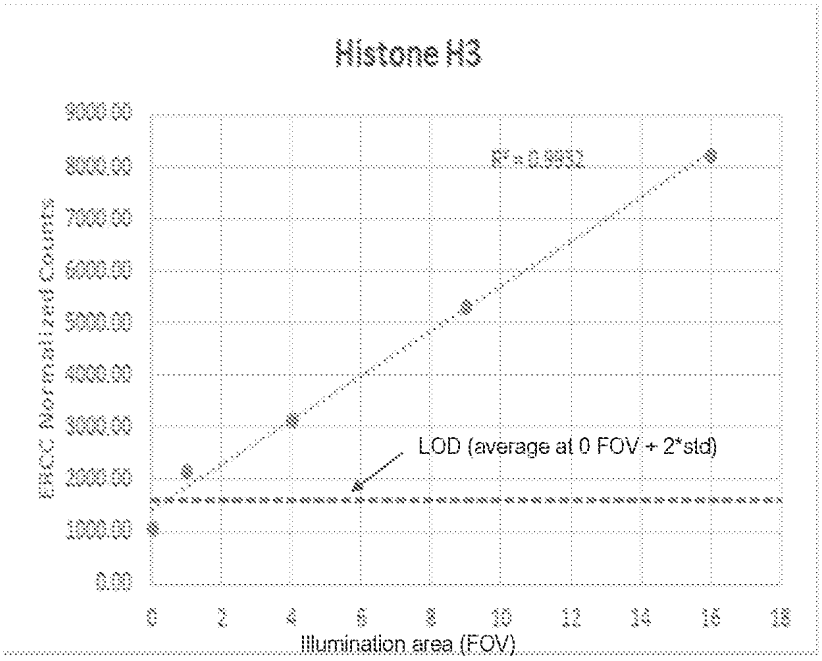

Histone H3
R² = 0.9932
LOD (average at 0 FOV + 2*std)
illumination area (FOV)
0
2
4
6
8
10
12
14
16
18

Figure 37

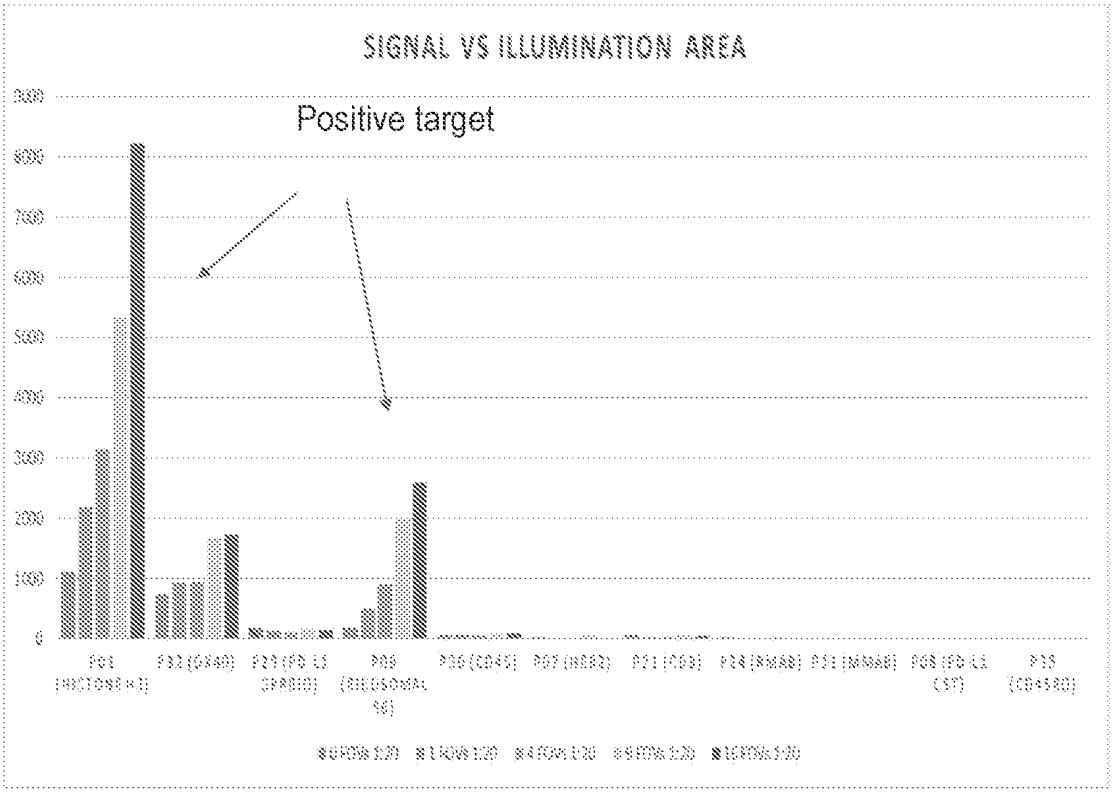
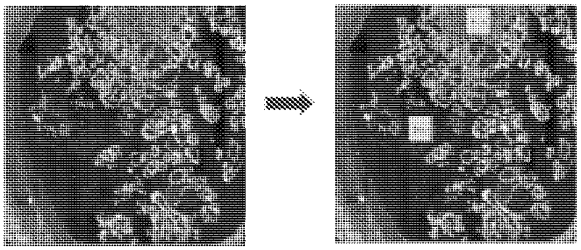

SIGNAL VS ILLUMINATION AREA
Positive target
8000
7000
6000
5000
4000
3000
2000
1000
0
P01 (HISTONE H3)
P32 (GRAM)
P23 (FOXP3)
P05 (DIFFUSONAL 96)
P36 (CD45)
P07 (HER2)
P21 (CD3)
P24 (PANAE)
P31 (MMaA8)
P08 (PD-L1 (SP))
P35 (CD45RO)

Figure 38



Fluorescence imaging of large area

UV illumination for specific area (grey square)

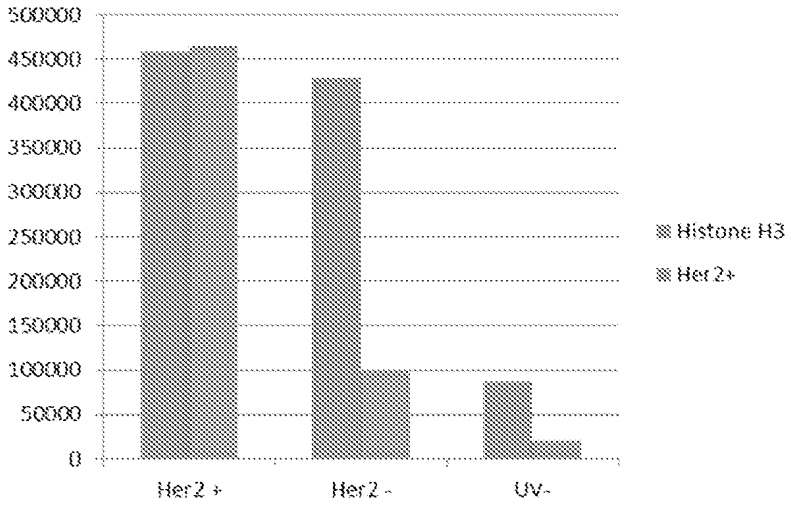
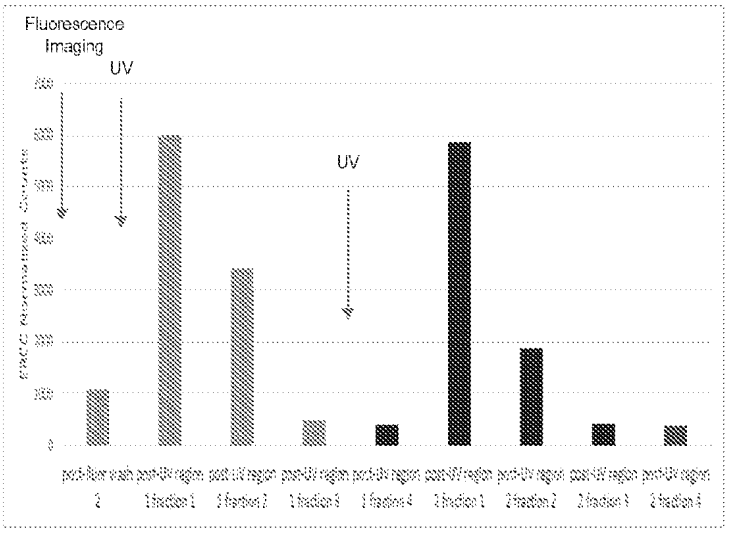
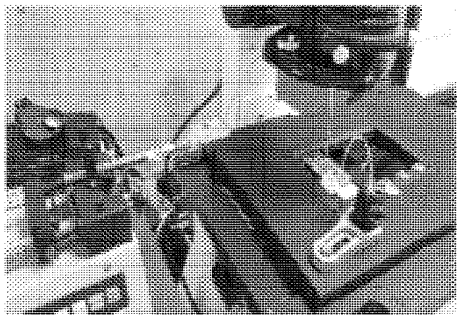
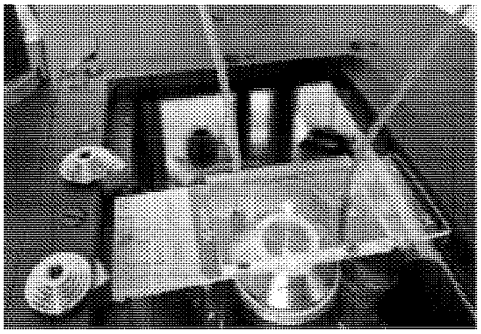


Figure 39


Fluorescence Imaging
UV
UV





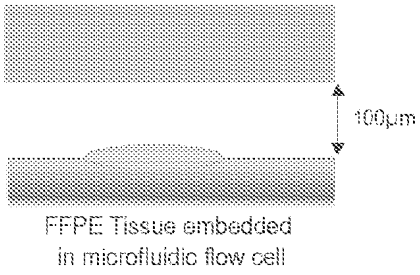

100µm
FFPE Tissue embedded
in microfluidic flow cell

Figure 40

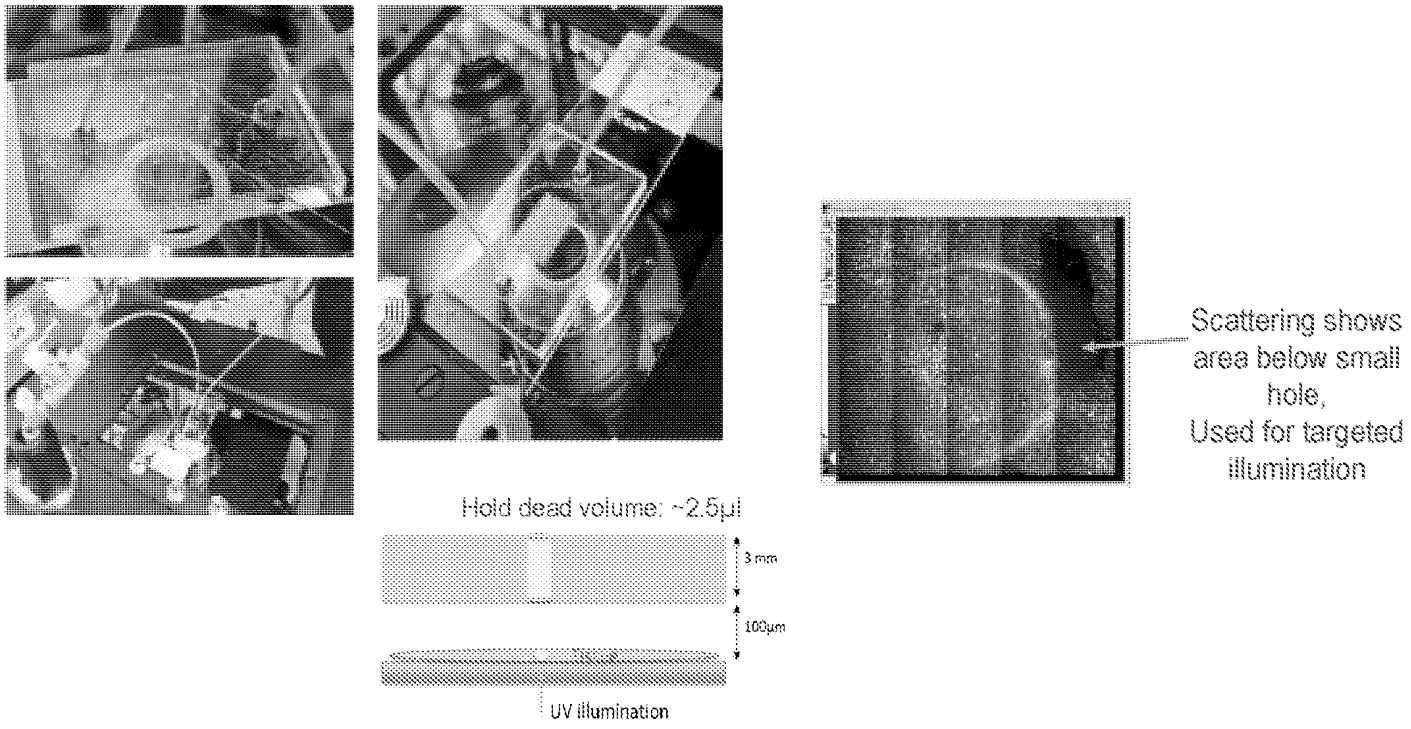

Scattering shows area below small hole, Used for targeted illumination
Hold dead volume: ~2.5µl
3 mm
100µm
UV illumination

Figure 41A

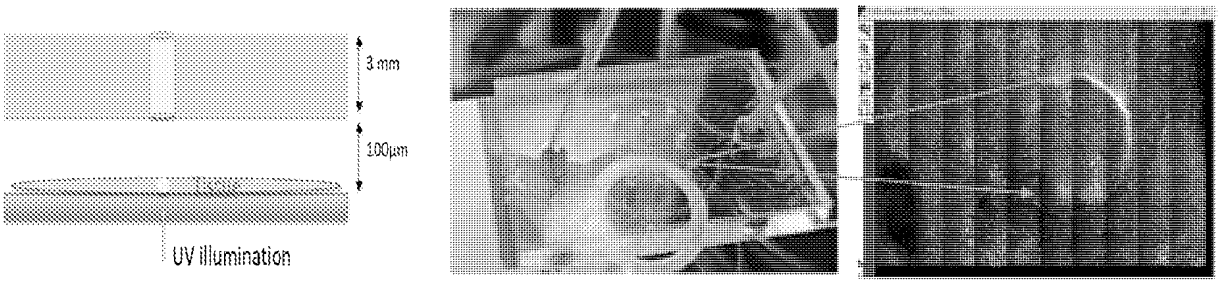


Figure 41B

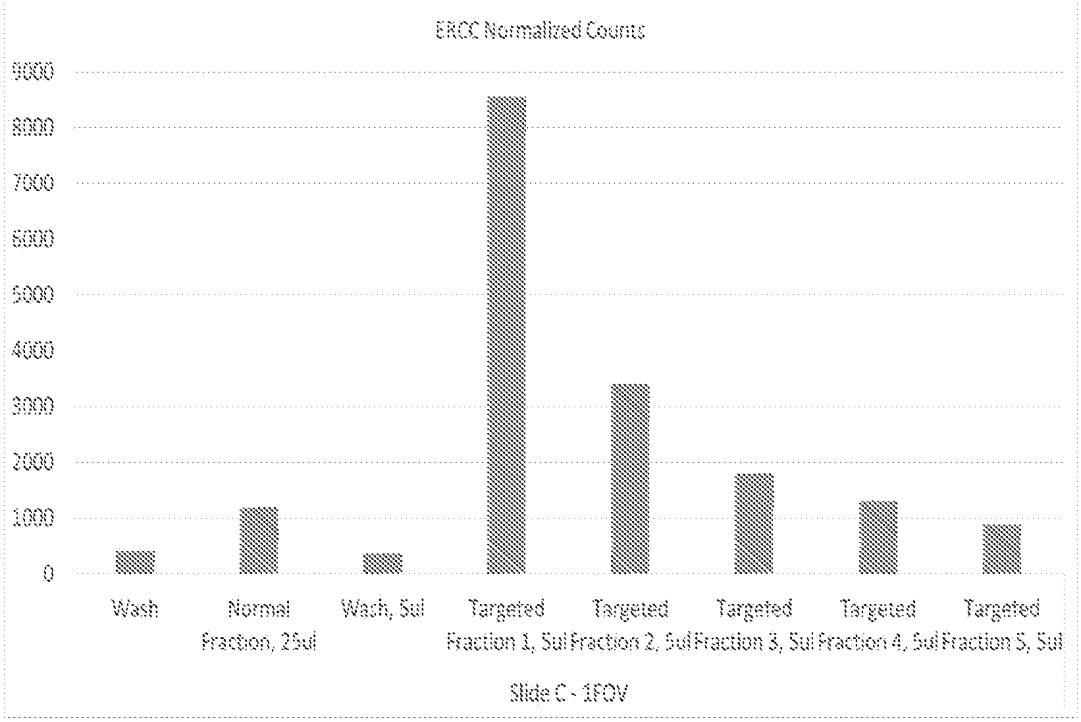

ERCC Normalized Counts
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Wash
Normal Fraction, 25ul
Wash, 5ul
Targeted Fraction 1, 5ul
Targeted Fraction 2, 5ul
Targeted Fraction 3, 5ul
Targeted Fraction 4, 5ul
Targeted Fraction 5, 5ul
Slide C - 1FOV

Figure 41C

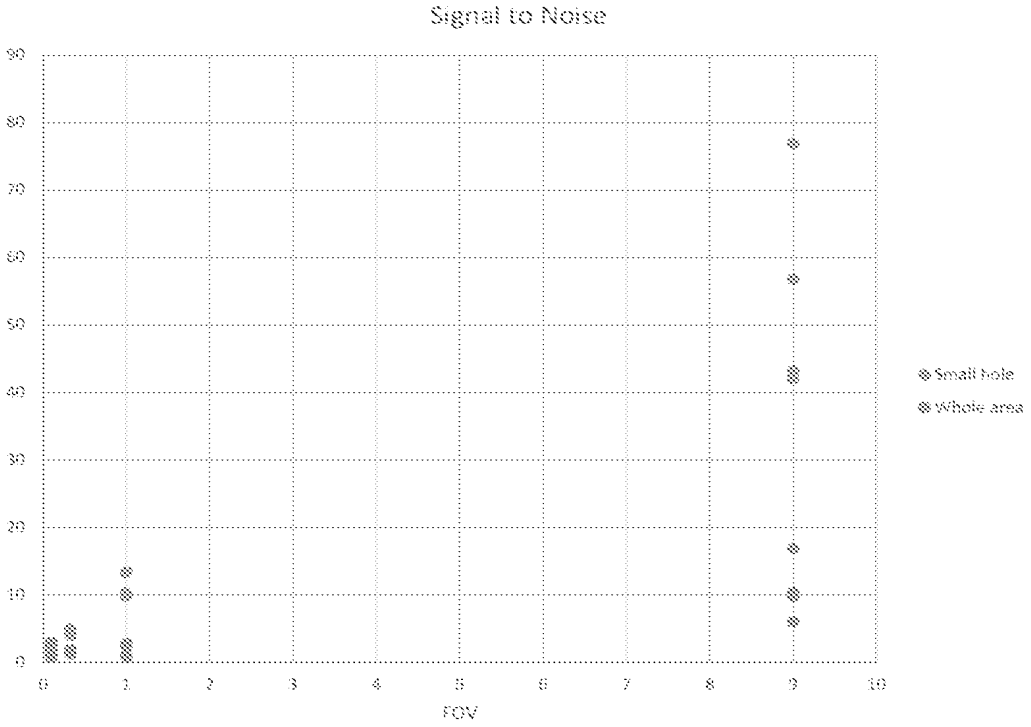
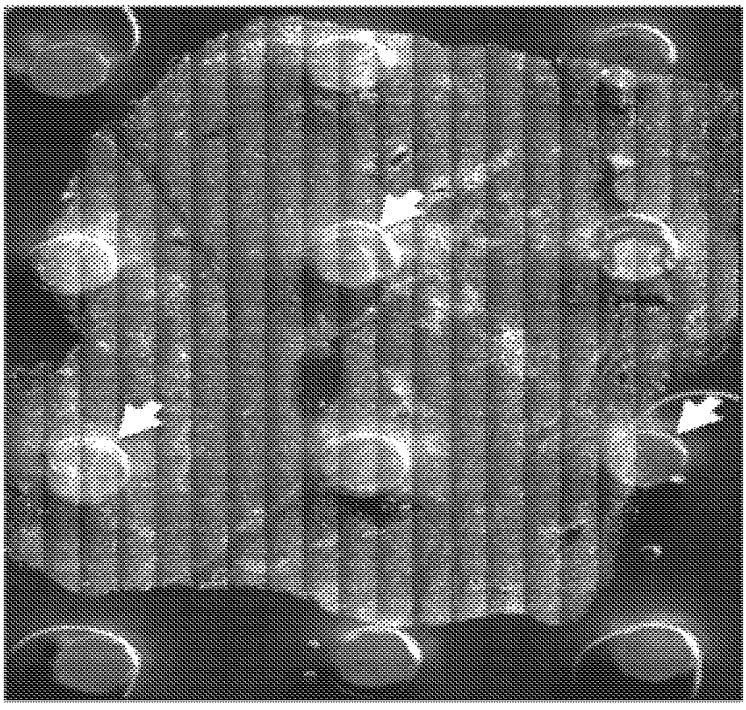

Signal to Noise
90
80
70
60
50
40
30
20
10
0
0
1
2
3
4
5
6
7
8
9
10
FOV
Small hole
Whole area

Figure 42A



12-hole format

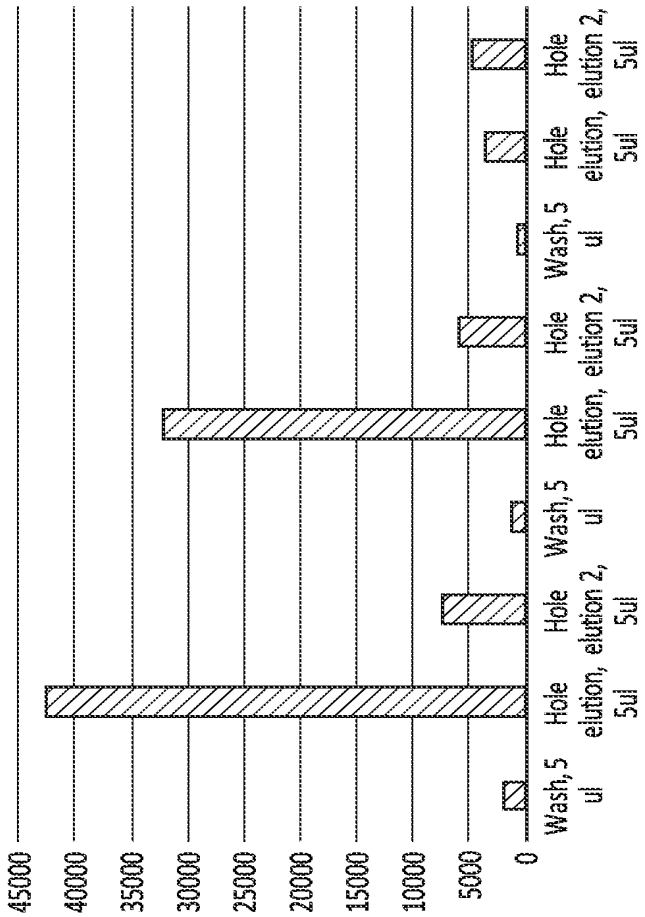
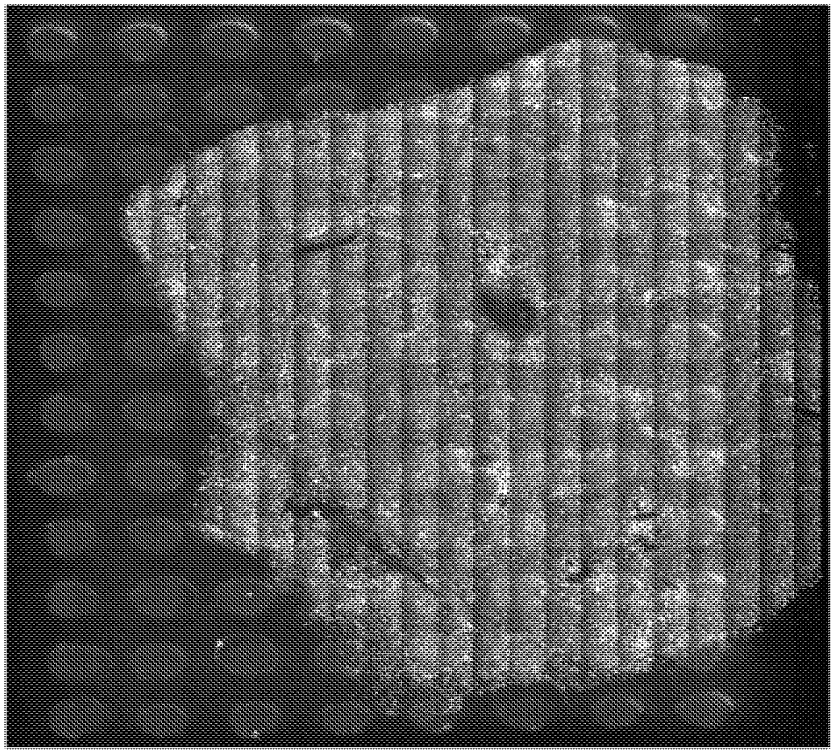
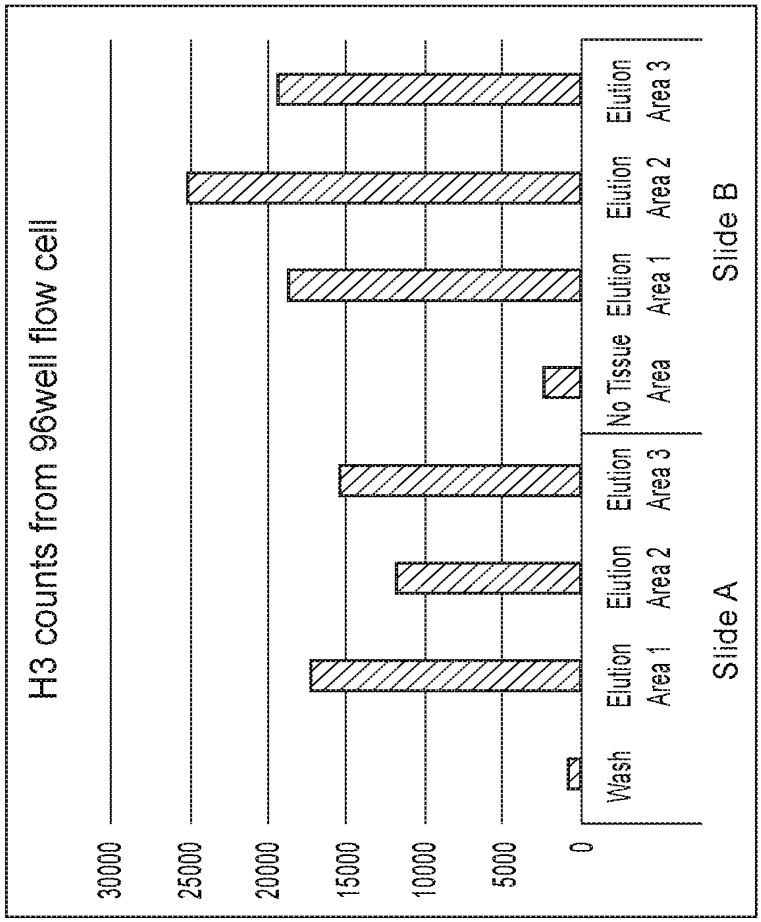


Figure 42B

96-hole format





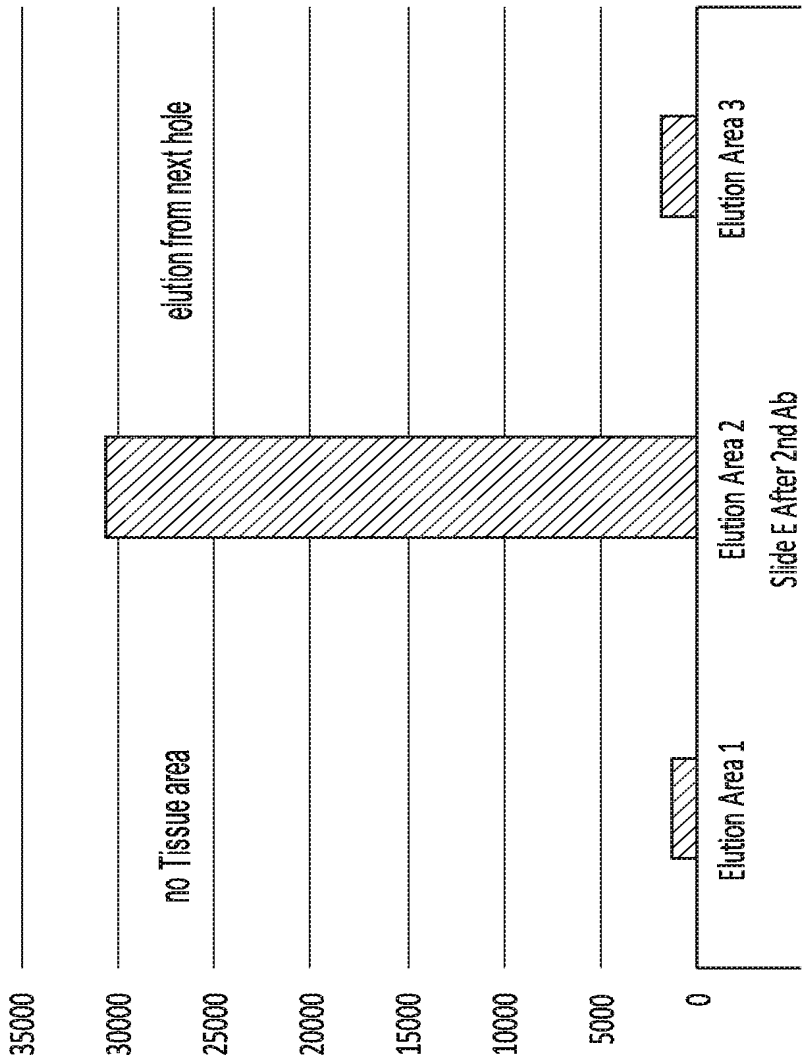

H3 counts
35000
30000
25000
20000
15000
10000
5000
0
no Tissue area
elution from next hole
Elution Area 1
Elution Area 2
Elution Area 3
Slide E After 2nd Ab

Figure 42C

Figure 43A

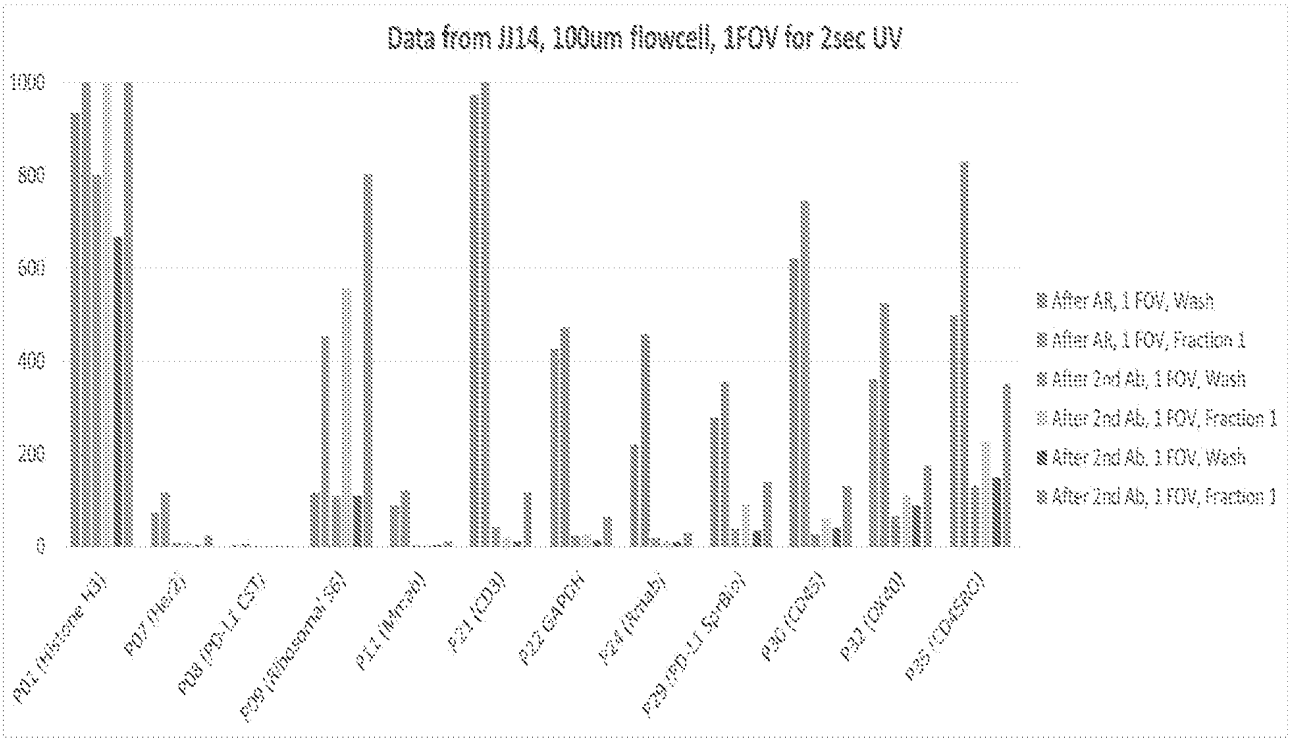

Data from JJ14, 100um flowcell, 1FOV for 2sec UV
1000
800
600
400
200
0
After AB, 1 FOV, Wash
After AB, 1 FOV, Fraction 1
After 2nd Ab, 1 FOV, Wash
After 2nd Ab, 1 FOV, Fraction 1
After 2nd Ab, 1 FOV, Wash
After 2nd Ab, 1 FOV, Fraction 1
P01: (Histone H3)
P07 (Ms-Ki)
P08 (PD-L1 CST)
P09 (Ribosomal S6)
P11 (Ki-67)
P21 (CD3)
P22: (S4 PDH
P24 (Smab)
P29 (PD-L1 SpvBio)
P30 (CD45)
P32 (CK+AE)
P35: (CD45RO)

Figure 43B

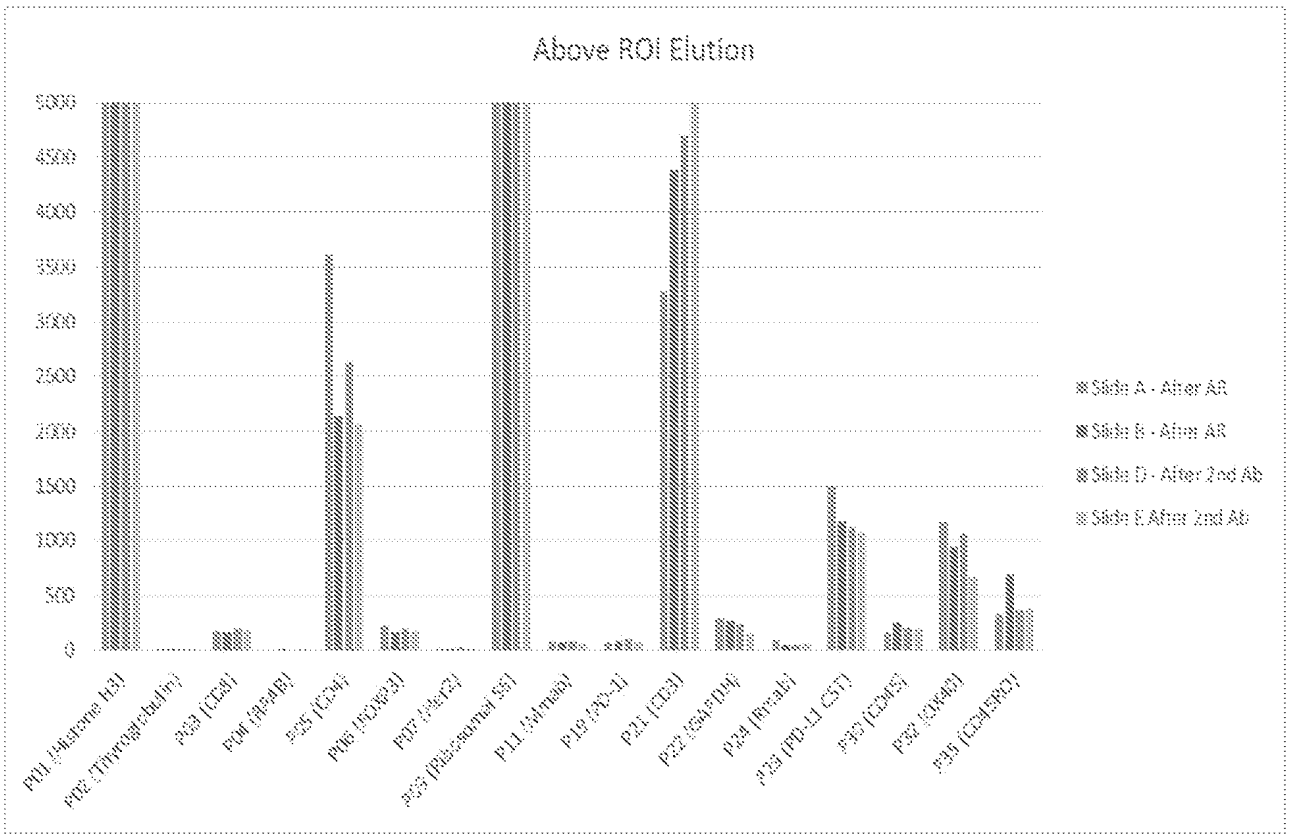

Above ROI Elution
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Slide A - After AR
Slide B - After AR
Slide D - After 2nd Ab
Slide E After 2nd Ab
P01 (Melanoma HMB)
P02 (Thyroglobulin)
P03 (CD8)
P04 (MelA)
P05 (CD4)
P06 (FOXP3)
P07 (PanCK)
P08 (PR/Stromal SP1)
P11 (Survivin)
P19 (PD-L1)
P21 (CD3)
P22 (SOX10)
P24 (Ki-67)
P29 (PD-L1 CST)
P30 (CD68)
P32 (CD163)
P35 (CD45RO)

Figure 44

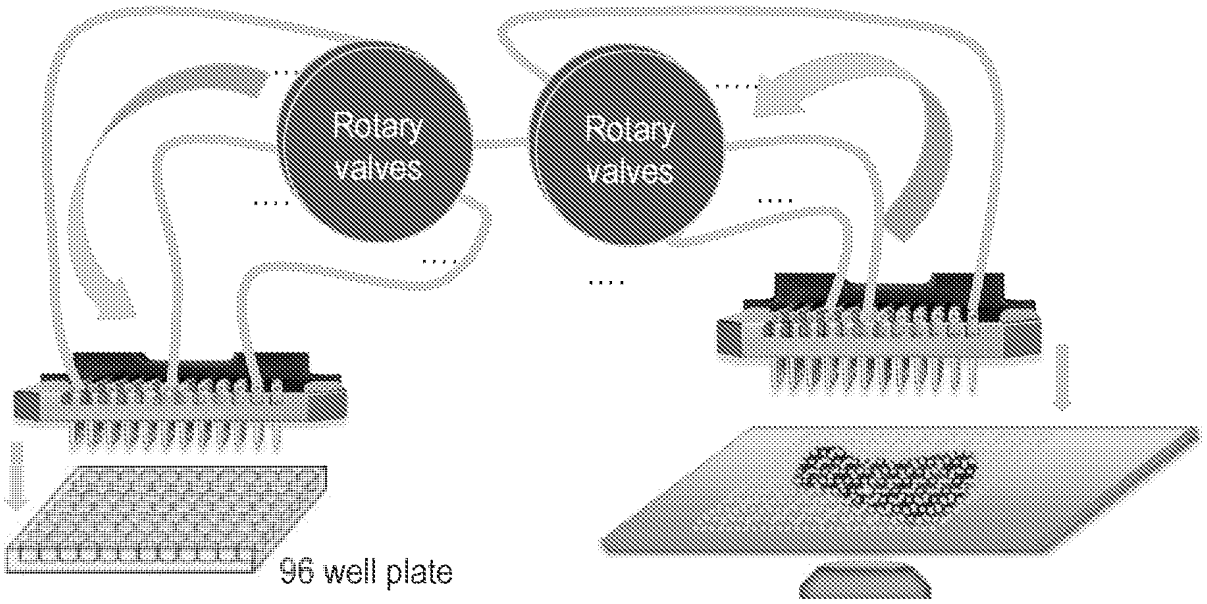


Figure 45

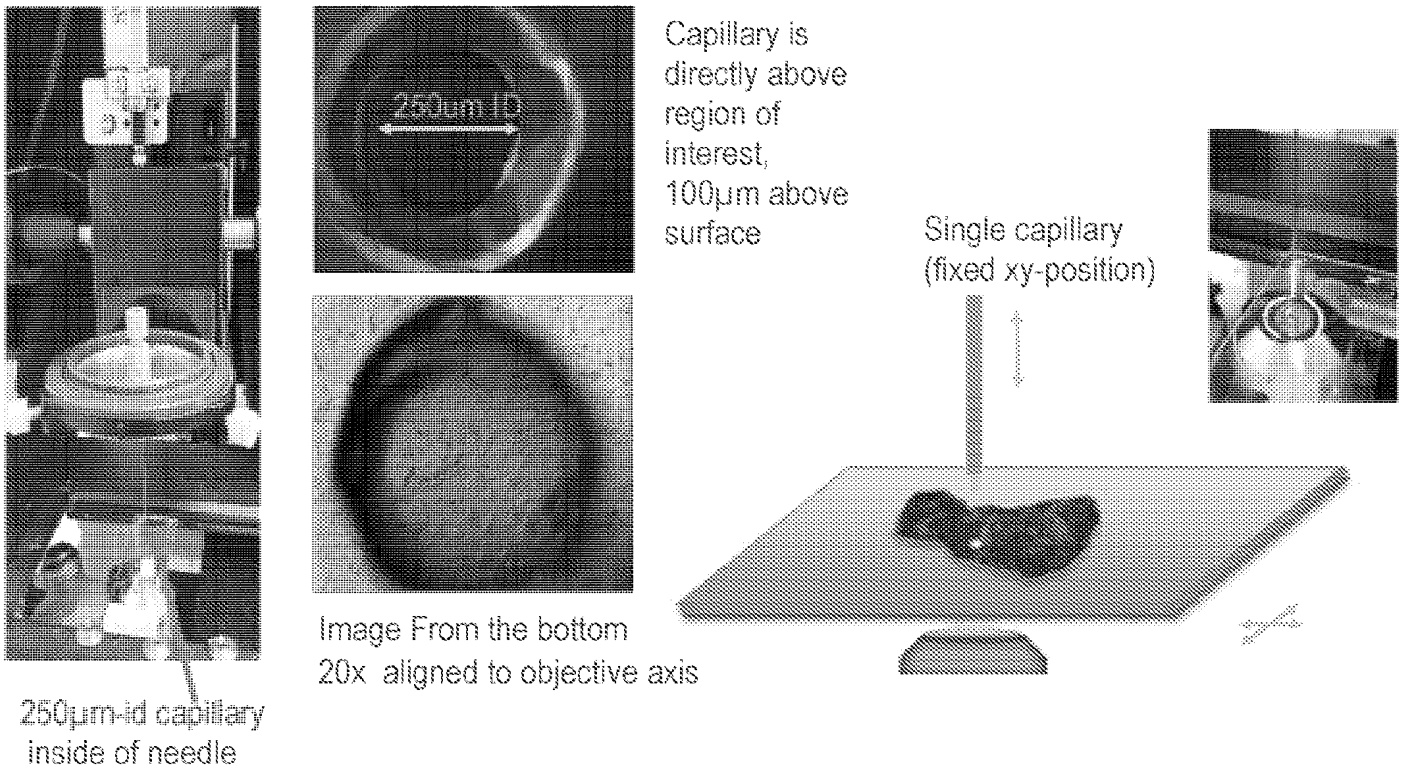

250µm ID
Capillary is directly above region of interest, 100µm above surface
Single capillary (fixed xy-position)
Image From the bottom 20x aligned to objective axis
250µm-id capillary inside of needle

Figure 46A

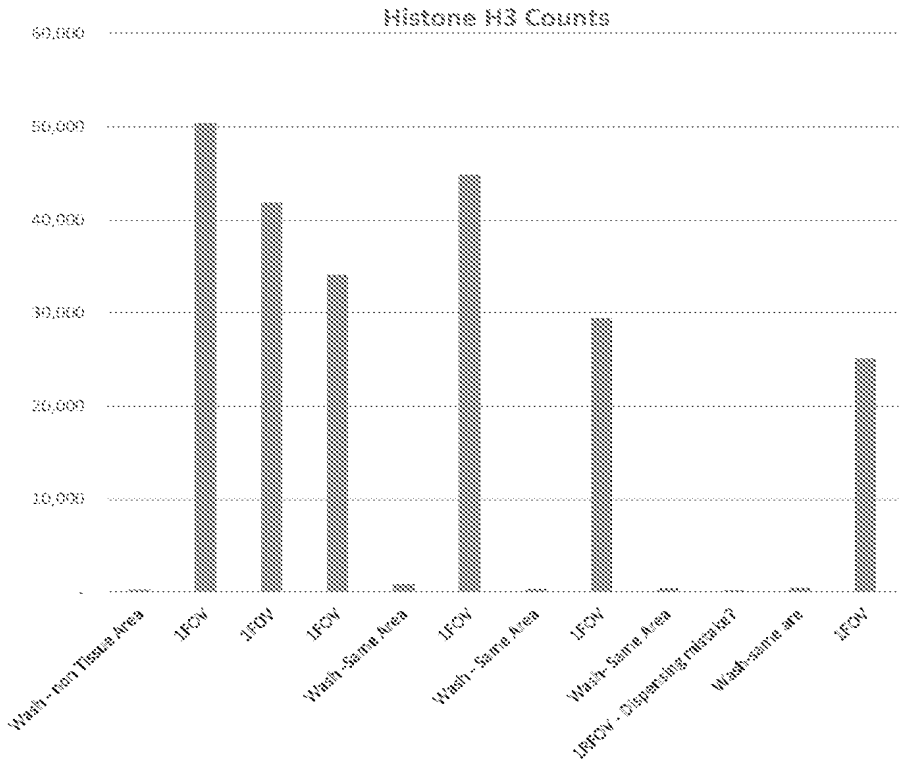

Histone H3 Counts
60,000
50,000
40,000
30,000
20,000
10,000
-
Wash - non Tissue Area
1FOV
1FOV
1FOV
Wash -Same Area
1FOV
Wash - Same Area
1FOV
Wash- Same Area
1FOV - Dispersing material?
Wash-same are
1FOV

Figure 46B

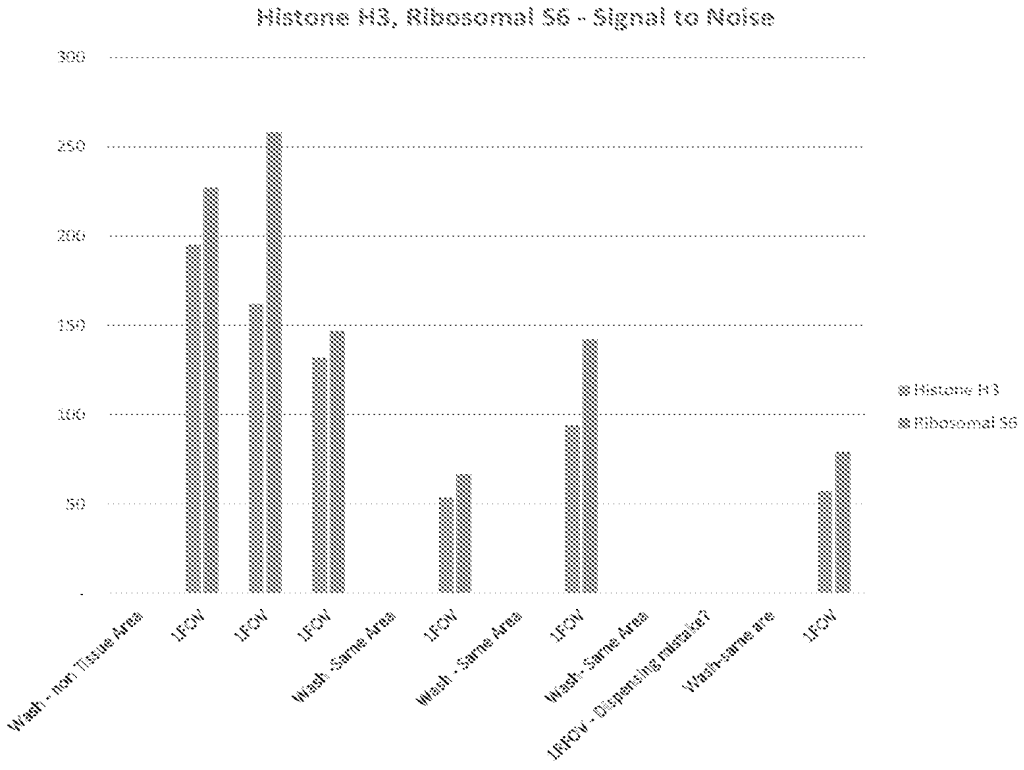

Histone H3, Ribosomal S6 - Signal to Noise
300
250
200
150
100
50
-
Histone H3
Ribosomal S6
Wash - non Tissue Area
1FOV
1FOV
1FOV
Wash - Same Area
1FOV
Wash - Same Area
1FOV
Wash - Same Area
1FOV - Dispensing mistake?
Wash-same are
1FOV

Figure 47

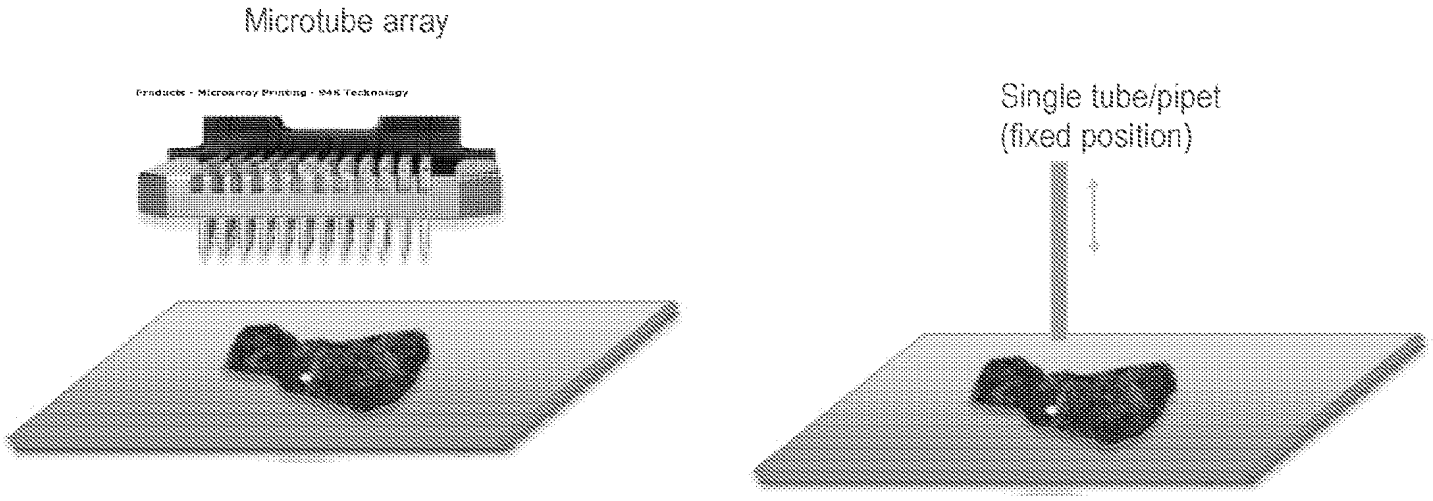


Figure 48

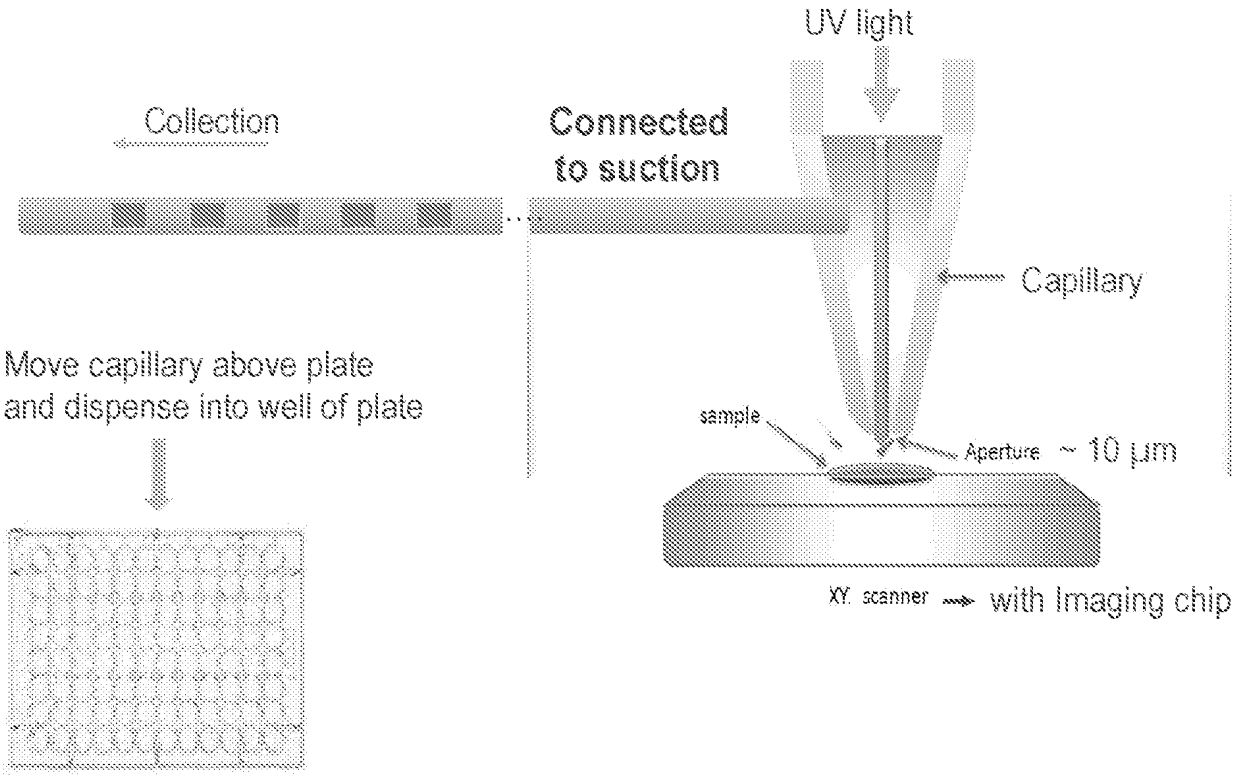

UV light
Collection
Connected to suction
Capillary
Move capillary above plate and dispense into well of plate
sample
Aperture
~ 10 µm
XY scanner
with Imaging chip

Figure 49

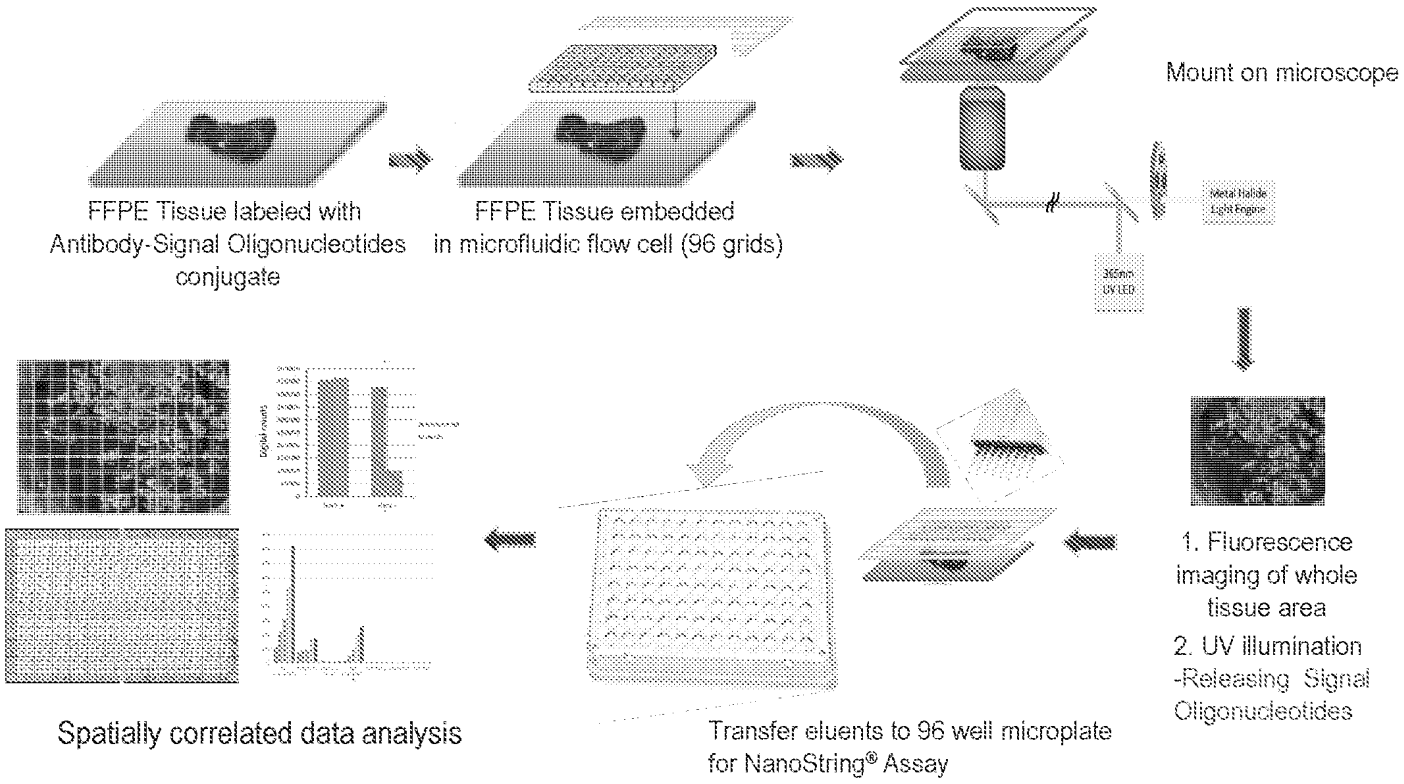

FFPE Tissue labeled with Antibody-Signal Oligonucleotides conjugate
FFPE Tissue embedded in microfluidic flow cell (96 grids)
Mount on microscope
Metal Halide Light Engine
365nm UV LED
1. Fluorescence imaging of whole tissue area
2. UV illumination -Releasing Signal Oligonucleotides
Transfer eluents to 96 well microplate for NanoString® Assay
Spatially correlated data analysis

Figure 50

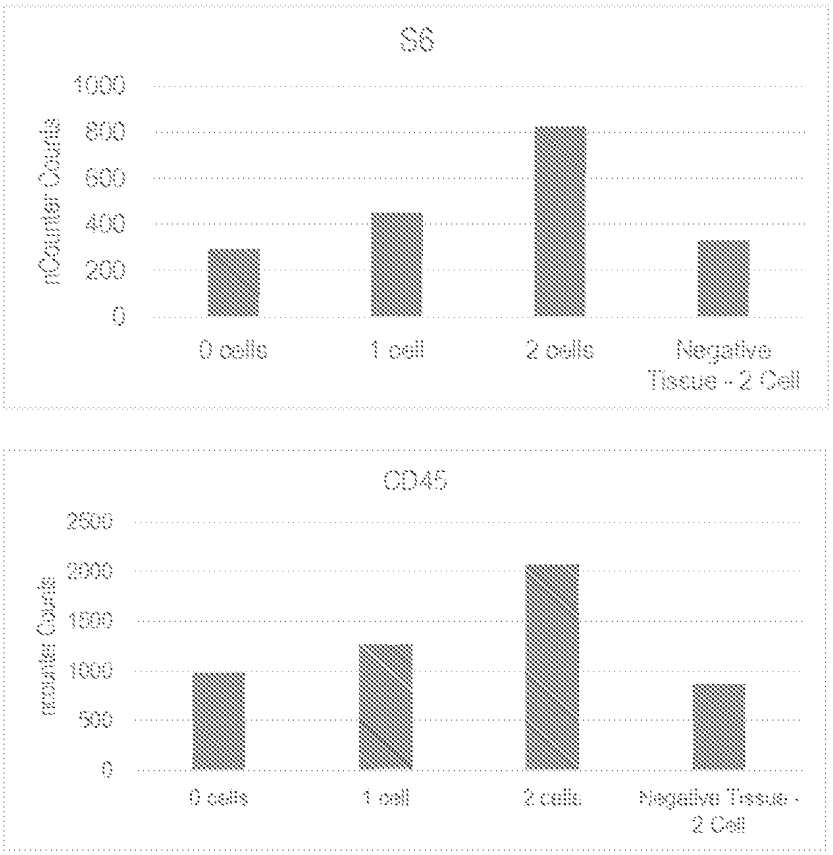

S6
nCounter Counts
1000
800
600
400
200
0
0 cells
1 cell
2 cells
Negative Tissue - 2 Cell
CD45
nCounter Counts
2500
2000
1500
1000
500
0
0 cells
1 cell
2 cells
Negative Tissue - 2 Cell

Figure 51

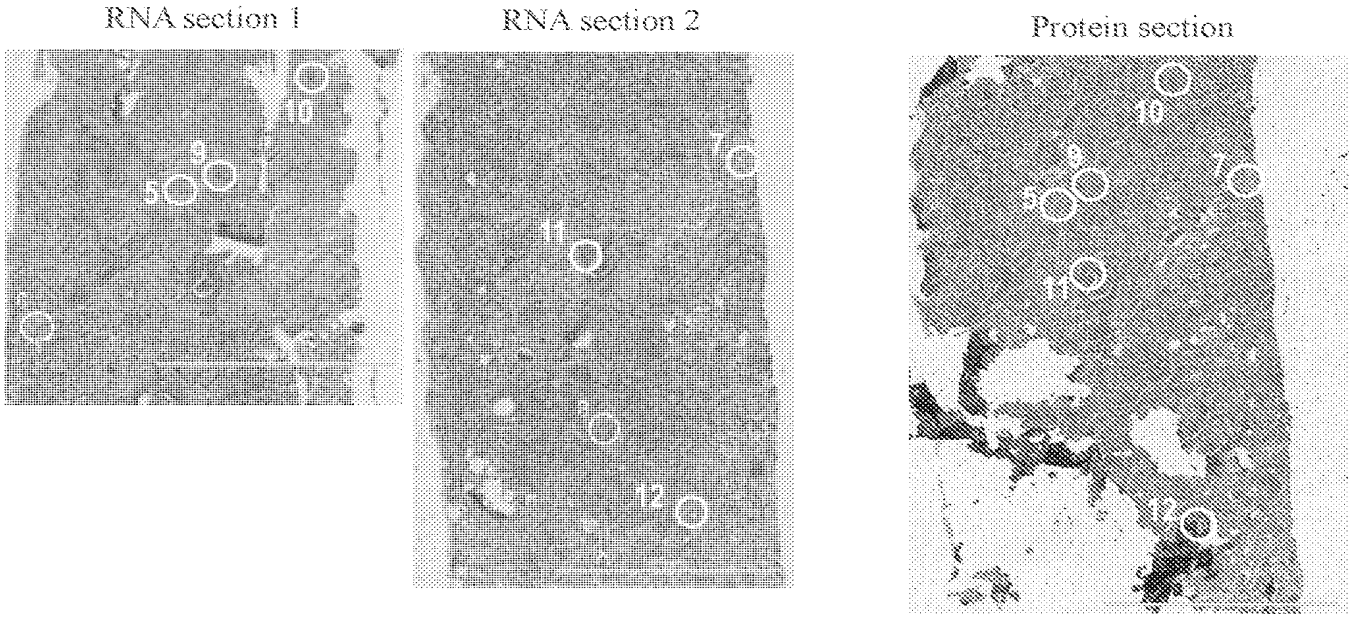


Figure 52

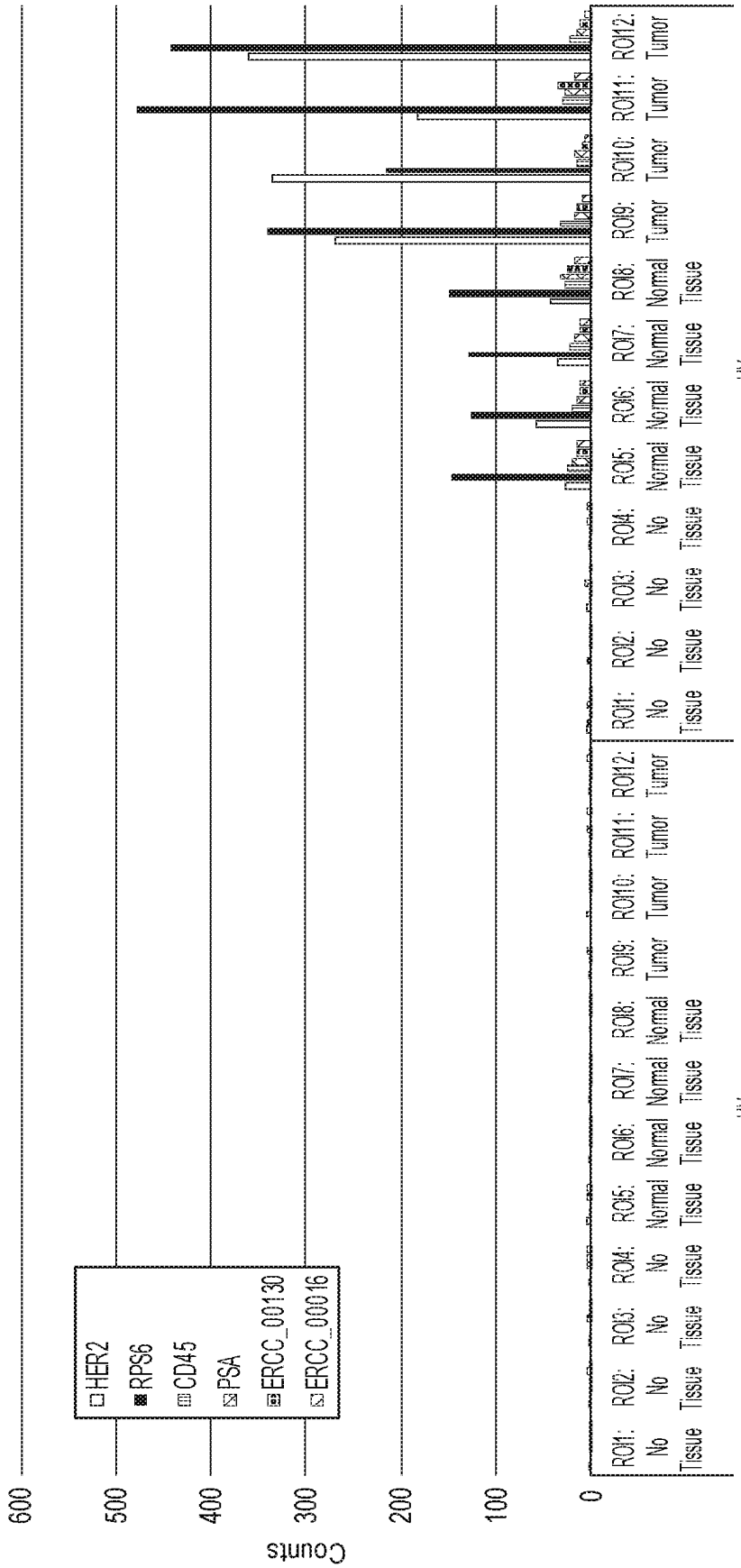

RNA Probe Counts in HER2 3+ Breast Tissue
HER2
RPS6
CD45
PSA
ERCC_00130
ERCC_00016
Counts
600
500
400
300
200
100
0
ROI1: No Tissue
ROI2: No Tissue
ROI3: No Tissue
ROI4: No Tissue
ROI5: Normal Tissue
ROI6: Normal Tissue
ROI7: Normal Tissue
ROI8: Normal Tissue
ROI9: Tumor Tissue
ROI10: Tumor Tissue
ROI11: Tumor Tissue
ROI12: Tumor Tissue
-UV
+UV

Figure 53

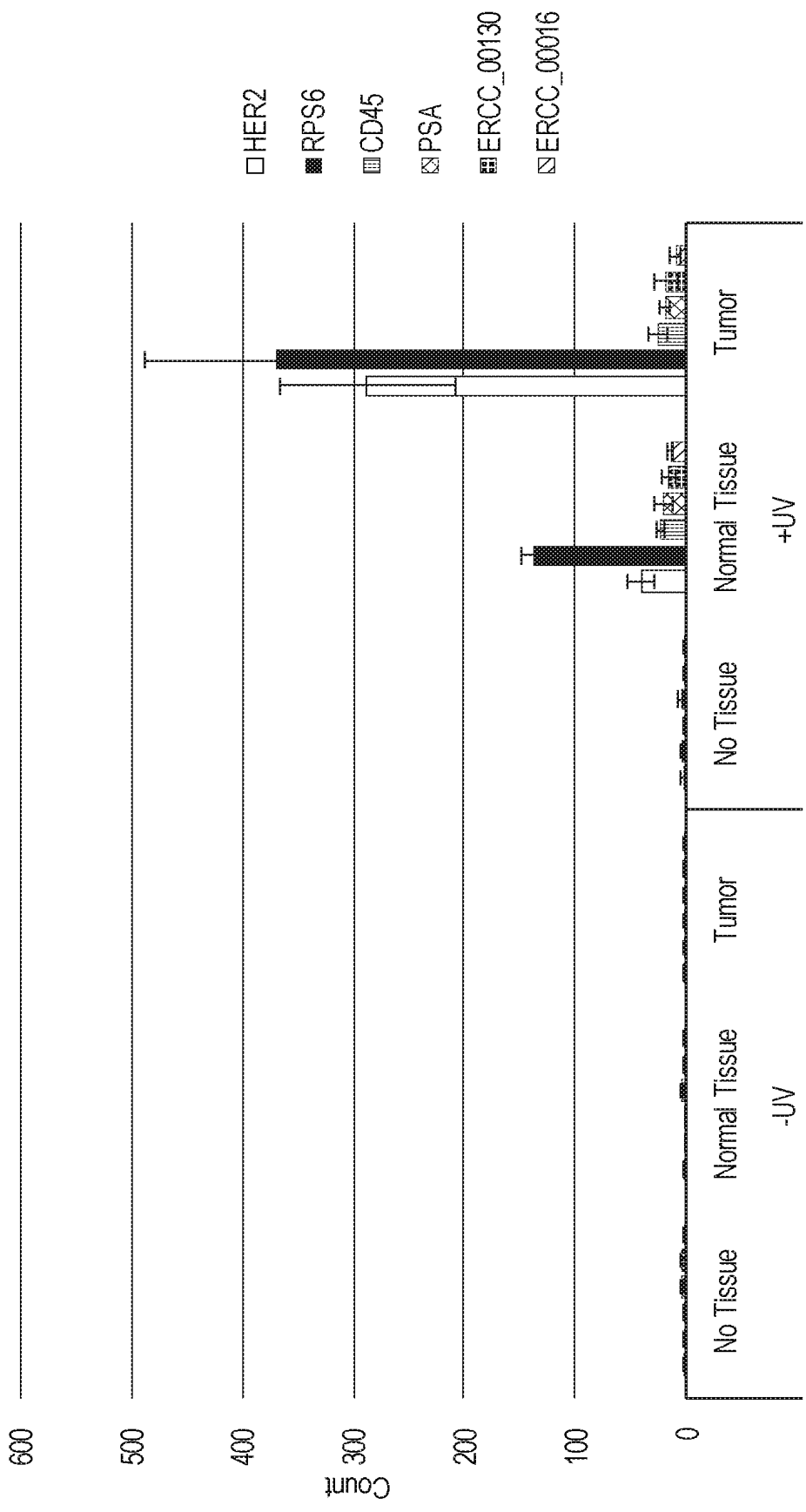

RNA Probe Counts in HER2 3+ Breast
Count
600
500
400
300
200
100
0
HER2
RPS6
CD45
PSA
ERCC_00130
ERCC_00016
No Tissue
Normal Tissue
Tumor
-UV
No Tissue
Normal Tissue
Tumor
+UV

Figure 54

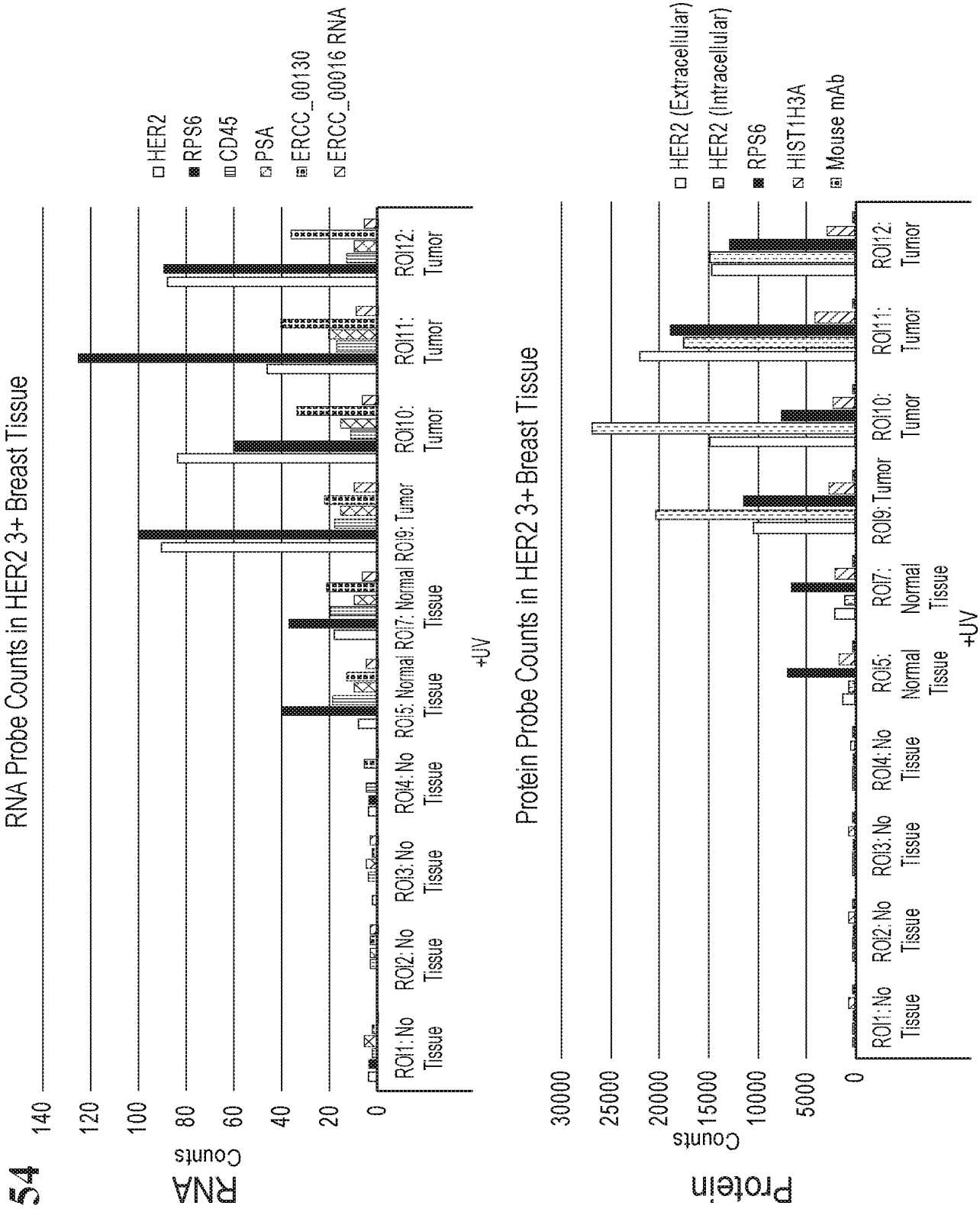

RNA Probe Counts in HER2 3+ Breast Tissue
RNA Counts
HER2
RPS6
CD45
PSA
ERCC_00130
ERCC_00016 RNA
ROI1: No Tissue
ROI2: No Tissue
ROI3: No Tissue
ROI4: No Tissue
ROI5: Normal Tissue
ROI7: Normal Tissue
ROI9: Tumor
ROI10: Tumor
ROI11: Tumor
ROI12: Tumor
+UV
Protein Probe Counts in HER2 3+ Breast Tissue
Protein Counts
HER2 (Extracellular)
HER2 (Intracellular)
RPS6
HIST1H3A
Mouse mAb
+UV

Figure 55

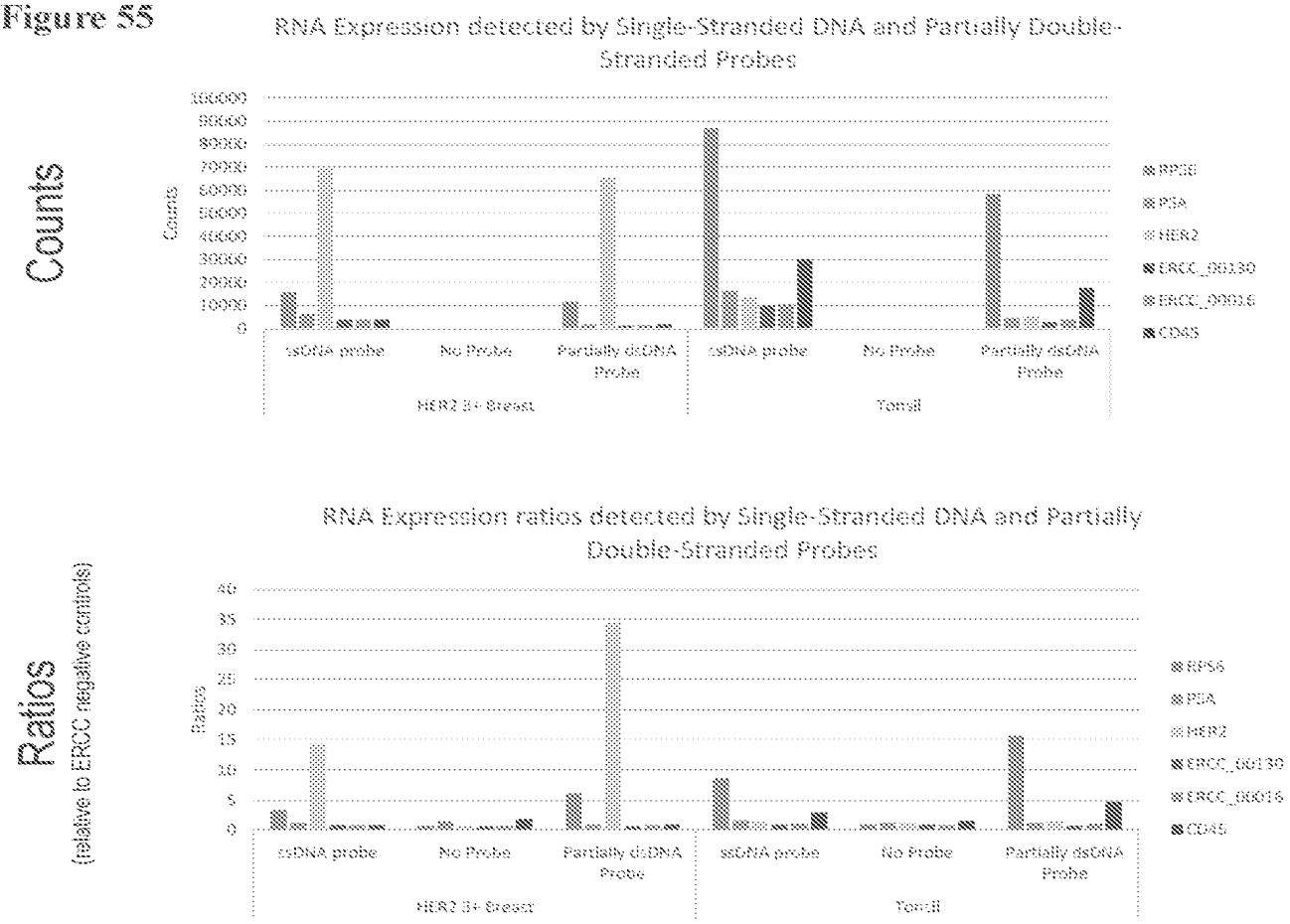

RNA Expression detected by Single-Stranded DNA and Partially Double-Stranded Probes
Counts
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
ssDNA probe
No Probe
Partially dsDNA Probe
ssDNA probe
No Probe
Partially dsDNA Probe
HER2 3+ Breast
Tonsil
RPS6
PSA
HER2
ERCC_00130
ERCC_00016
CD45
RNA Expression ratios detected by Single-Stranded DNA and Partially Double-Stranded Probes
Ratios
(relative to ERCC negative controls)
40
35
30
25
20
15
10
5
0
ssDNA probe
No Probe
Partially dsDNA Probe
ssDNA probe
No Probe
Partially dsDNA Probe
HER2 3+ Breast
Tonsil
RPS6
PSA
HER2
ERCC_00130
ERCC_00016
CD45

Figure 56

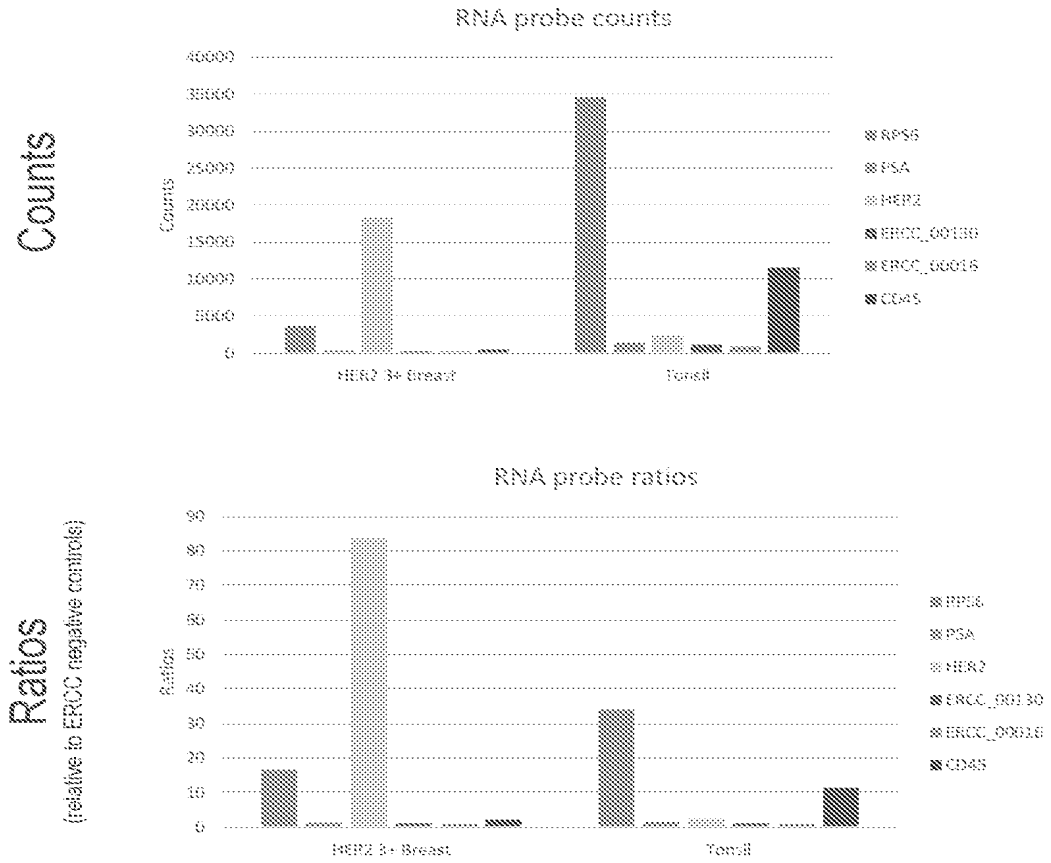

RNA probe counts
Counts
40000
35000
30000
25000
20000
15000
10000
5000
0
HER2 3+ Breast
Tonsil
RPS6
PSA
HER2
ERCC_00130
ERCC_00018
CD45
RNA probe ratios
Ratios
(relative to ERCC negative controls)
90
80
70
60
50
40
30
20
10
0
HER2 3+ Breast
Tonsil
RPS6
PSA
HER2
ERCC_00130
ERCC_00018
CD45

Figure 57

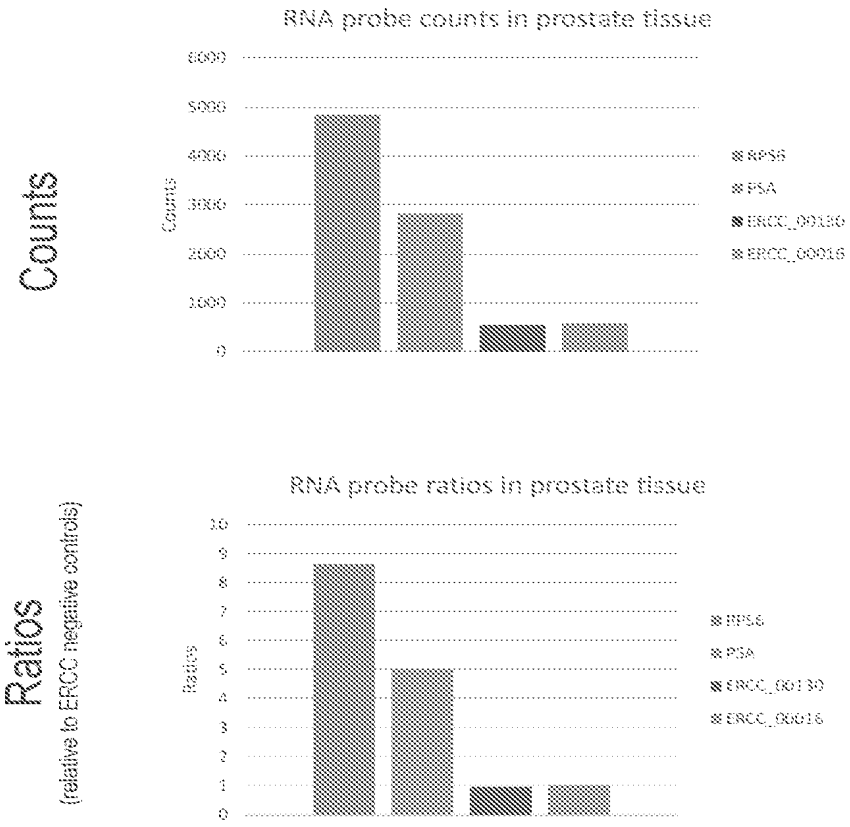
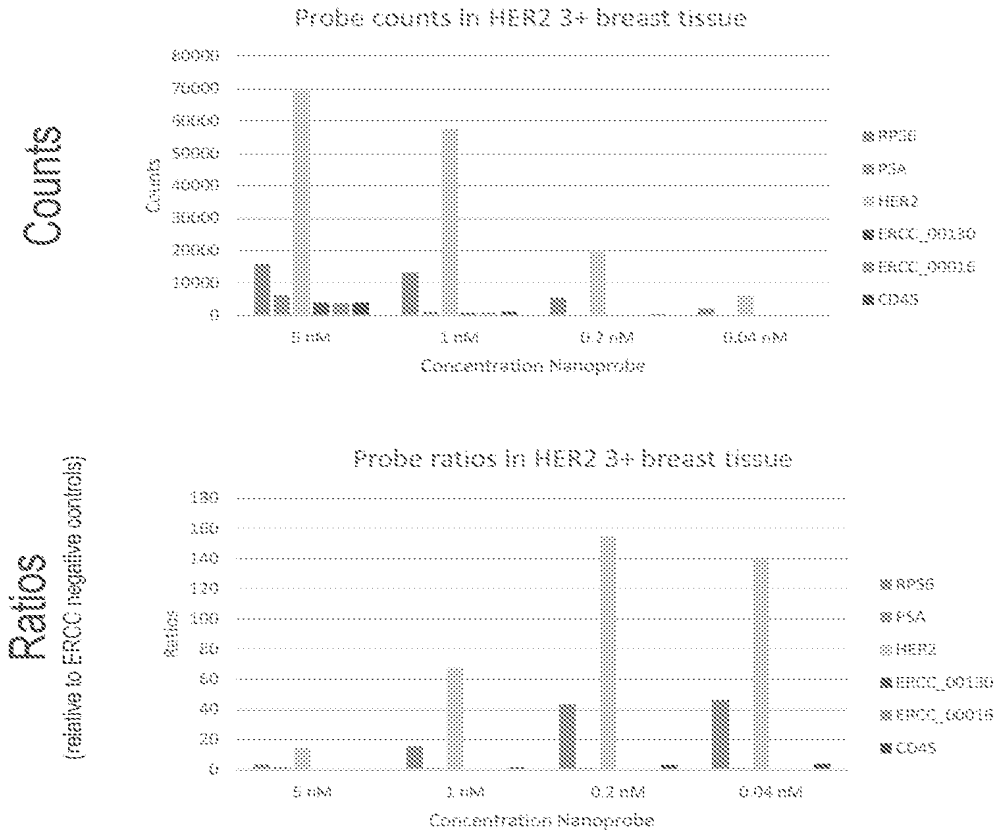

RNA probe counts in prostate tissue
Counts
6000
5000
4000
3000
2000
1000
0
RPS6
PSA
ERCC_00130
ERCC_00016
RNA probe ratios in prostate tissue
Ratios
(relative to ERCC negative controls)
10
9
8
7
6
5
4
3
2
1
0
RPS6
PSA
ERCC_00130
ERCC_00016

Figure 58


Counts
Probe counts in HER2 3+ breast tissue
Counts
80000
70000
60000
50000
40000
30000
20000
10000
0
5 nM
1 nM
0.2 nM
0.04 nM
Concentration Nanoprobe
RPS6
PSA
HER2
ERCC_00130
ERCC_00016
CD45
Ratios
(relative to ERCC negative controls)
Probe ratios in HER2 3+ breast tissue
Ratios
180
160
140
120
100
80
60
40
20
0
5 nM
1 nM
0.2 nM
0.04 nM
Concentration Nanoprobe
RPS6
PSA
HER2
ERCC_00130
ERCC_00016
CD45

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/042460

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2008/038734 A1 (SORGE JOSEPH A [US] ET AL) 14 February 2008 (2008-02-14) para. 16, 134, 135, 176, 507, 526-530, 543-544, 611-613; fig. 4	1-4,14, 15, 17-19, 22,27, 30-34
X	US 2011/151451 A1 (LEMAIRE REMI [FR] ET AL) 23 June 2011 (2011-06-23) para. 1, 10, 14-24, 31, 38, 41-49, 162, 178; fig. 1, 2	1-53
X,P	WO 2015/128272 A2 (VENTANA MED SYST INC [US]; HOFFMANN LA ROCHE [CH]) 3 September 2015 (2015-09-03) the whole document	1-53

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 October 2016	14/10/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Ripaud, Leslie

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011/245111 A1 (CHEE MARK S [US]) 6 October 2011 (2011-10-06) para. 7-9, 14, 15, 45	1-53
A	Crosetto: "Spatially resolved transcriptomics and beyond", Nature Review Genetics, 1 January 2015 (2015-01-01), pages 57-66, XP055284555, Retrieved from the Internet: URL:http://www.nature.com/nrg/journal/v16/n1/pdf/nrg3832.pdf [retrieved on 2016-06-29] the whole document	1-53
A	GEISS GARY K ET AL: "Direct multiplexed measurement of gene expression with color-coded probe pairs", NATURE BIOTECHNOLOGY, GALE GROUP INC, vol. 26, no. 3, 1 March 2008 (2008-03-01), pages 317-325, XP002505107, ISSN: 1087-0156, DOI: 10.1038/NBT1385 the whole document	1-53
A	WO 2014/200767 A1 (GEN HOSPITAL CORP [US]) 18 December 2014 (2014-12-18) the whole document	1-53
A	THIERY-LAVENANT GWENDOLINE ET AL: "Targeted Multiplex Imaging Mass Spectrometry in Transmission Geometry for Subcellular Spatial Resolution", JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY, ELSEVIER SCIENCE INC, US, vol. 24, no. 4, 9 February 2013 (2013-02-09), pages 609-614, XP035354275, ISSN: 1044-0305, DOI: 10.1007/S13361-012-0563-Z [retrieved on 2013-02-09] the whole document	1-53

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2016/042460

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2008038734 A1	14-02-2008	AU 2008210549 A1	07-08-2008
		CA 2680347 A1	07-08-2008
		EP 2121986 A2	25-11-2009
		JP 2010525792 A	29-07-2010
		US 2008038734 A1	14-02-2008
		WO 2008094902 A2	07-08-2008
US 2011151451 A1	23-06-2011	AT 520022 T	15-08-2011
		AT 528644 T	15-10-2011
		CA 2611266 A1	04-01-2007
		CA 2611297 A1	18-01-2007
		EP 1889047 A2	20-02-2008
		EP 1891426 A1	27-02-2008
		EP 2287602 A1	23-02-2011
		EP 2322920 A1	18-05-2011
		EP 2322921 A1	18-05-2011
		EP 2325631 A1	25-05-2011
		ES 2371203 T3	28-12-2011
		JP 5220593 B2	26-06-2013
		JP 2008542783 A	27-11-2008
		JP 2008542784 A	27-11-2008
		US 2008203289 A1	28-08-2008
		US 2011151451 A1	23-06-2011
		WO 2007000669 A2	04-01-2007
		WO 2007007192 A1	18-01-2007
WO 2015128272 A2	03-09-2015	AU 2015222268 A1	14-07-2016
		CA 2940764 A1	03-09-2015
		WO 2015128272 A2	03-09-2015
US 2011245111 A1	06-10-2011	AU 2011237729 A1	25-10-2012
		CA 2794522 A1	13-10-2011
		CN 102834526 A	19-12-2012
		CN 105385756 A	09-03-2016
		DK 2556171 T3	14-12-2015
		EP 2556171 A1	13-02-2013
		ES 2555106 T3	29-12-2015
		HK 1182143 A1	26-08-2016
		HR P20151294 T1	11-03-2016
		JP 5893607 B2	23-03-2016
		JP 2013523151 A	17-06-2013
		KR 20130100051 A	09-09-2013
		PT 2556171 E	21-12-2015
		SI 2556171 T1	31-03-2016
		SM T201500301 B	25-02-2016
		US 2011245111 A1	06-10-2011
		WO 2011127099 A1	13-10-2011
WO 2014200767 A1	18-12-2014	AU 2014278537 A1	07-01-2016
		CA 2915033 A1	18-12-2014
		EP 3008201 A1	20-04-2016
		JP 2016525344 A	25-08-2016
		WO 2014200767 A1	18-12-2014

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	Not Yet Assigned

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT
UNDER 37 C.F.R. §§ 1.56, 1.97, AND 1.98

In accordance with the duty of disclosure set forth in 37 C.F.R. §1.56, Applicant hereby submits the following information in conformance with 37 C.F.R. §§1.97 and 1.98. It is respectfully requested that the information be expressly considered during the prosecution of this application, and the references be made of record therein and appear among the "References Cited" on any patent to issue therefrom.

[X] Pursuant to 37 C.F.R. §1.98, a copy of each non-US patent document cited in the attached Form used in lieu of PTO/SB/08 is enclosed.

[] No copies of the publications listed on the attached Form used in lieu of PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98(d) because the publications were previously cited by or submitted to the Office in prior Application Serial No(s). ____ to which the above-identified application claims priority under 35 U.S.C. §120.

[X] No copies of any U.S. patents or U.S. patent application publications listed on the attached Form used in lieu of PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98.

[] Publication(s) listed on the attached Form used in lieu of PTO/SB/08 were cited in a foreign search or examination report corresponding to application serial no. and mailed on .

[] Enclosed is a copy of a non-English publication(s) _____ Pursuant to §609 of the M.P.E.P., Applicant submits the attached foreign search or examination report, which cites such non-English language publication(s).

[] Enclosed is a copy of a non-English publication(s) _____ English language publication _____ (copy enclosed) claims priority from this non-English publication.

[] Enclosed is an explanation of non-English publication(s) _____ for which an English translation is not available.

[] Enclosed is an English translation of non-English publication(s) _____ cited in the attached Form used in lieu of PTO/SB/08.

[] Enclosed is a copy of pending patent Application Serial No. _____.

[X] Applicants wish to make the Examiner aware of the following co-pending U.S. applications:

US Patent Application No. 15/211,230, filed July 15, 2016.

US Patent Application No. 15/211,236, filed July 15, 2016.

[X] In accordance with **37 C.F.R. §1.97(b)**, no additional fee for submission of this Information Disclosure Statement is required, as it is filed within any one of the following time periods:

[] within three months from the filing date of this national application other than a CPA under 37 C.F.R. § 1.53(d);

[] within three months from the date of entry of the national stage as set forth in 37 C.F.R. §1.491 in this international application;

[X] before the mailing date of a first office action on the merits; or

[] before the mailing of a first office action after the filing of a request for continued examination under 37 C.F.R. § 1.114.

In accordance with 37 C.F.R. § 1.97(g), this Information Disclosure Statement shall not be construed as to mean that a search has been made.

In accordance with 37 C.F.R. § 1.97(h), the filing of this Information Disclosure Statement shall not be construed to be an admission that the information cited in the statement is, or is considered to be material to patentability as defined by 37 C.F.R § 1.56(b).

Remarks

It is respectfully requested that the Examiner consider the above-noted information and return an initialed copy of the attached Form used in lieu of PTO/SB/08 to the undersigned.

The Examiner is invited to contact the undersigned by telephone if it is felt that a telephone interview would advance the prosecution of the present application.

Applicant believes no fee is due with this response.

Dated: May 8, 2019

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (202) 842-7899

By: / Matthew Pavao /
Matthew Pavao, Ph.D.
Reg. No. 50,572
Robert Powers
Reg. No. 77,802

203107874 v1

Electronic Acknowledgement Receipt

EFS ID:	35948432
Application Number:	16272487
International Application Number:	
Confirmation Number:	6864
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Customer Number:	141429
Filer:	Matthew Pavao/Sylvana Householder
Filer Authorized By:	Matthew Pavao
Attorney Docket Number:	NATE-037001US 321329-2523
Receipt Date:	08-MAY-2019
Filing Date:	11-FEB-2019
Time Stamp:	10:25:37
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Information Disclosure Statement (IDS) Form (SB08)	NATE037001US_321329_2523_sb08.pdf	282259 f2bf499d55b2c5debe4714325ca6f36a7dac3268	no	7

Warnings:

Information:					
This is not an USPTO supplied IDS fillable form					
2	Foreign Reference	WO9714028A2.pdf	9196089 643afca95e74f9d4de04c1595b30ecb0fe2182b2	no	294
Warnings:					
Information:					
3	Foreign Reference	WO9707245A1.pdf	3451992 5825a05188d35a8605230ea7211c341e76383091	no	64
Warnings:					
Information:					
4	Foreign Reference	WO9918434A1.pdf	3754705 e478296a939eddd44c6b0873dbfeaf71095398f4	no	82
Warnings:					
Information:					
5	Foreign Reference	WO0073777A1.pdf	4967967 a4e8fd700dffe0ec13a4d220c55bb5fb0fd4370d	no	103
Warnings:					
Information:					
6	Foreign Reference	WO0100875A2.pdf	2415947 7fb3864dbea0fbe2c924ef8d207206625c242394	no	51
Warnings:					
Information:					
7	Foreign Reference	WO2012106385A2.pdf	6597382 e5a47a5fa2c9e4b0bff04e75a0ba104d799af561	no	107
Warnings:					
Information:					
8	Foreign Reference	WO2013122996A1.pdf	2275596 5a511faca8da707c1bfc2ed6c8d22e5233bfb0b6	no	45
Warnings:					
Information:					

9	Foreign Reference	WO2014200767A1.pdf	10143022 d4bcc1b341efc2a7482b3d062398da0f10a093c7	no	162
Warnings:					
Information:					
10	Foreign Reference	WO2015128272A2.pdf	9861474 d72c57355e4d65943e081427078aa9595de2abb2	no	143
Warnings:					
Information:					
11	Foreign Reference	WO2017015099A1.pdf	5679777 c7f2209ad816de4f7a0845d4fffbec07c34cfa4c	no	107
Warnings:					
Information:					
12	Non Patent Literature	Agasti.pdf	2714353 93cff69923469f9baa71eee328dc856b1c9b723d	no	4
Warnings:					
Information:					
13	Non Patent Literature	Alfano.pdf	18622239 10429d7057d027e0ae4673583552f76b055fdad8	no	342
Warnings:					
Information:					
14	Non Patent Literature	Angelo.pdf	1503596 8a7a047973fd01705a1930ea12dde9b9da582aa7	no	9
Warnings:					
Information:					
15	Non Patent Literature	Cesano.pdf	272593 143367b46d6ce2f44e1900a4f9d2d2bbc98cc1d2	no	5
Warnings:					
Information:					

16	Non Patent Literature	Chow.pdf	651047 a476d10a8de89a22e0e694f5f335a8707b8d53a4	no	7
Warnings:					
Information:					
17	Non Patent Literature	Crossetto.pdf	2286276 5691d2b881a226b0b53890f1a142df0e10ce7d5f	no	10
Warnings:					
Information:					
18	Non Patent Literature	Dierck.pdf	1207619 1e30a3bdfdf1bbf61b9117d42858ef94ca123358	no	8
Warnings:					
Information:					
19	Non Patent Literature	Drummond.pdf	1307779 bead31ffa29202d4d522c91e641f2e080ed0a0c8	no	8
Warnings:					
Information:					
20	Non Patent Literature	Ferguson.pdf	727661 8876f481103402794b5050c8b031da44177a913d	no	7
Warnings:					
Information:					
21	Non Patent Literature	Frei.pdf	1101091 105fb16866c7d8845680f16dfd9f16e6490bcaf7	no	9
Warnings:					
Information:					
22	Non Patent Literature	Geiss.pdf	1573709 f4e8d4d48043b36a2dbd1e42d82a5c30e139d570	no	10
Warnings:					
Information:					

23	Non Patent Literature	Gullberg.pdf	922504 028c1d89d338258abb8c7022c6a31af76337a4ec	no	5
Warnings:					
Information:					
24	Non Patent Literature	Lind.pdf	6600250 a756beadf1871b3367e266b79454d20e7244e3da	no	10
Warnings:					
Information:					
25	Non Patent Literature	PosterDigitalSpatialProfilingPowerPointPresentation.pdf	829927 87ee889a6409b5fa78c2a705fd107dac60b96ac8	no	1
Warnings:					
Information:					
26	Non Patent Literature	Steemers.pdf	787679 da86267d9651e268d40bf55b68504b60dedb0fa2	no	4
Warnings:					
Information:					
27	Non Patent Literature	ThieryLavenant.pdf	1272726 a738d8098aa8add6dae8b07fd9723d143b745fef	no	6
Warnings:					
Information:					
28	Non Patent Literature	Ulall.pdf	2613753 a2631e6f0b1775394f70dff40c09b3ffcca34491	no	22
Warnings:					
Information:					
29	Non Patent Literature	Werner.pdf	1157332 a6166dc34ba82a4253a2b9edf885698a7a7bc7a7	no	12
Warnings:					
Information:					

30	Non Patent Literature	Zinchuk.pdf	1696596 36f7b16e995775cd758aafd47bb7924a847052a2	no	11
Warnings:					
Information:					
31	Non Patent Literature	Zollinger_Abstract.pdf	1016073 b5dfb207771fc767306f842a870e614c82296aea	no	2
Warnings:					
Information:					
32	Transmittal Letter	NATE037001US_321329_2523_IDS_Transmittal.pdf	108213 a1b73203f17e498ace95584622aab35a7fa1344a	no	3
Warnings:					
Information:					
		Total Files Size (in bytes):	107599226		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/272,487	02/11/2019	Joseph M. Beechem	NATE-037001US 321329-2523	6864

141429 7590 06/28/2019

Cooley LLP/NanoString Technologies, Inc
1299 Pennsylvania Avenue NW
Suite 700
Washington, DC 20004

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
06/28/2019	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zIPPatentDocketingMailboxUS@cooley.com

UNITED STATES DEPARTMENT OF COMMERCE
U.S. Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450

APPLICATION NO./ CONTROL NO.	FILING DATE	FIRST NAMED INVENTOR/ PATENT IN REEXAMINATION	ATTORNEY DOCKET NO.
16/272,487	02/11/2019	Beechem et al.	NATE-037001US 321

	EXAMINER	
Cooley LLP/NanoString Technologies, Inc 1299 Pennsylvania Avenue NW Suite 700 Washington, DC 20004	FRANK LU	
	ART UNIT	PAPER
	1600	20190624

DATE MAILED: _____

Please find below and/or attached an Office communication concerning this application or proceeding.

Commissioner for Patents

The Petition for Color Drawings filed on 2/11/19 has been GRANTED.
Should there be any questions with respect to this decision, please contact Marianne Seidel by telephone at 571-272-0584.

/MARIANNE C SEIDEL/ Quality Assurance Specialist, Art Unit 1600	/ANNE M. GUSSOW/ Supervisory Patent Examiner, Art Unit 1600

PTO-90C (Rev.04-03)

Technology Center Petition Count Sheet

Application No.	Applicant(s)
16/272,487	Beechem et al.

Petition Code	Petition Type
☐ 601	Petition to Make Special, Age or Health (37 CFR 1.102, MPEP § 708.02)
☐ 604	Petition to Make Special, Accelerated Examination (37 CFR 1.102, MPEP § 708.02)
☐ 607	Petition to Make Special, For Reasons not Provided in 601-606 and 639-640 (37 CFR 1.102, MPEP § 708.02)
☐ 608	Petition to Reopen After a Board Decision (37 CFR 1.198)
☐ 609	Petition for Review of Final Restriction Requirement (37 CFR 1.144)
☐ 610	Petition Invoking Authority of Commissioner under 37 CFR 1.181 Not Specified in Codes 611-617
☐ 611	Petition Relating to Prematureness of Final Rejection (37 CFR 1.181, MPEP § 706.07)
☐ 612	Petition Relating to Refusal to Enter an Amendment (37 CFR 1.181 and 1.127, MPEP § 714.09)
☐ 613	Petition Relating to Withdraw of holding of Abandonment (37 CFR 1.137, MPEP § 711.03)
☐ 614	Petition Relating to Withdraw of holding of Abandonment (37 CFR 1.137, MPEP § 7.14.19)
☐ 615	Petition Relating to Formal Sufficiency / Propriety of Affidavits (37 CFR 1.131, 1.132, 1.608, MPEP § 608.04(C)
☐ 617	Petition Relating to Refusal to Enter an Amendment (37 CFR 1.312)
☐ 620	Petition for Extension of Time (37 CFR 1.136 (b))
☐ 622	Petition Concerning Appeal Application before Transfer of Jurisdiction to Board of Appeals (MPEP § 1206)
☐ 623	Petition for Second or Subsequent Suspension of Action (37 CFR 1.103, MPEP § 709)
☐ 634	Petition to Correct Inventorship in a Patent not in Interference (37 CFR 1.324, MPEP § 1481)
☐ 635	Petition to Correct Inventorship in a Patent Application (37 CFR 1.48)
☐ 636	Petition to Change Inventorship in a Patent (37 CFR 1.324)
☐ 637	Petition to Withdraw from Issue, before Payment of an Issue Fee (37 CFR 1.313(a))
☐ 639	Petition for Make Special Prospective Manufacture (37 CFR 1.102, MPEP § 708.02)
☐ 646	Petition for Make Special Accelerated Examination
☑ 690	Petition Related Color Drawings of Photographs
☐ 699	Petition for Matters Before Technology Center

Information Required for OPESS Processing

☑ GRANTED ☐ GRANTED-IN-PART ☐ DENIED ☐ DISMISSED

☑ Mail Decision and Enter Decision in Petition Database/PALM
☐ Copy Decision, Send to Scanning
☐ Fees Need to be Charged
☐ _____

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371(C) DATE	FIRST NAMED APPLICANT	ATTY. DOCKET NO./TITLE
16/272,487	02/11/2019	Joseph M. Beechem	NATE-037001US 321329-2523

CONFIRMATION NO. 6864

141429
Cooley LLP/NanoString Technologies, Inc
1299 Pennsylvania Avenue NW
Suite 700
Washington, DC 20004

PUBLICATION NOTICE

OC00000010426249

Title:BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

Publication No.US-2019-0249248-A1
Publication Date:08/15/2019

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Public Records Division. The Public Records Division can be reached by telephone at (571) 272-3150 or (800) 972-6382, by facsimile at (571) 273-3250, by mail addressed to the United States Patent and Trademark Office, Public Records Division, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently https://portal.uspto.gov/pair/PublicPair. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Managment, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	F. W. M. Lu
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1.	2001/0002315	05-31-2001	Schultz et al.	
	2.	2001/0007775	07-12-2001	Seul et al.	
	3.	2001/0023078	09-20-2001	Bawendi et al.	
	4.	2001/0029049	10-11-2001	Walt et al.	
	5.	2001/0034034	10-25-2001	Bruchez et al.	

213808694 v1

Examiner Signature:		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Electronic Acknowledgement Receipt

EFS ID:	37552525
Application Number:	16272487
International Application Number:	
Confirmation Number:	6864
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Customer Number:	141429
Filer:	Robert E. Powers/Sylvana Householder
Filer Authorized By:	Robert E. Powers
Attorney Docket Number:	NATE-037001US 321329-2523
Receipt Date:	24-OCT-2019
Filing Date:	11-FEB-2019
Time Stamp:	15:33:16
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	NATE-037-001US_321329_2523_IDSTransmittal.pdf	111577 ea1dcba21c481fa54b5001de6f1589a53ea5d5ec	no	3

Warnings:

Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	NATE-037-001US_321329_2523_sb08.pdf	201669 0779874992d7442937bf839214deb9873d2b2ff9	no	1
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
			Total Files Size (in bytes):	313246	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	F. W. M. Lu

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

VIA EFS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT
UNDER 37 C.F.R. §§ 1.56, 1.97, AND 1.98

In accordance with the duty of disclosure set forth in 37 C.F.R. §1.56, Applicant hereby submits the following information in conformance with 37 C.F.R. §§1.97 and 1.98. It is respectfully requested that the information be expressly considered during the prosecution of this application, and the references be made of record therein and appear among the "References Cited" on any patent to issue therefrom.

[] Pursuant to 37 C.F.R. §1.98, a copy of each non-US patent document cited in the attached Form used in lieu of PTO/SB/08 is enclosed.

[] No copies of the publications listed on the attached Form used in lieu of PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98(d) because the publications were previously cited by or submitted to the Office in prior Application Serial No(s). ____ to which the above-identified application claims priority under 35 U.S.C. §120.

[X] No copies of any U.S. patents or U.S. patent application publications listed on the attached Form used in lieu of PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98.

[] Publication(s) listed on the attached Form used in lieu of PTO/SB/08 were cited in a foreign search or examination report corresponding to application serial no. and mailed on .

[] Enclosed is a copy of a non-English publication(s) _____ Pursuant to §609 of the M.P.E.P., Applicant submits the attached foreign search or examination report, which cites such non-English language publication(s).

[] Enclosed is a copy of a non-English publication(s) _____ English language publication _____ (copy enclosed) claims priority from this non-English publication.

[] Enclosed is an explanation of non-English publication(s) _____ for which an English translation is not available.

[] Enclosed is an English translation of non-English publication(s) _____ cited in the attached Form used in lieu of PTO/SB/08.

[] Enclosed is a copy of pending patent Application Serial No. _____.

[X] Applicants wish to make the Examiner aware of the following co-pending U.S. applications:

US Patent Application No. 16/596,587, filed October 8, 2019

US Patent Application No. 16/596,596, filed October 8, 2019.

[X] In accordance with **37 C.F.R. §1.97(b)**, no additional fee for submission of this Information Disclosure Statement is required, as it is filed within any one of the following time periods:

[] within three months from the filing date of this national application other than a CPA under 37 C.F.R. § 1.53(d);

[] within three months from the date of entry of the national stage as set forth in 37 C.F.R. §1.491 in this international application;

[X] before the mailing date of a first office action on the merits; or

[] before the mailing of a first office action after the filing of a request for continued examination under 37 C.F.R. § 1.114.

In accordance with 37 C.F.R. § 1.97(g), this Information Disclosure Statement shall not be construed as to mean that a search has been made.

In accordance with 37 C.F.R. § 1.97(h), the filing of this Information Disclosure Statement shall not be construed to be an admission that the information cited in the statement is, or is considered to be material to patentability as defined by 37 C.F.R § 1.56(b).

Remarks

It is respectfully requested that the Examiner consider the above-noted information and return an initialed copy of the attached Form used in lieu of PTO/SB/08 to the undersigned.

The Examiner is invited to contact the undersigned by telephone if it is felt that a telephone interview would advance the prosecution of the present application.

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 50-1283, under Order No. NATE-037/001US 321329-2523 from which the undersigned is authorized to draw.

Dated: October 24, 2019

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (202) 842-7899

By: /Robert E. Powers/
Robert E. Powers, Ph.D.
Reg. No. 77,802
Matthew Pavao, Ph.D.
Reg. No. 50,572

213808873 v1

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	F. W. M. Lu
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1.	10,501,777	12-10-2019	Beechem et al.	
	2.	10,640,816	05-05-2020	Beechem et al.	
	3.	2020/0040385	02-06-2020	Beechem et al.	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (*if known*)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	4.	JP 2008-542783 A	11-27-2008	Universite des Sciences et Technologie		Corresponds to WO 2007/000669 A2
	5.	WO 2007/000669 A2	01-04-2007	Centre National de la Recherche Scientifique (CNRS) et al.		

232841174 v1

Examiner Signature:		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2008-542783
(P2008-542783A)
(43)公表日　平成20年11月27日(2008.11.27)

(51)Int.Cl.		F I		テーマコード（参考）
G01N 27/62	***(2006.01)***	G01N 27/62	Y	2G041
G01N 33/53	***(2006.01)***	G01N 33/53	Y	
G01N 27/64	***(2006.01)***	G01N 27/62	V	
		G01N 27/64	B	

審査請求 未請求　予備審査請求 未請求　（全 46 頁）

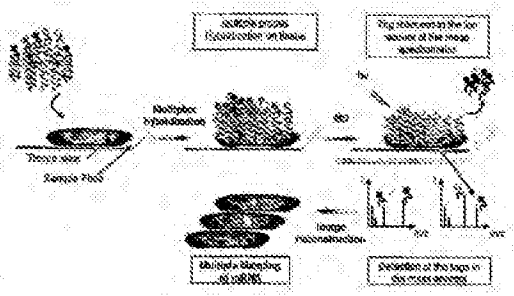
(21)出願番号　特願2008-515317 (P2008-515317)
(86)(22)出願日　平成18年6月7日 (2006.6.7)
(85)翻訳文提出日　平成20年2月7日 (2008.2.7)
(86)国際出願番号　PCT/IB2006/002309
(87)国際公開番号　WO2007/000669
(87)国際公開日　平成19年1月4日 (2007.1.4)
(31)優先権主張番号　60/687,848
(32)優先日　平成17年6月7日 (2005.6.7)
(33)優先権主張国　米国 (US)

(71)出願人　594016872
サントル、ナショナール、ド、ラ、ルシェルシュ、シアンティフィク、（セーエヌエルエス）
フランス国パリ、リュ、ミケ、ランジュ、3
(71)出願人　507402635
ユニベルシテ・デ・シヤーンス・エ・テクノロジー・ドゥ・リル
UNIVERSITE DES SCIENCES ET TECHNOLOGIES DELILLE
フランス国、エフ－５９６５０　ビルヌーブ・ダスク（番地なし）

最終頁に続く

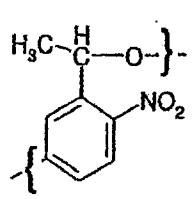
(54)【発明の名称】組織切片の質量分析法による分析のための光解離または断片化により開裂可能なリンカーとの接合体の使用

(57)【要約】
本発明は、少なくとも１種類の（Ａ－Ｘ）$_n$－Ｂ接合体（式中、Ａは既知分子量のタグ分子であり、Ｘは試料の脱離／イオン化の際に開裂されるリンカーであり、ｎは少なくとも１の整数であり、Ｂは上記タグ分子に特異的に結合する結合分子である）を用いて、組織切片における少なくとも１種類のターゲット分子マップを決定する方法に関する。ＭＡＬＤＩ質量分析法を用いるときには、上記リンカー分子Ｘは、上記ＭＡＬＤＩレーザーの波長で光開裂可能であるときには、試料のレーザー照射の際の光解離によって開裂することができる。あるいは、ＵＶ－ＭＡＬＤＩ、ＩＲ－ＭＡＬＤＩ、ＳＩＭＳまたはＤＥＳＩ質量分析法を用いるときには、上記リンカー分子Ｘは、試料の脱離／イオン化の際に断片化によって開裂することができる。

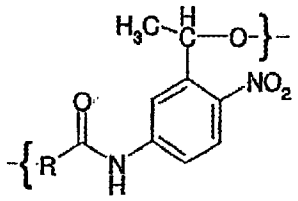


【特許請求の範囲】
【請求項１】
組織切片における少なくとも１種類のターゲット分子マップを決定する方法であって、
ａ）　上記組織切片を少なくとも１種類の（Ａ－Ｘ）$_n$－Ｂ接合体
（上記式中、
Ａは、既知分子量のタグ分子であり、
Ｘは、試料の脱離／イオン化中に開裂されるリンカーであり、
ｎは、少なくとも１の整数であり、
Ｂは、上記ターゲット分子に特異的に結合する結合分子であり、かつ
それぞれ異なるＢ分子が異なるＡタグ分子に連結される）
とハイブリダイズさせること；
ｂ）　組織切片表面を走査し、それぞれの隣接スポットを質量分析計で分析すること（ここで、上記リンカーＸは、試料の脱離／イオン化中に開裂され、それぞれのスポットで得られるデータは保存される）；および
ｃ）　それぞれの異なるタグ分子の分子質量ウインドウで得られたデータを解析して、異なる検討したターゲット分子の数と同じ数の組織切片のマップを作製すること
を含んでなる、方法。
【請求項２】
各ターゲット分子が、核酸、特にｍＲＮＡ分子、ペプチド、タンパク質、特に受容体およびリガンド、抗体、抗原、アプタマー、ハプテン、並びに有機化合物からなる群から独立して選択される、請求項１に記載の方法。
【請求項３】
少なくとも１種類のターゲット分子が、ｍＲＮＡ分子である、請求項２に記載の方法。
【請求項４】
少なくとも１種類のターゲット分子が、ペプチド、タンパク質、抗原またはハプテンである、請求項２に記載の方法。
【請求項５】
ターゲット分子に特異的に結合する各Ｂ分子が、核酸、特にオリゴヌクレオチド、ペプチド、タンパク質、特に受容体およびリガンド、抗体、抗原、ハプテン並びに有機化合物からなる群から独立して選択される、請求項１～４のいずれか一項に記載の方法。
【請求項６】
ターゲット核酸に特異的に結合する各Ｂ分子が、ターゲットｍＲＮＡ配列に相補的な配列を有する核酸プローブである、請求項５に記載の方法。
【請求項７】
ターゲットペプチド、タンパク質、抗原またはハプテンに特異的に結合する各Ｂ分子が、上記ペプチド、タンパク質、抗原またはハプテンに対して向けられた抗体である、請求項５に記載の方法。
【請求項８】
各Ａタグ分子が、ペプチド、核酸、糖類、ポリマー、脂質および有機化合物からなる群から選択される、請求項１～７のいずれか一項に記載の方法。
【請求項９】
少なくとも１種類のＡタグ分子が、ペプチドである、請求項８に記載の方法。
【請求項１０】
ＭＡＬＤＩ質量分析法を用い、Ｘリンカー分子がＭＡＬＤＩレーザーの波長で光開裂可能である、請求項１～９のいずれか一項に記載の方法。
【請求項１１】
Ｘリンカー分子が、

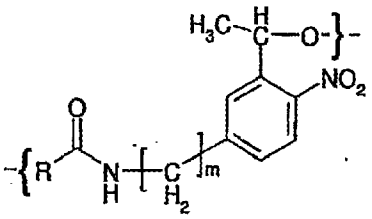
【化１】



,

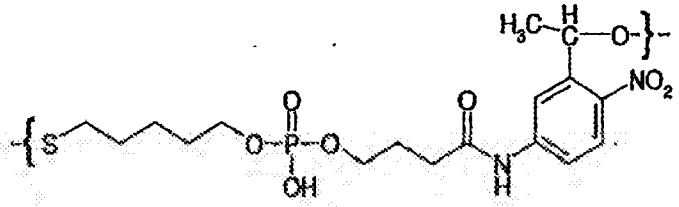


,



,

および

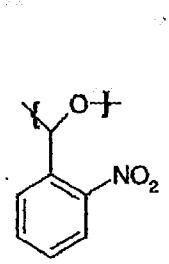


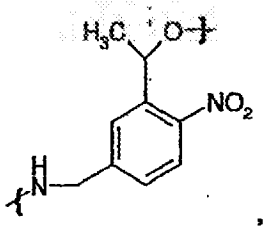
（上記式中、Ｒは、Ｃ１－Ｃ６アルキル基であり、ｍは、１～４の整数である）からなる群から選択される残基を含んでなる、請求項１０に記載の方法。

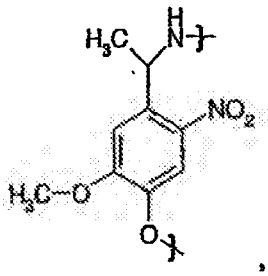
【請求項１２】

Ｘリンカー分子が、

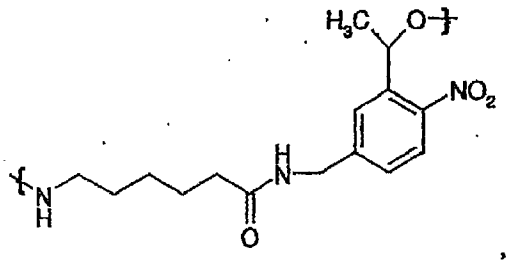
【化２】



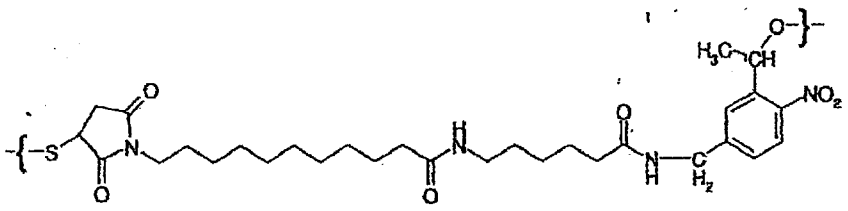


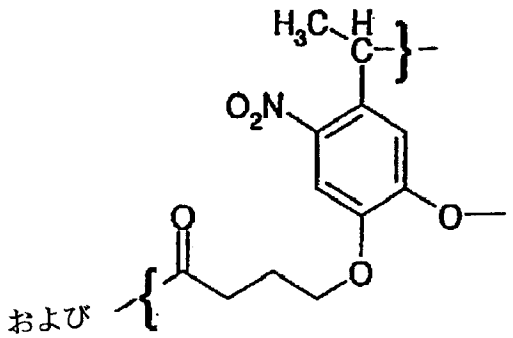


,



,



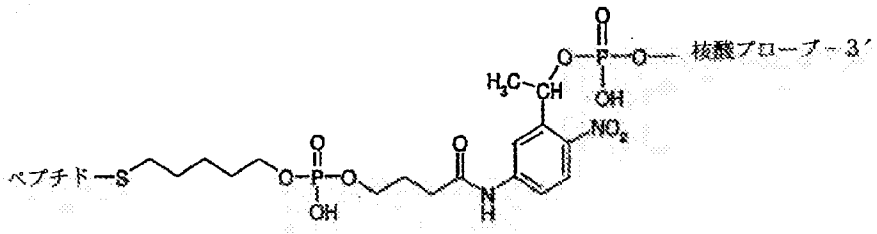


からなる群から選択される残基を含んでなる、請求項１０に記載の方法。

【請求項１３】

ターゲット分子が、ｍＲＮＡ分子であり、結合Ｂ分子が、ｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子が、ペプチドであり、ｎが、１であり、（Ａ－Ｘ）－Ｂ接合体が、下記の構造：

【化３】

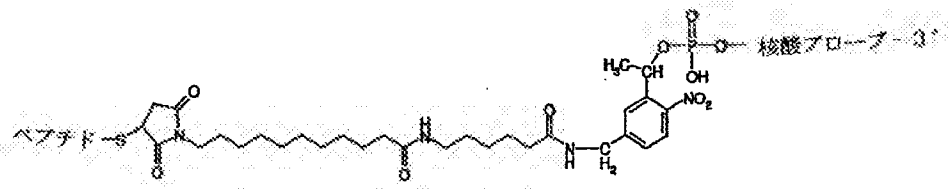


を有する、請求項１０に記載の方法。

【請求項１４】

ターゲット分子が、ｍＲＮＡ分子であり、結合Ｂ分子が、ｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子が、ペプチドであり、ｎが、１であり、（Ａ－Ｘ）－Ｂ接合体が、下記の構造：

【化４】

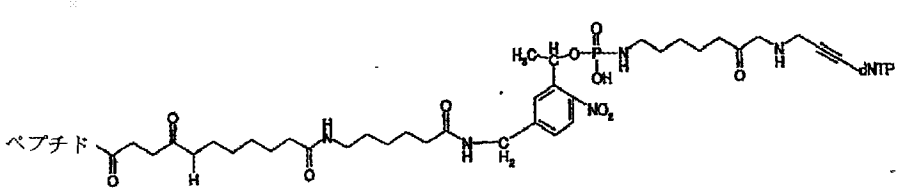


を有する、請求項１０に記載の方法。

【請求項１５】

ターゲット分子が、ｍＲＮＡ分子であり、結合Ｂ分子が、ｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子が、ペプチドであり、ｎが、１より大きく、核酸プローブが、下記の構造：

【化５】

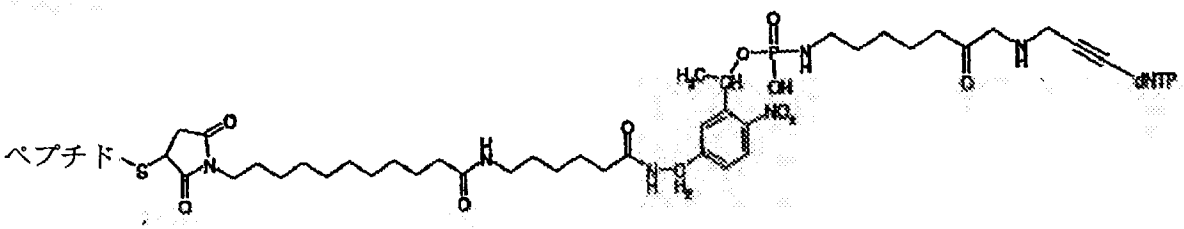


の少なくとも１種類の修飾塩基を含んでなる、請求項１０に記載の方法。

【請求項１６】

ターゲット分子が、ｍＲＮＡ分子であり、結合Ｂ分子が、ｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子が、ペプチドであり、ｎが、１より大きく、核酸プローブが、下記の構造：

【化６】



の少なくとも１種類の修飾塩基を含んでなる、請求項１０に記載の方法。

【請求項１７】

少なくとも５種類の異なるｍＲＮＡ分子が、組織切片上でマッピングされる、請求項１３～１６のいずれか一項に記載の方法。

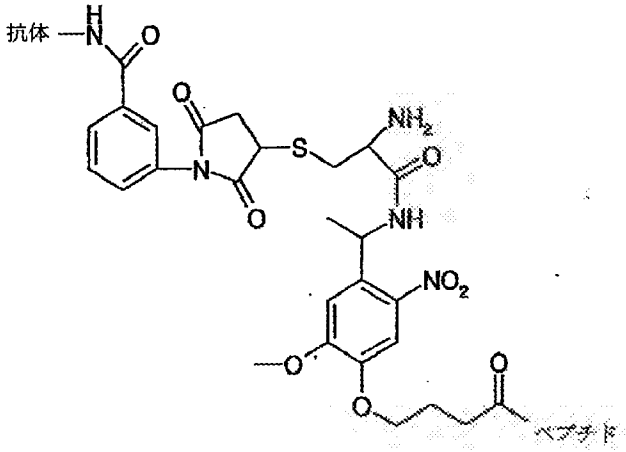
【請求項１８】

更に、組織切片のそれぞれに対応するタンパク質発現マップを作製するための、それぞれのｍＲＮＡ分子に対応するタンパク質の分子質量ウインドウで得られたデータを分析することからなる工程ｄ）を含んでなる、請求項１３～１７のいずれか一項に記載の方法。

【請求項１９】

ターゲット分子が、ペプチド、タンパク質、またはハプテンであり、結合Ｂ分子が、上記ターゲット分子に対して向けられた抗体であり、Ａタグ分子が、ペプチドであり、ｎが、１であり、（Ａ－Ｘ）－Ｂ接合体が、下記の構造：

【化７】



を有する、請求項１０に記載の方法。

【請求項２０】

上記Ｘリンカー分子が、試料のイオン化中に断片化により開裂される、請求項１～９のいずれか一項に記載の方法。

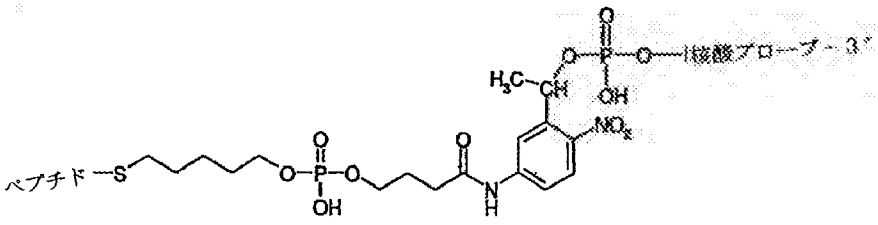
【請求項２１】

上記Ｘリンカー分子が、請求項１１または１２に定義された通りである、請求項２０に記載の方法。

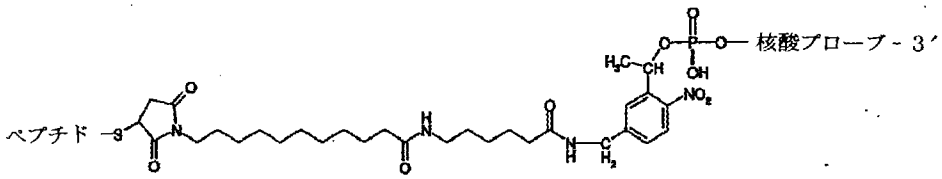
【請求項２２】

ターゲット分子が、ペプチド、タンパク質（抗体など）、抗原、またはハプテンであり、結合分子が、抗体または抗体断片であり、ｎが、１であり、タグ分子が、蛍光色素、酵

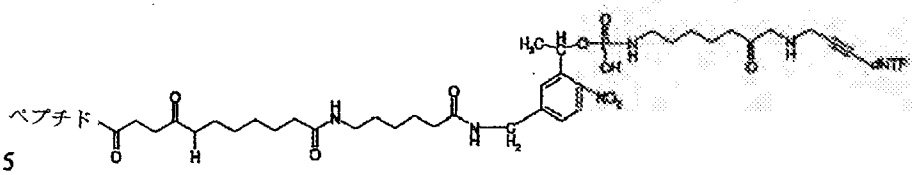
素、ビオチン、または金である、請求項２０に記載の方法。
【請求項２３】
ＵＶ－ＭＡＬＤＩ、ＩＲ－ＭＡＬＤＩ、ＳＩＭＳ、またはＤＥＳＩ質量分析法を用いる、請求項２０～２２のいずれか一項に記載の方法。
【請求項２４】
上記接合体が
【化８】



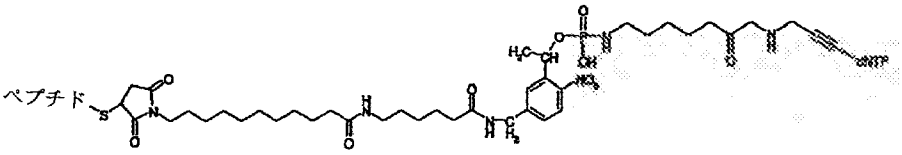
,



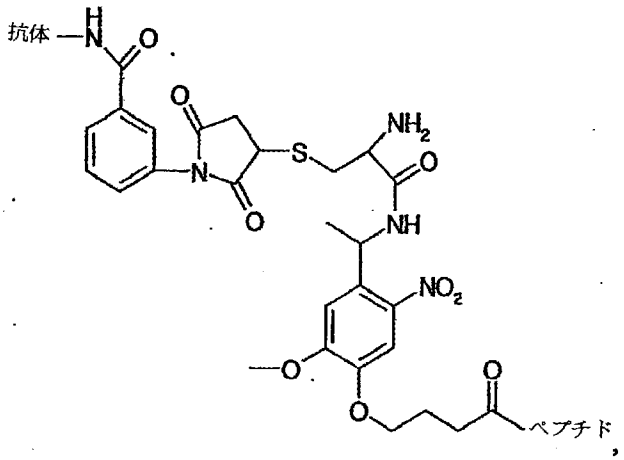
,



,

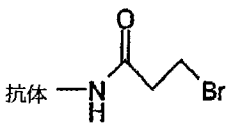


,



,

または



から選択される、請求項１、１０、または２０のいずれか一項に記載の方法での使用に適

した接合体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、少なくとも１種類の（Ａ－Ｘ）ｎ－Ｂ接合体（ｃｏｎｊｕｇａｔｅ）（式中、Ａは既知分子量のタグ分子であり、Ｘは試料の脱離／イオン化の際に開裂するリンカーであり、ｎは少なくとも１の整数であり、Ｂは上記タグ分子に特異的に結合する結合分子である）を用いて、組織切片における少なくとも１種類のターゲット分子マップを決定する方法に関する。ＭＡＬＤＩ質量分析法を用いるときには、上記リンカー分子Ｘは、上記ＭＡＬＤＩレーザーの波長で光開裂可能であるときには、試料のレーザー照射の際の光解離によって開裂することができる。あるいは、ＵＶ－ＭＡＬＤＩ、ＩＲ－ＭＡＬＤＩ、ＳＩＭＳまたはＤＥＳＩ質量分析法を用いるときには、上記リンカー分子Ｘは、試料の脱離／イオン化の際に断片化によって開裂することができる。
【０００２】
最近、トランスクリプトームおよびプロテオームの研究により、幾つかの種類の癌等の多種多様な疾患に関与する多くのタンパク質が同定されてきた。
【０００３】
しかしながら、これらの結果のほとんどは精製されて抽出された核酸またはタンパク質試料について得られたものであり、病因とされるタンパク質の組織部位についての情報は生理学的過程の理解に決定的なものであるが、この種の情報を生じるものではない。利用可能なデータのほとんどのもう一つの欠点は、目的とする分子のｍＲＮＡおよびタンパク質発現を同時に分析した研究はほんの僅かに過ぎないことであるが、これは特定のタンパク質の発現パターンを明らかにする上で重要である。従って、ｍＲＮＡとタンパク質発現マップとを組み合わせたデータを有することは、極めて有用なことである。
【０００４】
ｍＲＮＡの組織発現マッピングの最新の技術は、通常はｉｎ　ｓｉｔｕハイブリダイゼーション（ＩＳＨ）の後に放射性、蛍光または化学発光マーカーにカップリングした核酸プローブに依存している。組織切片でのタンパク質発現マッピングについては、通常の手法としては免疫組織化学および免疫蛍光が挙げられる。
【０００５】
組織切片発現マッピングのこれら総ての手法の主要な欠点は、同時に分析することができるターゲッティング分子の数が、同一実験では僅かに３または４個の異なるターゲット分子しか検討することができないので、蛍光プローブまたは抗体を用いても限定されていることである。
【０００６】
これに反して、質量分析法では、分子量によって生体分子の複雑な混合物の同時マルチプレックス分析を行うことができる。特に、マトリックス支援レーザー脱離／イオン化（ＭＡＬＤＩ）質量分析法は生物学的研究の分野において強力な手段となってきており、複雑な混合物由来の核酸、ペプチドおよびタンパク質の検出、同定および特性決定に用いられている。
【０００７】
特に、　Ｏｌｅｊｎｉｋと共同研究者らは、光開裂可能なペプチド－ＤＮＡ接合体の合成および特性決定を、マトリックス支援レーザー脱離／イオン化（ＭＡＬＤＩ）質量分析法による固定した合成ターゲットＤＮＡを検出するための光開裂可能な質量マーカー（ＰＣＭＭ）ハイブリダイゼーションプローブとしてのそれらの使用と共に報告している(Olejnik et al. Nucleic Acids Res. 1999 Dec 1;27(23):4626-31)。
【０００８】
ＷＯ　９８／２６０９５号には、生体分子ターゲットと特異的に相互作用するための質量標識化合物（ｍａｓｓ－ｌａｂｅｌｅｄ　ｃｏｍｐｏｕｎｄｓ）の合成および使用が記載されている。ＷＯ　００／６８４３４号には、ターゲット分子をシグナルでコード化し

た後コード化シグナルを解読することに基づく単一分析法で試料中の複数の検体を検出する方法が記載されている。

【０００９】

しかしながら、従来技術分野では、組織切片に含まれる核酸のＭＡＬＤＩ質量分析法（ＭＡＬＤＩ－ＭＳ）による分析の方法は報告されていない。

【００１０】

本発明者らは、ターゲット分子に特異的に結合する残基（ｍｏｉｅｔｙ）、既知分子量の残基（「タグ」残基）とイオン化過程中にＭＡＬＤＩレーザーによって直接切断される光開裂可能なリンカーとから構成される接合体を用いて、ＭＡＬＤＩ－ＭＳによって組織切片中の生体分子、特にｍＲＮＡ、を検出する新規な方法を組み立てた（図１参照）。この新規な方法により、組織切片中のｍＲＮＡの容易かつ正確な間接マッピングを行うことができ、広汎に分散した分子量のタグ残基を用いて多数の異なる生体分子を同時に分析することができる。

【００１１】

更に、この新規な方法を用いてｍＲＮＡの組織発現をマッピングするときには、２個の連続した組織切片でｍＲＮＡおよび対応するタンパク質発現の共通画像を得ることができる。実際に、幾つかの文献では、ＭＡＬＤＩ－ＭＳが組織切片におけるペプチドおよびタンパク質の直接分析の有効な手段となり得ることが示されている（Caprioli, R.M.; Farmer, T.B.; Gile, J. Anal. Chem. 1997, 69, 4751-4760; Stoeckli, M.; Farmer, T.B.; Caprioli, R.M. Nat. Med. 2001, 7, 493-496; Chaurand, P.; Schwartz,. S.A.; Caprioli, R.M. Anal Chem. 2004, 87A-93A）。

【００１２】

最後に、本発明者らは、光開裂可能なリンカーが、試料の脱離／イオン化の際の断片化によって開裂されるリンカーに置換される場合には、このような方法は質量分析法に置き換えることも可能であることを見出した。この場合には、ＭＡＬＤＩ（ＵＶ－ＭＡＬＤＩまたはＩＲ－ＭＡＬＤＩ）、ＳＩＭＳ（二次イオン質量分析法）またはＤＥＳＩ（脱離エレクトロスプレーイオン化）質量分析法、好ましくはＵＶ－ＭＡＬＤＩのみをＭＡＬＤＩの代わりに用いることができる。更に、本発明者らは、選択された光開裂可能なリンカーは試料の脱離／イオン化の際の断片化によっても開裂するので、ＵＶ－ＭＡＬＤＩ分析に適する光開裂可能なリンカーとの接合体は他の質量分析法の手法と共に用いることもできる。従って、本発明者らによって組み立てられた特定の接合体は、組織切片分析に適する任意の質量分析法による手法、特にＭＡＬＤＩ（ＵＶ－ ＭＡＬＤＩまたはＩＲ－ＭＡＬＤＩ）、ＳＩＭＳ（二次イオン質量分析法）またはＤＥＳＩ（脱離エレクトロスプレーイオン化）質量分析法と共に用いることができる。

【００１３】

従って、本発明は、組織切片における少なくとも１種類のターゲット分子マップを決定する方法であって、

ａ） 上記組織切片を少なくとも１種類の（Ａ－Ｘ）$_n$－Ｂ接合体

（上記式中、

Ａは、既知分子量のタグ分子であり、

Ｘは、試料の脱離／イオン化中に開裂されるリンカーであり、

ｎは、少なくとも１の整数であり、

Ｂは、上記ターゲット分子に特異的に結合する結合分子であり，かつ

それぞれ異なるＢ分子が異なるＡタグ分子に連結される）

とハイブリダイズさせること；

ｂ） 組織切片表面を走査し、それぞれの隣接スポットを質量分析計で分析すること（ここで、上記リンカーＸは試料のイオン化中に開裂し、それぞれのスポットで得られるデータは保存される）；および

ｃ） それぞれの異なるタグ分子の分子質量ウインドウで得られたデータを解析して、異なる検討したターゲット分子の数と同じ数の組織切片のマップを作製すること

を含んでなる、方法に関する。

【００１４】

本発明によれば、「組織切片」は、好ましくは下記の特性を有しており、すなわち、それを凍結またはパラフィン包埋することができ、その厚みは好ましくは哺乳類細胞直径、従って５－２０μｍのオーダーである。クライオスタットを用いて凍結組織から得られる凍結切片の場合には、ＯＣＴ（最適切断温度ポリマー）は好ましくは組織の固定のみに用いられるが凍結組織はＯＣＴに包埋されず、組織切片はＯＣＴと接触しないようにする。次に、組織切片を、金属、無機または有機材料、例えば金、鋼、ガラス繊維、ガラス、ナイロン６／６、シリコン、プラスチック、ポリエチレン、ポリプロピレン、ポリアミド、ポリフッ化ビニリデン、またはニッケルまたはＩＴＯのような透明性を保つ伝導性金属でコーティングした任意の厚みのガラススライスなどの更なるＭＡＬＤＩ分析に適する任意の材料から構成されるＭＡＬＤＩプレートに移すことができる。

【００１５】

本発明によれば、「ターゲット分子」とは、「結合分子」と呼ばれる別の分子に特異的に結合することができる目的の分子を意味する。このような直列の（タンデム）ターゲット／結合分子は、組織切片で特異的ハイブリダイゼーションを生じることができる場合には、任意の化学構造を示すことができる。多種多様なタンデムターゲット／結合分子並びにタンデム結合／ターゲット分子は、本発明の範囲に包含され、核酸／核酸、核酸／ペプチド、核酸／タンパク質、核酸／抗体、ペプチド／ペプチド、ペプチド／タンパク質、ペプチド／抗体、タンパク質／タンパク質（特に、リガンド／受容体）、タンパク質／糖質、抗原／抗体、ハプテン／抗体、有機化合物／受容体（図２のいずれかの例を参照）が挙げられる。

【００１６】

特に、ターゲット核酸配列は、一本鎖のターゲット核酸または二本鎖のターゲット核酸の鎖の一本に相補的である核酸配列を有する一本鎖核酸プローブを用いて特異的に検出することができる（図１参照）。ターゲットｍＲＮＡ分子の場合には、ｍＲＮＡ配列に相補的な核酸プローブを、結合分子として用いることができる。このような核酸プローブは、好ましくは２５０－５５０、更に好ましくは３００－５００、３５０－４５０、最も好ましくは約４００のヌクレオチド長を有する。

【００１７】

特定の核酸配列は、例えば転写因子、または特定のＤＮＡ配列に特異的な抗体もしくは抗体断片のようなターゲット配列に特異的に結合するタンパク質またはタンパク質断片（ペプチド）を用いて検出することができる。例えば、自己免疫抗ＤＮＡ抗体を用いることができる。

【００１８】

ペプチドおよびタンパク質の場合には、当該技術分野で知られている総てのペプチドリガンド／ペプチド受容体のタンデム分子は、本発明の範囲に包含される。これらのペプチドリガンド／ペプチド受容体のタンデム分子としては、ペプチド抗原／抗体または抗体断片、並びに任意のホルモン／ホルモン受容体、サイトカイン／サイトカイン受容体タンデム、ケモカイン／ケモカイン受容体、アプタマー／ペプチド、アプタマー／タンパク質が挙げられる。細胞移動に関与する膜糖質およびそれらのタンパク質受容体も、本発明の範囲にある。

【００１９】

また、任意の種類の抗原（例えば、核酸、ハプテン、ペプチドまたはタンパク質）およびそれらの特異抗体は、本発明によるタンデムターゲット／結合および結合／ターゲット分子に包含される。特に、接合体における結合分子として抗体または抗体断片を使用することによって、ＭＡＬＤＩ－ＭＳ分析を、抗体に光開裂可能に結合したタグ分子により明らかにする手法として用いる、組織切片における新規な種類の免疫細胞化学が可能になる。

【００２０】

有機化合物は、本発明による方法を用いてマッピングすることもできる。特に、投与した有機薬剤のイン・ビボでの分布は、本発明による方法を用いて観察することができる。

【００２１】

本発明による上記方法の特定の態様では、それぞれのターゲット分子は、核酸、特にｍＲＮＡ分子、ペプチド、タンパク質、特に受容体およびリガンド、抗体、抗原、ハプテン、および有機化合物から構成される群から独立して選択される。好ましい態様では、少なくとも１つのターゲット分子はｍＲＮＡ分子である。もう一つの好ましい態様では、少なくとも１つのターゲット分子はペプチド、タンパク質、抗原またはハプテンから選択される。

【００２２】

本発明による任意の上記方法の特定の態様では、ターゲット分子に特異的に結合する各結合Ｂ分子は、核酸、特にオリゴヌクレオチド、ペプチド、タンパク質、特に受容体およびリガンド、抗体、抗原、ハプテン、アプタマーおよび有機化合物から構成される群から独立して選択される。好ましい態様では、少なくともターゲット分子がｍＲＮＡ分子であるときには、ターゲット核酸に特異的に結合する各Ｂ分子は、上記ターゲットｍＲＮＡ配列に相補的な配列を有する核酸プローブである。もう一つの好ましい態様では、少なくともターゲット分子がペプチド、タンパク質、抗原またはハプテンであるときには、ターゲットペプチド、タンパク質、抗原またはハプテンに特異的に結合する各Ｂ分子は、そのペプチド、タンパク質、抗原またはハプテンに対して向けられた抗体である。特に、抗体または抗体断片を結合分子Ｂとして用いるときには、上記抗体またはその断片が特異的に向けられたターゲット分子は、ＩｇＭ、ＩｇＤ、ＩｇＧまたはＩｇＥなどの特定の抗体サブクラスのフレームワーク領域となることができる。あるいは、抗体または抗体断片が結合分子Ｂとして用いられるときには、抗体またはその断片が特異的に向けられるターゲット分子は、例えば、ウサギ、マウス、ラット、ヤギ、ハムスター、ヒツジおよびヒトのような特定の種に生じた抗体のフレームワーク領域となることができる。これらの２つの場合には、任意の所望のターゲット分子に対して向けられる特定のサブクラスの一次未修飾抗体であって、次に本発明による抗体－タグ接合体によって認識される一次抗体を用いて、２段階間接検出を行うことができる（実施例２および図３参照）。このような接合体は、任意の抗原特異性を有する任意の一次抗体が接合体の抗体によって認識されるサブクラスに属する場合には、この一次抗体と共に用いることができるので、極めて有利である。

【００２３】

本発明によれば、「タグ分子」は、ＭＡＬＤＩ（ＵＶ／ＩＲ）、ＳＩＭＳおよびＤＥＳ１を用いる質量分析法によって検出可能な既知分子量の分子を表す。ＭＡＬＤＩ質量分析計では、通常は広範囲のｍ／ｚ比で化合物を検出することができる。例えば、ＭＡＬＤＩ－飛行時間（ＴＯＦ）型分析装置は、１，０００，０００までのｍ／ｚ比の化合物を検出することができる。適当なタグ分子は、ターゲットおよび結合分子の特異的結合を妨げる必要がない。従って、好ましくは、適当なタグ分子は、ターゲットおよび結合分子の結合の立体障害を回避するような立体容積に制限される。従って、これら上記の理由により、適当なタグ分子は、好ましくは、ｍ／ｚ比は１０，０００を下回る。適当なタグ分子としては、ペプチド、核酸、糖類、ポリマー、脂質および有機化合物が挙げられる。これらの中、ターゲット分子と結合分子との間の複合体形成を明らかにするのに用いられる標識分子、例えば、通常のフルオレセインイソチオシアネート（ＦＩＴＣ）、Ｒ－フィコエリトリン（ＰＥ）、Ｃｙ３、Ｃｙ５、Ｃｙ７、ＰｅｒＣＰ、アロフィコシアニン（ＡＰＣ）、テキサスレッド（Ｔｅｘａｓ　Ｒｅｄ）、ＴＲＩＴＣ、ＰＥ－Ｃｙ５接合体、ＰＥ－Ｃｙ７接合体またはＡＰＣ－Ｃｙ７接合体のような蛍光色素；アルカリホスファターゼまたはペルオキシダーゼのような酵素；ビオチン；金；または総てのＭＡＬＤＩマトリックス、または純粋なレーザー脱離モードで分析することができる総ての化合物、および例えば予備イオン化分子を、タグ分子として用いることもできる。極めて小さな分子の検出に適する質量分析法の手法を用いるときには、原子をタグ分子として用いることもできる。

【００２４】

上記した総てのタグ分子は本発明で用いることができ、いかなる場合にも、用いられる質量分析法の手法によっては、質量分析法による分析の技術にある程度習熟した者であれば、どの種類のタグ分子を選択して検出段階を最適にするかが分かるであろう。例えば、ＭＡＬＤＩ質量分析法を用いるときには、タグ分子は好ましくはｍ／ｚ比が＜５０００である。あるいは、ＳＩＭＳを用いるときには、タグ分子は好ましくはｍ／ｚ比が＜５００であり、ＤＥＳＩを用いるときには、タグ分子は好ましくはｍ／ｚ比が＜５０００である。

【００２５】

本発明による任意の上記方法の特定の態様では、各Ａタグ分子は、ペプチド、核酸、糖類、および有機化合物から構成される群から選択される。好ましい態様では、少なくとも１種類のＡタグ分子はペプチドである。

【００２６】

本発明による上記方法の好ましい態様では、ターゲット分子はｍＲＮＡ分子であり、結合Ｂ分子はｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子はペプチドである。本発明による上記方法の好ましい態様では、ターゲット分子は、ペプチド、タンパク質（抗体など）、抗原またはハプテンであり、結合分子は抗体または抗体断片であり、タグ分子はペプチドである。本発明による上記方法のもう一つの好ましい態様では、ターゲット分子は、ペプチド、タンパク質（抗体など）、抗原またはハプテンであり、結合分子は抗体または抗体断片であり、タグ分子は、通常の抗体標識分子であり、例えばフルオレセインイソチオシアネート（ＦＩＴＣ）、Ｒ－フィコエリトリン（ＰＥ）、Ｃｙ３、Ｃｙ５、Ｃｙ７、ＰｅｒＣＰ、アロフィコシアニン（ＡＰＣ）、テキサスレッド、ＴＲＩＴＣ、ＰＥ－Ｃｙ５接合体、ＰＥ－Ｃｙ７接合体またはＡＰＣ－Ｃｙ７接合体のような蛍光色素；アルカリホスファターゼまたはペルオキシダーゼのような酵素；またはビオチンまたは金；または総てのＭＡＬＤＩマトリックス、または純粋なレーザー脱離モードで分析することができる総ての化合物、および例えば予備イオン化分子である。

【００２７】

本発明によれば、「上記組織切片を少なくとも１種類の（Ａ－Ｘ）。－Ｂ接合体とハイブリダイズさせる」とは、組織切片と接合体を、接合体の結合Ｂ分子が組織切片におけるそのターゲット分子に特異的に結合することができる条件で接触させる反応を意味する。結合およびターゲット分子の性質によっては、当業者は周知のハイブリダイゼーションプロトコルを利用することができる。実際に、結合分子が核酸プローブであるときには、組織切片におけるｉｎ　ｓｉｔｕハイブリダイゼーションは詳細に報告されており、従って当該技術に熟練した者には容易に利用することができる。組織切片での抗体染色も、当該技術に熟練した者には容易に利用することができるごく普通の技術である。

【００２８】

本発明による方法を用いると、組織切片における少なくとも１種類のターゲット分子のマップを得ることができる。組織切片におけるターゲット分子の「マップ」とは、この組織切片における上記ターゲット分子の発現を二次元的に表したものを意味する。この二次元的表現は、組織切片表面をＭＡＬＤＩ分析装置を用いて画定したスポット密度で走査し、それぞれの連続スポットについてＭＡＬＤＩ分析を行い、得られたデータとそれぞれのスポットの座標の両方を蓄積することによって得られる。スポット密度が高くなれば、得られるマップの精度は一層高くなる。ＭＡＬＤＩレーザーの直径は、一般的には系の焦点形成に応じて５０－２００μｍであるので、２個の隣接する照射スポットはレーザー光線の直径（すなわち、５０－２００μｍ）で離れているのが好ましい。精確なターゲット分子マップを得るには、隣接スポットは好ましくは最大で３００μｍ、最大で２００μｍ、更に好ましくは最大で１００μｍ、最大で８０μｍ、最大で６０μｍ、最大で５０μｍ、最大で４０μｍ、最も好ましくはＭＡＬＤＩレーザーの直径だけ離れている。

【００２９】

次に、それぞれのスポットのデータをタグ分子の分子ウインドウで分析し、タグ分子のシグナル強度をスポット座標に記録する。このようなイメージ再構築は、当該技術分野で

知られているまたは市販されている任意の適当なイメージ再構築ソフトウェアを用いて自動的に行うことができる。適当なソフトウェアの例は、ＲＳＩ（ＲＳＩ Ｃｏｒｐｏｒａｔｅ Ｈｅａｄｑｕａｒｔｅｒｓ，４９９０ パール・イースト・サークル，ボールダー，コロラド ８０３０１）によって商品化されているＩＤＬ（Ｉｎｔｅｒａｃｔｉｖｅ Ｄａｔａ Ｌａｎｇｕａｇｅ）ソフトウェア、ｆｌｅｘＩｍａｇｉｎｇ（Ｂｒｕｋｅｒ Ｄａｌｔｏｎｉｃｓ，ブレンメン，デラウェア）、ＭＩＴ（Ｍ．Ｓｔｏｅｃｋｌｉ，Ｎｏｖａｒｔｉｓ，バーゼル，スイス国）である。

【００３０】

組織切片における少なくとも１種類のターゲット分子マップを決定するための本発明による方法では、数個の異なるターゲット分子を同時にマッピングすることができる。実際に、幾つかの異なるターゲット分子の検出を行うには、異なるタグＡ分子を有し、よって異なる分子量を示す、幾つかの接合体を用いることで十分である。広く分散した分子量を有するタグ分子を用いることによって、本発明による任意の上記方法を用いて、同一組織切片における多数の異なるターゲット分子の発現を同時にマッピングすることができる。特定の態様では、少なくとも２、少なくとも３、少なくとも５、少なくとも８、少なくとも１０、少なくとも１５、少なくとも２０、少なくとも３０、少なくとも４０、少なくとも５０、少なくとも７５または少なくとも１００個の異なるターゲット分子を、同時にマッピングすることができる。特に、ｍＲＮＡターゲット分子の場合には、少なくとも２、少なくとも３、少なくとも５、少なくとも８、少なくとも１０、少なくとも１５、少なくとも２０、少なくとも３０、少なくとも４０、少なくとも５０、少なくとも７５または少なくとも１００個の異なるターゲットｍＲＮＡ分子を、同一組織切片で同時にマッピングすることができる。

【００３１】

ターゲット分子がｍＲＮＡ分子である任意の上記方法の特定の態様では、この方法は、更に、組織切片のそれぞれ対応するタンパク質発現マップを作製するためのそれぞれのｍＲＮＡ分子に対応するタンパク質の分子質量ウインドウで得られたデータを分析することからなる工程ｄ）を含んでなることがある。

【００３２】

第一の好ましい態様では、ＭＡＬＤＩ質量分析法を用い、Ｘリンカー分子はＭＡＬＤＩレーザーの波長で光開裂可能である。従って、本発明は、組織切片における少なくとも１個のターゲット分子マップを決定する方法であって、

ａ）　上記組織切片を少なくとも１種類の（Ａ－Ｘ）$_n$－Ｂ接合体
（上記式中、
Ａは、既知分子量のタグ分子であり、
Ｘは、ＭＡＬＤＩレーザーの波長で光開裂可能なリンカーであり、
ｎは、少なくとも１の整数であり、
Ｂは、上記ターゲット分子に特異的に結合する結合分子であり、
それぞれの異なるＢ分子は異なるＡタグ分子に連結している）
とハイブリダイズさせること；

ｂ）　組織切片表面を走査し、それぞれの隣接スポットを質量分析計で分析すること（ここで、ＭＡＬＤＩレーザーは、タグ分子Ａを放出し、かつ試料のイオン化を誘発するのに用いられ、それぞれのスポットで得られるデータは保存される）：および

ｃ）　それぞれの異なるタグ分子の分子質量ウインドウで得られたデータを解析して、異なる検討したターゲット分子の数と同じ数の組織切片のマップを作製すること
を含んでなる、方法に関する。

【００３３】

本発明によれば、「ＭＡＬＤＩレーザーの波長で光開裂可能なリンカー」とは、２つの他の化学残基（ｃｈｅｍｉｃａｌ ｍｏｉｅｔｙ）を分離し、ＭＡＬＤＩレーザーへの暴露下で少なくとも１つの部位で開裂することができる化合物を表す。大抵のＭＡＬＤＩレーザーは、通常は３００－５００ｎｍの紫外（ＵＶ：５００ｎｍより短波長）波長を有

する。例えば、多くのＵＶ－ＭＡＬＤＩ分析装置は、３３７ｎｍの波長のパルス窒素レーザーを有する。従って、本発明によるリンカーは、特定の態様では、２５０－５００ｎｍの波長、好ましくは３２０－３６０ｎｍまたは３２０－３５０ｎｍの波長、更に好ましくは３３７ｎｍの波長への暴露下で少なくとも１つの部位で効率的に開裂するので、ＵＶ－ＭＡＬＤＩレーザーはリンカーを開裂し、かつ試料をイオン化するという両方の作用をする。他のＭＡＬＤＩ分析装置は、赤外（ＩＲ：　７７０ｎｍより長波長）レーザーを示す。例えば、Ｎｄ：ＹＡＧレーザー（波長＝１０６０ｎｍ）、Ｅｒ：ＹＡＧレーザー（波長＝２９４０ｎｍ）、中赤外光パラメトリック発振器（ＯＰＯ）（波長＝２９４０ｎｍ）またはＴＥＡ－CO_2レーザー　（波長＝１０６００ｎｍ）をＩＲ－ＭＡＬＤＩレーザーとして用いることができる。他の特定の態様では、本発明によるリンカーは、従って１０００－１１００ｎｍの波長、好ましくは１０６０ｎｍの波長、または２９００－３０００ｎｍの波長、好ましくは２９４０ｎｍの波長、または１０５００－１０７００ｎｍの波長、好ましくは１０６００ｎｍの波長への暴露下で少なくとも１つの部位で効率的に開裂し、それぞれＮｄ：ＹＡＣ、Ｅｒ：ＹＡＧまたはＴＥＡ－CO_2　ＩＲ－ＭＡＬＤＩレーザーはリンカーを開裂し、かつ試料をイオン化するという両方の作用をする。

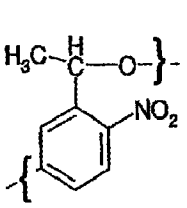
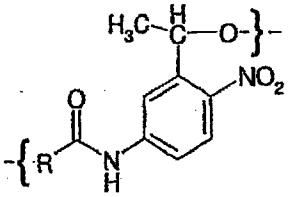
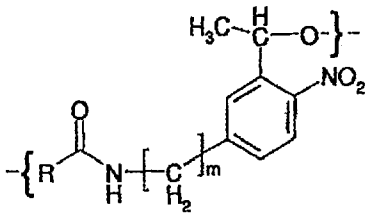
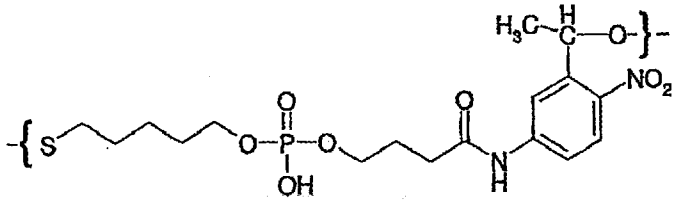
【００３４】

ＵＶ－ＭＡＬＤＩレーザーの場合には、幾つかのリンカーが報告されており（Olejnik et al. Nucleic Acids Res. 1999 Dec 1;27(23):4626-31; BaiX, et al. Nucleic Acids Res. 2004 Jan 26;32(2):535-41; Wenzel T et al. Nucleosides Nucleotides Nucleic Acids. 2003 May-Aug;22(5-8): 1579-81）、有機分子に光開裂可能なリンカーを導入するのに有用な幾つかの試薬が、特に、ＡｍｂｅｒＧｅｎ（登録商標）（１１０６　コモンウェルスアベニュー　ボストン，マサチューセッツ　０２２１５，米国）、Ｌｉｎｋ　Ｔｅｃｈｎｏｌｏｇｉｅｓ（３　マラードウエイ，ストラスサイクル・ビジネス・パーク，ベルシル，ラナークシアＭＬＡ　３ＢＦ，スコットランド）、Ｉｎｔｅｇｒａｔｅｄ　ＤＮＡ　Ｔｅｃｈｎｏｌｏｇｉｅｓ　（１７１０　コマーシャル・パーク゜コーラルビル，アイオワ　５２２４１，米国）、Ｇｌｅｎ　Ｒｅｓｅａｒｃｈ（２２８２５　デービス・ドライブ，スターリング，バージニア　２０１６４，米国）、Ｅｕｒｏｇｅｎｔｅｃ（ＥＵＲＯＧＥＮＴＥＣ　ｓ．ａ．　Ｈｅａｄｑｕａｒｔｅｒｓ．　リエージュ・サイエンス・パーク．　リュ・ボア・サン・ジアン５．　４１０２　セライング．　ベルギー国）などから市販されている。

【００３５】

ＭＡＬＤＩ質量分析法とＭＡＬＤＩレーザーの波長で光開裂可能なリンカー分子Ｘとを用いる本発明による任意の上記方法の特定の態様では、Ｘリンカー分子は、

【化１】

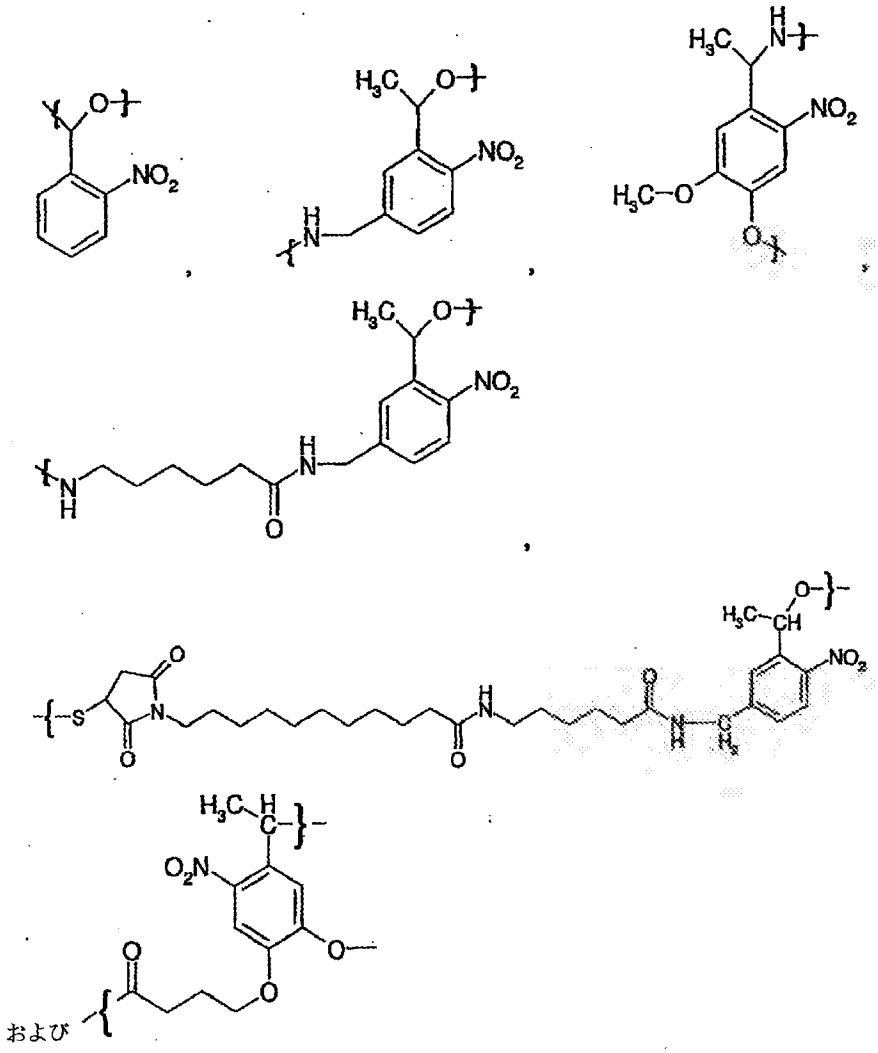



および


（上記式中、ＲはＣ１－Ｃ６アルキル基であり、ｍは１－４の整数である）
からなる群から選択される残基である。

【００３６】

あるいは、ＭＡＬＤＩレーザーの波長で光開裂可能な上記リンカー分子Ｘを用いる態様では、Ｘリンカー分子は、

【化２】



からなる群から選択される残基である。

【００３７】

タグ分子Ａと結合分子Ｂを連結するリンカー分子Ｘは、このような残基を含んでなる市販のリンカーを用いてＡとＢの間で得ることができる。例えば、本発明による方法で用いるための光開裂可能なリンカーとの接合体を調製するのに適したリンカーは、

ＰＣリンカー（製品番号１０－４９２０－０２）、

ＰＣスペーサーホスホルアミダイト（製品番号１０－４９１３－０２）、および

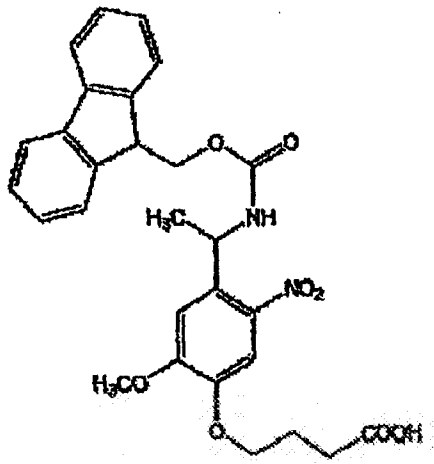
ＰＣアミノ修飾因子（製品番号１０－４９０６－０２）

などをＧｌｅｎ Ｒｅｓｅａｒｃｈ Ｃｏｒｐｏｒａｔｉｏｎ（２２８２５ デービス・ドライブ，スターリング，バージニア ２０１６４，米国）から、または

例えば、リンカー Ｓｕｌｆｏ－ＫＭＵＳ（製品番号＃２１１１１）をＰｉｅｒｃｅ（Ｐｉｅｒｃｅ Ｂｉｏｔｅｃｈｎｏｌｏｇｙ，Ｉｎｃ．Ｃｕｓｔｏｍｅｒ Ｓｅｒｖｉｃｅ Ｄｅｐａｒｔｍｅｎｔ Ｐ．Ｏ．Ｂｏｘ １１７ ロックフォード，イリノイ，６１１０５ 米国）から、または

商標Ｎｏｖａｂｉｏｃｈｅｍ（登録商標）、例えば、リンカー

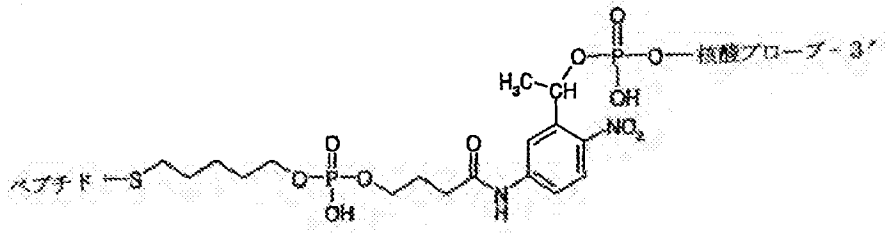
【化３】



（製品番号０１－６０－００４２）をＭｅｒｃｋ　Ｂｉｏｓｃｉｅｎｃｅｓ　Ｌｔｄ（プールバード・インダストリアル・パーク，　パッジ・ロード，　ビーストン，　ノッティンガム　ＮＧ９　２ＪＲ，　英国）から、または
　Ｃｈｒｏｍａｌｉｎｋ　Ｂｉｏｔｉｎ　３５４Ｓという名称の製品番号Ｂ１００１であって、３５４ｎｍの波長で開裂可能であるものをＳｏｌｕｌｉｎｋから得ることができる。

【００３８】
　ターゲット分子がｍＲＮＡ分子であり、結合Ｂ分子がｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子がペプチドである、もう一つの好ましい態様では、ｎが１であり、（Ａ－Ｘ）$_n$－Ｂ接合体、および得られる（Ａ－Ｘ）－Ｂ接合体が下記の構造

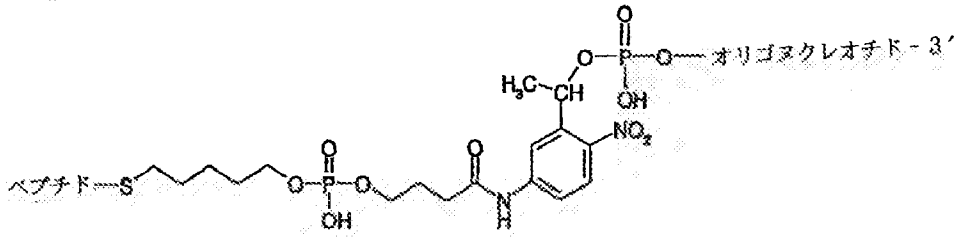
【化４】



を有する。

【００３９】
　このような接合体は、例えば、下記の構造

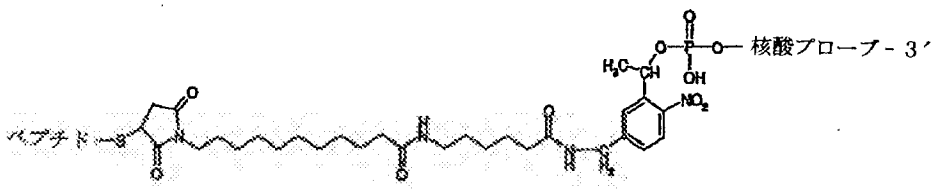
【化５】



のフォワード（５’）プライマーを用いて所望の核酸配列を増幅することによって得ることができる。

【００４０】
　ターゲット分子がｍＲＮＡ分子であり、結合Ｂ分子がｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子がペプチドである、もう一つの好ましい態様では、ｎは１であり、（Ａ－Ｘ）$_n$－Ｂ接合体、および得られる（Ａ－Ｘ）－Ｂ接合体は下記の構造

【化６】

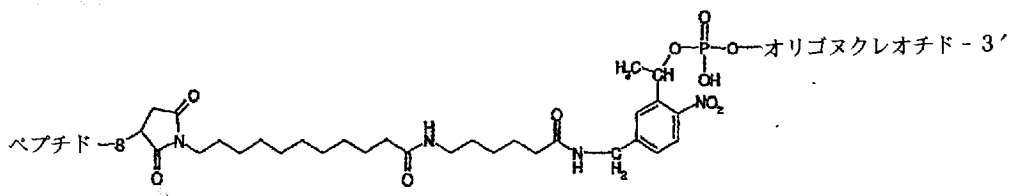


を有する。

【００４１】

このような接合体は、例えば、下記の構造

【化７】

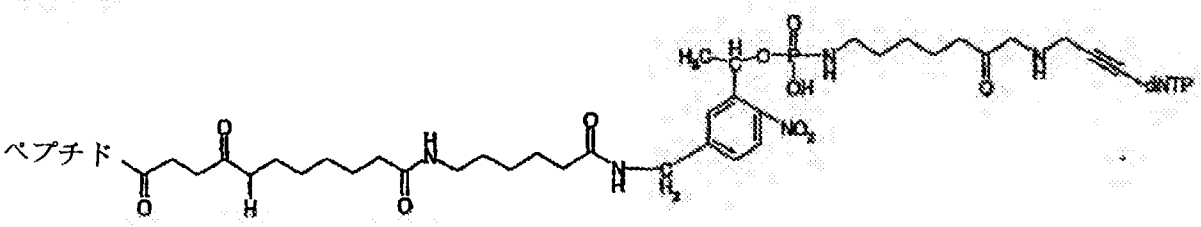


を有するフォワード（５’）プライマーを用いて所望の核酸配列を増幅することによって得ることができる。

【００４２】

あるいは、ターゲット分子がｍＲＮＡ分子であり、結合Ｂ分子がｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子がペプチドである、もう一つの好ましい態様では、ｎは１より大きく、核酸プローブは、下記の構造：

【化８】



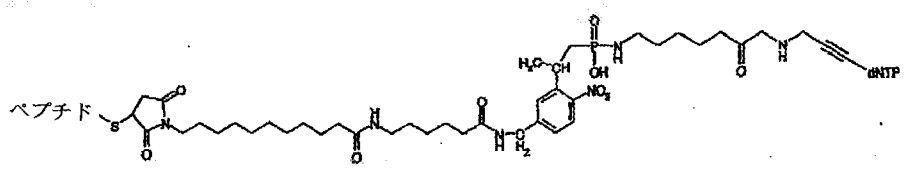
（上記式中、「ｄＮＴＰ」はｄＡＴＰ、ｄＣＴＰ、ｄＧＴＰ、ｄＴＴＰおよびｄＵＴＰなどの任意のデオキシヌクレオチド三リン酸を意味する）

の少なくとも１種類の修飾塩基を含んでなる。好ましい態様では、修飾したｄＵＴＰ塩基が用いられる。

【００４３】

ターゲット分子がｍＲＮＡ分子であり、結合Ｂ分子がｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子がペプチドである、もう一つの好ましい態様では、ｎは１より大きく、核酸プローブは、下記の構造：

【化９】



（上記式中、「ｄＮＴＰ」はｄＡＴＰ、ｄＣＴＰ、ｄＧＴＰ、ｄＴＴＰおよびｄＵＴＰなどの任意のデオキシヌクレオチド三リン酸を意味する）

の少なくとも１種類の修飾塩基を含んでなる。好ましい態様では、修飾したｄＵＴＰ塩基が用いられる。

【００４４】

このような修飾ペプチドタグ付きｄＵＴＰを用いると、ペプチドタグ付き核酸ハイブリダイゼーションプローブは、ｄＡＴＰ、ｄＣＴＰ、ｄＧＴＰおよびペプチドタグ付きｄＵＴＰの存在下にて簡単なＰＣＲ増幅を用いて極めて簡単に生成させることができる。この方法により、特異的ハイブリダイゼーションプローブを任意のターゲットｍＲＮＡ配列について容易に合成することができる。

【００４５】

更に、修飾ペプチドタグ付きｄＵＴＰを用いて合成した所定のハイブリダイゼーションプローブはその配列にＵ塩基の数と同数のタグペプチドを有するので、これによりシグナル増幅を行うことができる。

【００４６】

最後に、修飾ペプチドタグ付きｄＵＴＰを用いて合成したハイブリダイゼーションプローブを用いると、組織切片におけるｍＲＮＡ発現を定量的に分析することができる。実際に、１または数個の検討したｍＲＮＡの対応するタグによって生じるシグナルを、参照ｍＲＮＡ配列（例えば、ＨＰＲＴのアクチンのようなハウスキーピング遺伝子）について得られたものと比較することができる。それぞれのハイブリダイゼーションプローブにおけるＵ塩基の数は知られているので、それぞれの検討したｍＲＮＡ配列と参照ｍＲＮＡ配列との間の発現比を計算することができる。

【００４７】

ターゲット分子がｍＲＮＡ分子であり、結合Ｂ分子がｍＲＮＡ配列に相補的な配列を有する核酸プローブであり、Ａタグ分子がペプチドである方法では、上記接合体を用いて、幾つかの異なるターゲットｍＲＮＡ分子を同時にマッピングすることができる。実際に、幾つかの異なるターゲット分子の検出を行うには、異なる核酸プローブと異なる分子量を示す異なるタグペプチドを有する幾つかの接合体を用いることで十分である。広く分散した分子量を有するタグペプチドを用いることによって、任意の上記接合体を用いて同一組織切片における多数の異なるターゲットｍＲＮＡ分子の発現を同時にマッピングすることができる。特定の態様では、少なくとも２、少なくとも３、少なくとも５、少なくとも８、少なくとも１０、少なくとも１５、少なくとも２０、少なくとも３０、少なくとも４０、少なくとも５０、少なくとも７５、または少なくとも１００個の異なるターゲットｍＲＮＡ分子を，同時にマッピングすることができる。

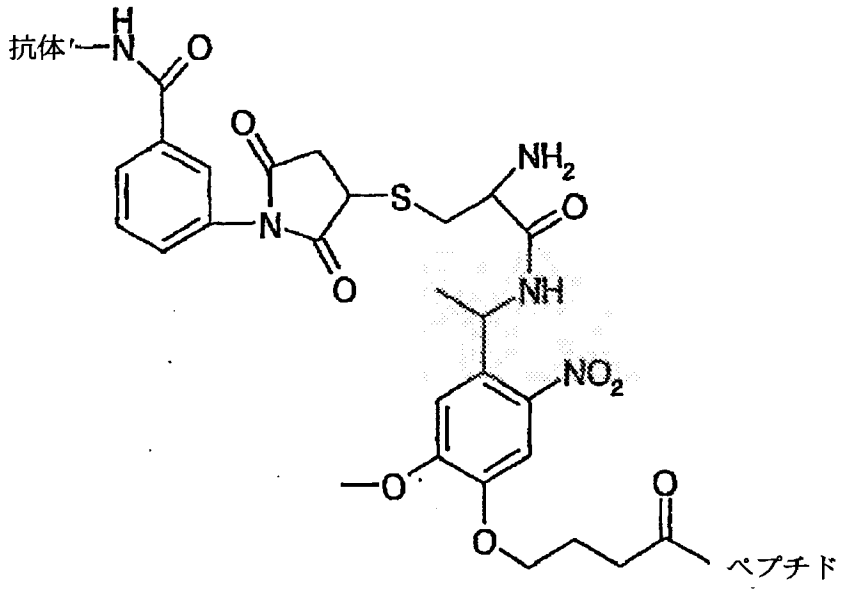
【００４８】

更に、保管されたスペクトルデータは、組織切片の直接ペプチド／タンパク質分析も示し、上記接合体を用いるターゲットｍＲＮＡ分子の組織切片分析法は、更に、組織切片のそれぞれに対応するタンパク質発現マップを作製するためのそれぞれのｍＲＮＡ分子に対応するタンパク質の分子質量ウインドウで得られたデータを分析することからなる最終工程ｄ）を含んでなることがある。

【００４９】

ターゲット分子がペプチド、タンパク質（抗体など）、またはハプテンである好ましい態様では、結合Ｂ分子は上記ターゲット分子に対して向けられた抗体であり、Ａタグ分子はペプチドであり、ｎは１であり、（Ａ－Ｘ）－Ｂ接合体は下記の構造：

【化１０】



を有する。

【００５０】

ＭＡＬＤＩ質量分析法を用いる本発明による任意の方法では、ＭＡＬＤＩ－ＭＳ分析に用いるマトリックスは、任意の古典的ＭＡＬＤＩマトリックス（例えば、ＣＨＣＡ、２，５－　ＤＨＢ、ＳＡ、ＨＡＢＡ、３－ＨＰＡなど）または２，４－ＤＮＰＨでよい。

【００５１】

「マトリックス」とは、検体と混合するとき、レーザー照射によって良好に脱離され固相結晶から気相または蒸気相にイオン化され、分子イオンとして加速される、結晶性のマトリックス包埋検体分子を生じる任意の材料を意味する。一般に用いられるＭＡＬＤＩ－ＭＳマトリックスは、通常はレーザー波長で吸収する小さな酸性化合物であり、ニコチン酸、桂皮酸、２，５－ジヒドロキシ安息香酸（２，５－ＤＨＢ）、α－シアノ－４－ヒドロキシ桂皮酸（ＣＨＣＡ）、３，５－ジメトキシ－４－ヒドロキシ桂皮酸（シナピン酸またはＳＡ）、３－メトキシ－４－ヒドロキシ桂皮酸（フェルラ酸）、３，４－ジヒドロキシ桂皮酸（カフェイン酸）、２－（４－ヒドロキシフェニルアゾ）安息香酸（ＨＡＢＡ）、３－ヒドロキシピコリン酸（ＨＰＡ）、２，４，６－トリヒドロキシアセトフェノン（ＴＨＡＰ）および２－アミノ－４－メチル－５－ニトロピリジンが挙げられる。これらのマトリックスを調製するためのプロトコルは当該技術分野で周知であり、これらのマトリックスの大半は市販されている。ペプチド／タンパク質分析に現在一般に用いられるマトリックスとしては、α－シアノ－４－ヒドロキシ桂皮酸（ＣＨＣＡ）、２，５－ジヒドロキシ安息香酸（２，５－ＤＨＢ）およびシナピン酸（ＳＡ）が挙げられる。ＤＮＰＨは２，４－ジニトロフェニルヒドラジンであり、アルデヒドおよびケトンの検出に用いられる。

【００５２】

本発明の第二の好ましい態様では、試料の脱離／イオン化の際に断片化によって開裂するリンカー分子Ｘを用い、その分析はＭＡＬＤＩ質量分析法に限定されないが、一般的には組織切片分析に適する任意の質量分析法の技術に換えることができる。特に、ＵＶ－ＭＡＬＤＩ、ＩＲ－ＭＡＬＤＩ、ＳＩＭＳまたはＤＥＳＩ質量分析法を用いることができる。

【００５３】

実際に、化合物の間接的検出には、装置の線源における脱離およびイオン生成のごく初期にタグ分子を放出する必要がある。マトリックス分子によるエネルギー吸収を介したイオン脱離／イオン化の目的でレーザーを用いるＭＡＬＤＩ質量分析法を用いるときには、

タグ分子の放出を誘発する幾つかの方法を考えることができる。第一の最も簡潔な方法は、ＭＡＬＤＩレーザー照射によるタグ分子の光解離を用いることである（図１および図３参照）。これは極めて特異的な放出を意味するが、必然的に一層複雑な構造および合成も意味する。

【００５４】

しかしながら、タグ分子を放出する方法として、迅速断片化を考えることもできる。この現象は一般に質量分析法の技術に存在しており、従ってその使用はＭＡＬＤＩおよび他の質量分析法の線源と両立する。特に、この断片化現象を、線源が一次イオンビームであるＳＩＭＳ（二次イオン質量分析法）またはＤＥＳＩ（脱離エレクトロスプレーイオン化）と称される最近に報告された技術のような他の質量分析法の技術と共に利用することができる。更に、ＭＡＬＤＩ／ＭＳについては、これら２種類の技術を、組織切片の直接分析に用いることができる（Touboul, D.; Kolmer, F.; Niehis, E.; Brunelle, A.; Laprevote, O., Journal of the American Society for Mass Spectrometry, (JASMS) 2005, 15, 1608-1618)（Cooks, R.G.M.; Ouyang, Z. Takats, Z.; Wiseman, J.M.; Science, 17 March 2006, Vol 311, 1566-1570）。

【００５５】

従って、リンカー分子を介してタグ分子に連結した特異的結合分子の接合体を用いる本発明のコンセプトは，迅速断片化によって開裂されるリンカーを用いることにより、タグ分子の迅速断片化放出を生じるという条件で、ＳＩＭＳまたはＤＥＳＩのようなＭＡＬＤＩ以外の質量分析法の技術に取り換えることができる。この現象を考えると、迅速断片化によって開裂さえされれば、光開裂可能である必要はないので、更に単純なリンカーなしの接合体をデザインすることができ、これは容易に得ることができる。実際に、あらゆる種類のリンカーを断片化によって開裂することができ、任意のリンカーの能力を容易に試験することができる。

【００５６】

更に、ＭＡＬＤＩレーザーの波長で光開裂可能であると上記したリンカーは、迅速断片化によっても開裂されることが本発明者によって見出された。あるいは、このようなリンカーは、ＭＡＬＤＩ以外の質量分析法の技術を用いるときなどの上記の本発明による任意の方法、特にＳＩＭＳまたはＤＥＳＩ質量分析法、についての接合体で用いることができる。

【００５７】

しかしながら、迅速断片化現象を利用することによって用いることができる（Ａ－Ｘ）„－Ｂ接合体は、光解離および断片化によって開裂することができる上記の特定の適当なリンカーに限定されないことを明確に理解しなければならない。実際に、Ａタグ分子とＢ結合分子を容易に連結することができる任意のリンカーの迅速断片化によって開裂する能力は、単に合成した接合体を溶液中で分析し、開裂したＡタグ分子の存在を観察することによって容易に試験される。この方法により、任意の可能性のあるリンカーは、本発明での使用について容易に試験し、承認しまたは拒絶することができる。更に、質量分析法の技術に熟練した者であれば、可能性のあるリンカーの大多数について、それらの化学構造から断片化により開裂されるかされないか分かるので、容易な上記試験を行うことはほとんどの場合に一層必要なくなるであろう。

【００５８】

特に、迅速断片化を用いてリンカーを開裂するときには、通常の販売さえされている標識抗体を接合体として用いることができる。例えば、蛍光接合抗体を用いることができ、蛍光分子は、タグ分子として用いられる。特に、ＦＩＴＣ接合、ＰＥ接合、ＰｅｒＣＰ接合、ＡＰＣ－接合、Ｃｙ３接合、Ｃｙ５接合、Ｃｙ７接合、テキサスレッド接合、ＴＲＩＴＣ接合、ＰＥ－Ｃｙ５接合、ＰＥ－Ｃｙ７接合、またはＡＰＣ－Ｃｙ７接合抗体を用いることができる。あるいは、酵素アルカリホスファターゼまたはペルオキシダーゼのような他の分子に接合した抗体を用いることもでき、酸素はタグ分子として用いられる。この可能性は、極めて広い範囲の抗原特異性を有するあらゆる種類の抗体が標識抗体として市

販されているという決定的な利点を有する。更に、標識分子を通常のリンカーを介して任意の抗体に結合させる技術は、周知の日常的技術である。この方法により、組織切片に存在する生体分子の間接質量分析法を、本発明による方法を用いて非常に多数の異なる生体分子について行うことができる。

【００５９】

従って、好ましい態様では、迅速断片化を用いるときには、ターゲット分子はペプチド、タンパク質（抗体など）、抗原、またはハプテンであり、結合分子は抗体または抗体断片であり、ｎは１であり、タグ分子は蛍光色素、酵素、ビオチンまたは金である。特にＳＩＭＳ質量分析法を用いるときには、Ｂｒのような原子を抗体に接合させることもできる（実施例４参照）。この場合には、Ｓｉｇｍａから製品番号リンカーＥ１７６９で市販されているＥＤＡＣと称するリンカーを用いて、抗体をＢｒ原子に接合させることができる（実施例４参照）。

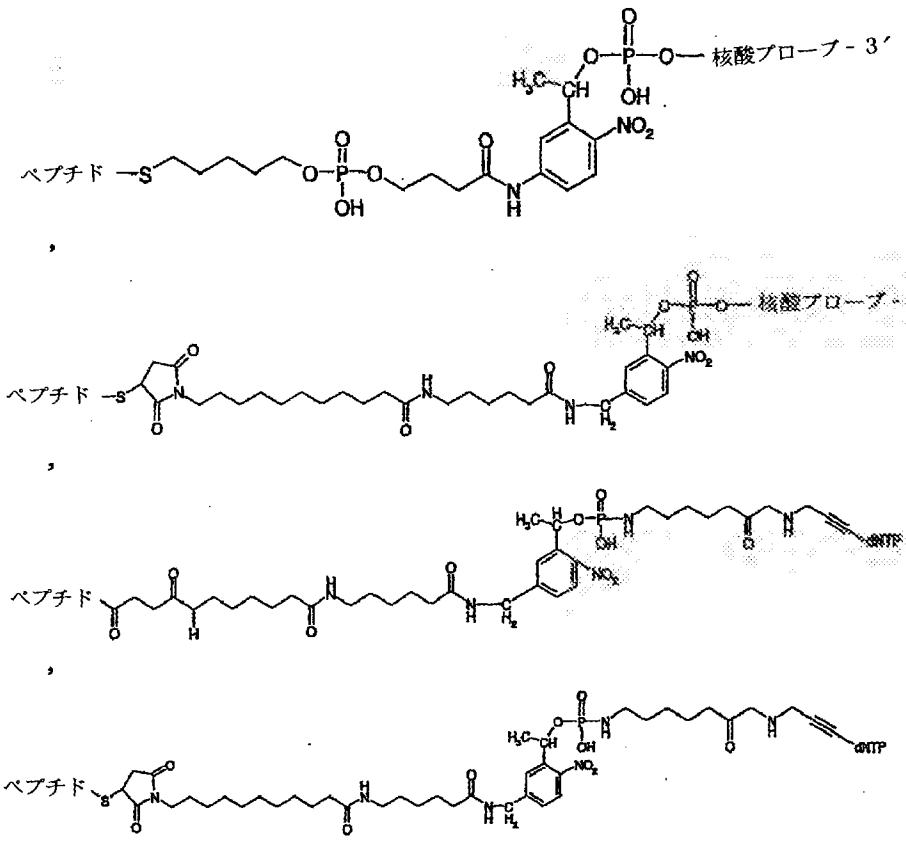
【００６０】

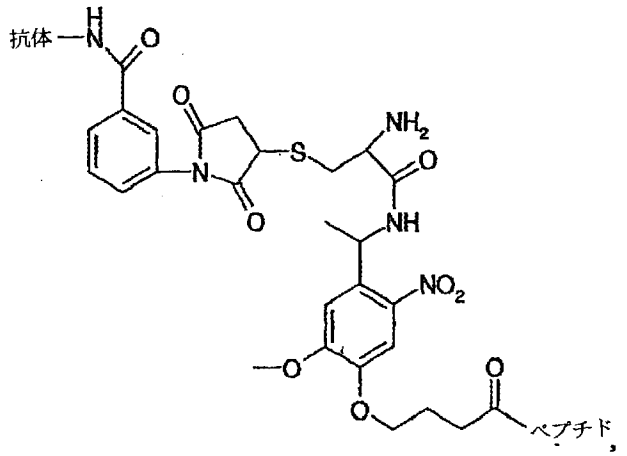
更に、迅速断片化によって開裂されるこのようなリンカー分子Ｘを用いるときには、潜在的に任意の質量分析法の技術を、組織切片分析に適しているという条件で（これはいかなる熟練者にとっても直ちに明らかである）用いて、本発明による方法を実行することができる。特に、ＵＶ－ＭＡＬＤＩまたはＩＲ－ＭＡＬＤＩ以外の幾つかの追加の質量分析法の技術、例えば、ＳＩＭＳまたはＤＥＳＩ質量分析法を用いることができる。従って、迅速断片化によって開裂されるリンカー分子Ｘを用いる方法の好ましい態様では、ＵＶ－ＭＡＬＤＩ、ＩＲ－ＭＡＬＤＩ、ＳＩＭＳまたはＤＥＳＩ質量分析法が用いられる。

【００６１】

本発明は、更に本発明による任意の方法での使用に適する接合体に関し、上記接合体は

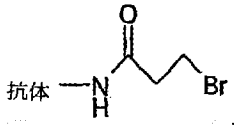
【化１１】





,

または



から選択される。

【００６２】

上記接合体のそれぞれは、個別的に本発明による接合体として好ましい。

【００６３】

本発明を概略的に説明してきたが、本発明の特徴と利点は、例示のためにのみ本明細書に提供されかつ特に断らない限り制限するものとは解釈されないある種の具体例および図を参照することによって一層の理解を深めることができる。

【００６４】

実施例１

ＭＡＬＤＩ－ＭＳによる組織切片におけるｍＲＮＡの間接検出のための光開裂可能なリンカーを有するオリゴペプチド接合体の使用

光開裂可能なリンカーを有するオリゴペプチド接合体について、ＭＡＬＤＩ－ＭＳによる組織切片における特異的ｍＲＮＡを間接検出することができる能力を試験した。

【００６５】

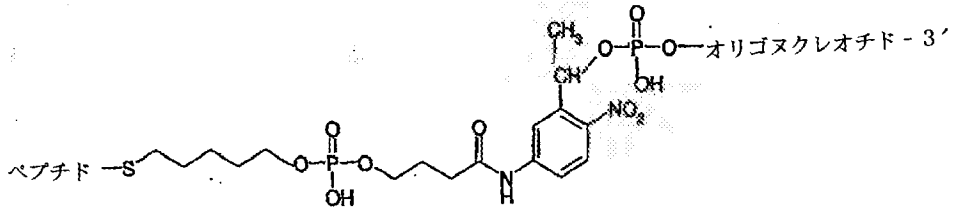
１．１　材料および方法

１．１．１　オリゴペプチド接合体

構造

検討を行ったオリゴペプチド接合体は、下記の構造を示した。

【化１２】



【００６６】

異なるペプチド残基を有する５種類のオリゴペプチド接合体を合成した。それぞれの接合体のオリゴヌクレオチドのペプチドアミノ酸配列、単一同位体（M_{mono}）および平均（M_{avg}）分子量、および核酸配列を、表２に示す。

【００６７】

【表１】

表２．検討したオリゴペプチド接合体の特徴

オリゴ'ペプチド	ペプチド配列（NからC末端）	M_{mono} (u.m.a)	M_{avg} (u.m.a)	オリゴヌクレオチド配列 (5'-3')
1	GRALGVFVC（配列番号：１）	918,50	919,14	CACGTACAGGATGTACAG（配列番号：６）
2	RPKPQQFFGLMC（配列番号：２）	1448,73	1449,79	TCGAGAGGTACATCGTG（配列番号：７）
3	RPPGFSPFRC（配列番号：３）	1160,58	1161,37	AAGCGGTACGAGTAGCA（配列番号：８）
フォワード	RPPGFSPFRC（配列番号：４）	1718,00	1719,00	GACGTACCAGGCGGTAGCTGCATTT（配列番号：９）
リバース	RPPGFSPFRC（配列番号：５）	1718,00	1719,00	CAGGACTCCCCAAAGGAGAACAGGA（配列番号：１０）

【００６８】

合成

オリゴペプチド接合体は、下記のプロトコルを用いて合成した。

【００６９】

ペプチドはＳｙｍｐｈｏｎｙ（Ｐｒｏｔｅｉｎ Ｔｅｃｈｎｏｌｏｇｉｅｓ Ｉｎｃ）上で合成し、Ｄｅｌｔａ－Ｐａｋ Ｃ１８ １５μｍ １００Ａカラム（Ｗａｔｅｒｓ）上で精製する。

【００７０】

オリゴヌクレオチドは、Ｅｘｐｅｄｉｔｅ（ＡＢＩ）上で３'から５'へ合成する。光開裂可能なリンカーを有するアミノ官能基を５'に付加した後、開裂と脱保護を行う。これらの工程は、２８％ＮＨ$_4$ＯＨ溶液を用いて暗所で２４時間行う。次に、アミノオリゴヌクレオチドをＤｅｌｔａ－Ｐａｋ Ｃ１８ １５μｍ ３００Ａカラム（Ｗａｔｅｒｓ）上で精製する。オリゴヌクレオチドのアミノ官能基を、マレイミド官能基を含んでなるヘテロ二官能性試薬にカップリングさせる。マレイミドオリゴヌクレオチドを水に可溶化し、溶液中のペプチド１．２当量に加える。混合物を１６時間撹拌する。

【００７１】

次に、オリゴペプチド接合体をＤｅｌｔａ－Ｐａｋ Ｃ１８ １５μｍ ３００Ａカラム（Ｗａｔｅｒｓ）上で精製し、質量分析法によって特性決定する。

【００７２】

１．１．２ ＭＡＬＤＩ－ＭＳ前の試料調製

幾つかの一般に用いられるマトリックスを、区別せずＭＡＬＤＩ－ＭＳ分析に用いた：α－シアノ－４－ヒドロキシ桂皮酸 （ＣＨＣＡ）、３－ヒドロキシピコリン酸 （ＨＰＡ）およびシナピン酸 （ＳＡ）。

【００７３】

ＣＨＣＡについては、マトリックス１０ｍｇをアセトニトリル／水（２：１，ｖ／ｖ，０．１％ＴＦＡ／Ｈ$_2$Ｏ）１ｍｌに溶解した。ＳＡについては、マトリックス２０ｍｇを同じ溶媒に溶解した。

【００７４】

ＳＡについては、マトリックス２０ｍｇを同じ溶媒に溶解した。

【００７５】

ＨＰＡについては、マトリックス５０ｍｇ／ｍｌを水に溶解した。

【００７６】

場合によっては、添加剤（クエン酸または酢酸アンモニウム）をマトリックスに加えた。

【００７７】

ＭＡＬＤＩ－ＭＳ分析の前に、試料溶液１μｌとマトリックス溶液１μｌを乾燥小滴製剤の手順に従ってＭＡＬＤＩプレート上で混合した（Karas, M.; Hillenkamp, F.; Anal. Chem. 1998, 60, 2299-2301）。

【００７８】

３種類の異なるＭＡＬＤＩプレート材料であるステンレススチール、金およびテフロンを試験した。

【００７９】

１．１．３　ＭＡＬＤＩ－ＭＳ分析

ＭＡＬＤＩ－ＴＯＦ質量スペクトルは、遅延引き出し（ＤＥ）を有し、かつ３３７ｎｍのパルスド窒素レーザーで操作するＶｏｙａｇｅｒ－ＤＥ　ＳＴＲ質量分析計（Ａｐｐｌｉｅｄ　Ｂｉｏｓｙｓｔｅｍｓ，　フラミンガム，　マサチューセッツ，　米国）で行った。

【００８０】

リニアモードでの分析

イノックス（ｉｎｏｘ）ＭＡＬＤＩプレートについては、取得パラメーターを下記のように設定した：　加速電圧：２０ｋＶ；第１のグリッド電圧：９４％；ガイドワイヤー電圧：０．０５％；引出遅延時間：１００－２５０ナノ秒。

【００８１】

金ＭＡＬＤＩプレートについては、取得パラメーターを下記のように設定した：　加速電圧：２５ｋＶ；第１のグリッド電圧：９６％；ガイドワイヤー電圧：０．０５％；引出遅延時間：６００ナノ秒。

【００８２】

リフレクターモードでの分析

加速電圧：２０ｋＶ，第１のグリッド電圧：７０％，ガイドワイヤー電圧：０．０５％，引出遅延時間：２００ナノ秒。

【００８３】

１．１．４　イメージ再構築のためのソフトウェア

イメージ再構築（ｉｍａｇｅ　ｒｅｃｏｎｓｔｒｕｃｔｉｏｎ）には、ソフトウェアｆｌｅｘＩｍａｇｉｎｇ　（Ｂｒｕｋｅｒ　ｄａｌｔｏｎｉｃｓ，ブレーメン，ドイツ国）を用いた。

【００８４】

１．１．５　オリゴペプチド接合体を用いるＰＣＲ増幅

プロエンケファリンのＰＣＲ増幅は、タグ付き（ペプチド－光開裂可能なリンカー－オリゴヌクレオチド）または通常の（光開裂可能なリンカーおよびペプチド残基のない）フォワード（５’－ＣＡＧ－ＧＡＣ－ＴＣＣ－ＣＣＡ－ＡＡＧ－ＧＡＧ－ＡＡＣ－ＡＧＧ－Ａ－３’，配列番号：９，表２参照）およびリバーズ（５’－ＧＡ－ＣＧＴ－ＡＣＣ－ＡＧＧ－ＣＧＧ－ＴＡＧ－ＣＴＧ－ＣＡＴ－ＴＴ－３’，配列番号：１０，表２参照）プライマーを用いて得た。

【００８５】

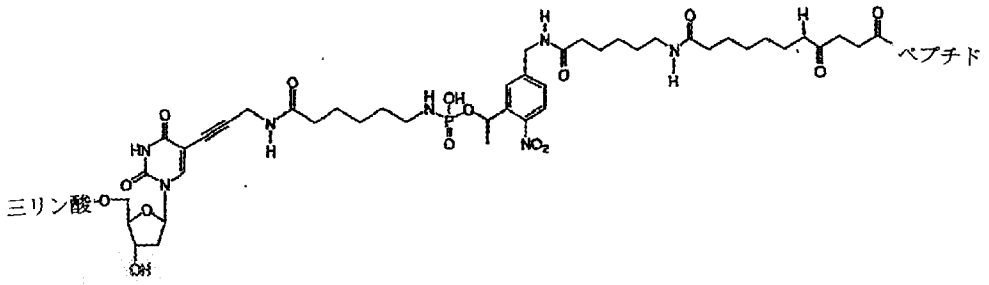
次に、増幅生成物を２％アガロースゲル中の電気泳動によって分析した。

【００８６】

１．１．６　光開裂可能なリンカーを用いるｄＵＴＰ－ペプチド接合体の合成

下式を有する修飾ペプチドタグ付きｄＵＴＰを合成した：

【化１３】

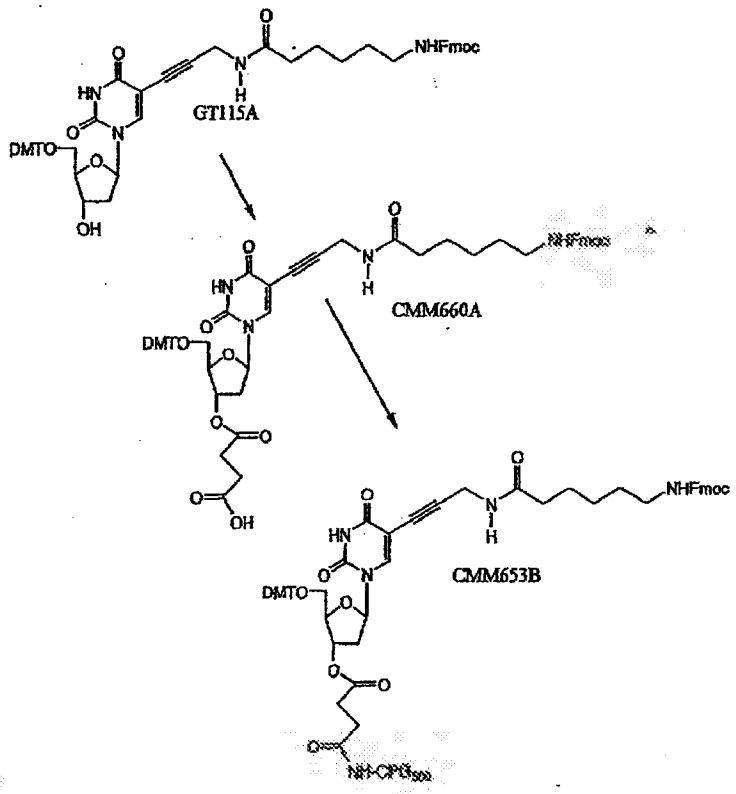


【００８７】

合成には、下記の合成工程図に従った。

【化１４】

スキーム１．Ｆｍｏｃ－アミノプロパルギルカプロイルｄＵから三リン酸塩



【００８８】

この三リン酸塩を調製するには、Ｆｍｏｃで保護したＣＰＧ樹脂が必要であった。前駆体は入手することができなかったので、ＧＴ１１５Ａ（１００ｍｇ）からスクシニレート（ｓｕｃｃｉｎｙｌａｔｅ）を調製した。試料は比較的純粋であったが、高流動性の非トリチル化化合物（Ｓｏｎｏｇａｓｈｉｒａ反応に由来し、以後の反応を妨げず、試料のｎｍｒスペクトルでは確認されない）を少量（ＴＬＣによる）含んでいた。このスクシネートを精製することはできなかったので、反応を若干変更した。定量的収率を得るには、通常は無水コハク酸２当量を反応に加えるが、これが完全に除去されないときには、ｃｐｇ樹脂のアミノ残基を官能化の際にブロックすることができる。従って、生成物の精確な純度が決定されないので、１．５当量を用いた。反応は完結しなかった（ＴＬＣからは、これはＵＶ（２５４ｎｍ）によるＴＬＣ上の成分の強度と発煙ＨＣｌで処理したときのＤＭＴカチオンの強度を比較することによって５０％を上回った）。スクシニル化されていない生成物は反応しないので、樹脂をこの混合物を用いて官能化した。樹脂は調製したが、

装填量（ｌｏａｄｉｎｇ）は極めて低く、５．４μｍｏｌ／ｇ（１８０ｍｇ）であった。

【００８９】

樹脂をＤＣＭで洗浄した２％　ＴＣＡ／ＤＣＭを用いて脱トリチル化し、ＤＭＴカチオンによる橙色が見られなくなるまでこの工程を繰り返した。

【００９０】

次いで、これを（アルゴン下にて吸引して）乾燥し、樹脂をピリジン／ＤＭＦ　１：３（０．４ｍｌ）に５分間浸漬した後、０．１Ｍ　Ｅｃｋｓｔｅｉｎ試薬をジオキサンに溶解したものを加えた（０．１ｍｌ）。反応を１５分間放置した後、樹脂を洗浄し（ジオキサン，ＭｅＣＮ）、乾燥した（アルゴン下にて吸引）。

【００９１】

次に、樹脂を０．５Ｍビス－（トリブチルアンモニウム）ピロホスフェートを無水ＤＭＦとトリ－ｎ－ブチルアミンに溶解したものに２０分間浸漬し、樹脂を洗浄し（ＤＭＦ，ＭｅＣＮ）、乾燥した（アルゴン下にて吸引）。

【００９２】

生成物を酸化し（ヨウ素／水／ピリジン／ＴＨＦ、３０分間）、洗浄し（ＭｅＣＮ）、乾燥した（アルゴン下にて吸引）。

【００９３】

Ｆｍｏｃ保護基を除去し（２０％ピペリジン／ＤＭＦ，０．５ｍｌ，２０分間）、樹脂を十分に洗浄し（ＤＭＦ，ＭｅＣＮ）、乾燥した（アルゴン下にて吸引）。これを次にＤＣＩで洗浄し、ＤＣＩ／光解離性アミノリンカーＣＥＰ　（１：１，０．５ｍｌ）の溶液を加え、反応を２０分間放置した。溶液を除去し、樹脂を洗浄し（ＭｅＣＮ）、乾燥した（アルゴン下にて吸引）。キャップＡ／キャップＢの混合物（１：１，０．５ｍｌ）を加え、樹脂を５分間浸漬した後、キャッピング試薬を除去し、上記の通りに樹脂を洗浄し、乾燥した。生成物を酸化し（Ｉ$_2$／ＴＨＦ／ピリジン／Ｈ$_2$Ｏ，５分間）、樹脂を上記の通りに樹脂を洗浄し、乾燥した。これを樹脂からｃＮＨ$_4$ＯＨを用いて室温にて３０分間開裂させた後、Ｄｉｏｎｅｘ　ＮｕｃｌｅｏＰａｃ１００　ＨＰＬＣカラム上でのアニオン交換ＨＰＬＣによって下記の溶媒系である緩衝液Ａ：１０％アセトニトリルを含む０．１Ｍ　ＮＨ$_4$Ｃｌおよび緩衝液Ｂ：　１０％アセトニトリルを含む１Ｍ　ＮＨ$_4$Ｃｌを用いて、流速２．５ｍｌ／分にて６Ｔｒｉｐｈｏｓ．ｍｔｈを用いて精製した。これにより、３画分（Ａ：－７分，Ｂ：－７．９分およびＣ：－１０．３分）を得た。３画分総てを一晩凍結乾燥した後、逆相ＨＰＬＣ（緩衝液Ａ：水；緩衝液Ｂ：アセトニトリル；　流速４ｍｌ／分）によって脱塩した。３画分を再度一晩凍結乾燥した後、水２００μｌに懸濁した。Ｍ．Ｓ．は、ＣＭＭ６６１Ａ　ｐｋ　１は確かに三リン酸塩ではないが、ＣＭＭ６６１ｐｋ　２または３である可能性があることを示した（Ｍ．Ｓ．プロフィールが極めて類似）。（ＣＭＭ６６２Ａは、ＣＭＭ６６１Ａ　ｐｋ　２から形成され、ＣＭＭ６６３Ａは、ＣＭＭ６６１Ａ　ｐｋ　３から形成された。）

【００９４】

次に、これらの試料を、以後の反応に用いた。重炭酸塩緩衝液（１０μｌ）およびマレイミドＮＨＳエステル（５０μｌ）をそれぞれの試料に加え、反応を一晩攪拌した。試料をｍｉｌｌｉＱ水（５００μｌ）で希釈し、濾過した。試料をＲＰ－ＨＰＬＣ　（緩衝液Ａ：　０．１Ｍ　ＴＥＡＡ，緩衝液Ｂ：　ＭｅＣＮ，流速　４ｍｌ／分）によりＭｅＣＮ５０．ｍｔｈを用いて精製し、ペプチドのカップリングをこれらの画分で行った。

【００９５】

１．１．７　修飾ペプチドタグ付きｄＵＴＰを組込んでいるプロエンケファリン核酸プローブ（４００ｂｐ）の合成

上記で合成した修飾ペプチドタグ付きｄＵＴＰを組込んでいるプロエンケファリン核酸プローブ（４００ｂｐ）（１．１．６項参照）を、１．１．５項に記載の通常のフォワードおよびリバースプライマーを用いるＲＴ－ＰＣＲによって合成した。

【００９６】

１．２　結果

１．２．１　様々な分析モードでの１個または多数のオリゴペプチド接合体の検出

３種類のオリゴペプチド接合体を、様々な分析モードで別々にまたは同時に分析した（図４）。

【００９７】

リニア正モード（図４Ａ）、リニア負モードおよびリフレクター正モードでの別々のオリゴペプチド接合体の分析は、様々なＭＡＬＤＩ分析モードで容易に検出することができる。

【００９８】

感度試験も行い、様々なＭＡＬＤＩ分析モードにおける検出閾値を評価した。観察閾値は、リニア正モードでは１００ｆｍｏｌでありシグナル／ノイズ比は８であり（図４Ｂ）、リフレクター正モードではｌｐｍｏｌでありシグナル／ノイズ比は５であった。

【００９９】

オリゴペプチド１、２および３を、更にリニア正モード、リニア負モードおよびリフレクター正モードで同時に分析した（図４Ｃ）。図４Ｃにおいて、それぞれのオリゴペプチド接合体について、異なるｍ／ｚ比の２本のピークが観察された。高い方のｍ／ｚ比のピーク（ペプチドｌ、３および２について、それぞれｍ／ｚ　１３４６．９５、１６０３．７４および１８９１．７２）は、予想開裂部位（光開裂可能なリンカーを含む）での光開裂ペプチドに対応する。低い方のｍ／ｚ比のピーク（ペプチド３および２について、それぞれｍ／ｚ　１１６１．８７および１４４９．８８）は、続いて気相で開裂してペプチドのみのプロトン化イオンを形成するものに対応すると思われる（光開裂可能なリンカーを含まない）。

【０１００】

いずれの場合にも、これらの結果は３種類の異なるオリゴペプチドを容易に同時検出することができることを示している。

【０１０１】

ＭＡＬＤＩプレート材料の、分解能、感度およびシグナル／ノイズ比に対する影響を、ステンレススチール、金およびテフロンＭＡＬＤＩプレートを用いてクエン酸または酢酸アンモニウムからなる添加剤の有無で検討した。

【０１０２】

結果は、どのような材料を用いても、クエン酸または酢酸アンモニウムがあってもなくても、光開裂可能なリンカーの開裂が可能であることを示している。

【０１０３】

更に、シグナル／ノイズ比およびシグナル強度が金のＭＡＬＤＩプレートで一層高くなると思われる（図４Ｄ）。

【０１０４】

１．２．２　固体支持体上でのオリゴペプチド接合体のマッピング

表面でのオリゴペプチド接合体のマッピングの可能性を評価するため、オリゴペプチド３からなるＸ文字のイメージをＭＡＬＤＩプレート上にマルチピペットを用いて描いた。

【０１０５】

次いで、対応する範囲をＭＡＬＤＩ－ＭＳによってスポット毎に分析し、ナトリウムカチオン化形態MNa^+でのオリゴペプチド３の光開裂断片に対応するイオンｍ／ｚ１６２６．３７の配分をイメージ構築ソフトウェアを用いて再構築した。

【０１０６】

図５は得られた再構築イメージを示し、直接開裂から簡単なイメージを得ることができることを立証している。

【０１０７】

１．２．３　オリゴペプチド接合体をプライマーとして用いるペプチドタグ付きプロエンケファリンｉｎ　ｓｉｔｕハイブリダイゼーションプローブの合成

オリゴペプチド接合体をフォワードおよび／またはリバースプライマーとして用いて一層大きなハイブリダイゼーションプローブを合成することが可能であることを確かめるた

めに、タグ付き（ペプチド－光開裂可能なリンカー－オリゴヌクレオチド）または通常の（光開裂可能なリンカーおよびペプチド残基のない）フォワード（５’－ＣＡＧ－ＧＡＣ－ＴＣＣ－ＣＣＡ－ＡＡＧ－ＧＡＧ－ＡＡＣ－ＡＧＧ－Ａ－３’）およびリバース（５’－ＧＡ－ＣＧＴ－ＡＣＣ－ＡＧＧ－ＣＧＧ－ＴＡＧ－ＣＴＧ－ＣＡＴ－ＴＴ－３’）プライマーを用いてＰＣＲ増幅を行った。

【０１０８】

生成された増幅生成物を、図６Ａに示す。得られた結果は、フォワードおよび／またはリバースペプチドタグ付きプライマーの使用が増幅を阻害せず、約４００ｂｐのプロエンケファリンｉｎ　ｓｉｔｕハイブリダイゼーションプローブが生じることを示している。

【０１０９】

シリカカラム精製の後、増幅されなかったプライマー二量体を除去し、純粋なＰＣＲ増幅生成物を得た（図６Ｂ）。

【０１１０】

精製したＰＣＲ増幅生成物のＭＡＬＤＩ分析を行った。タグ付きフォワードおよびリバースプライマー（図６Ｃ）およびタグ付きフォワードおよび通常のリバースプライマー（図６Ｄ）で得られた結果は、かなりの量のＰＥＧが含まれていても、ペプチドタグは開裂し、検出することができることを示している。実際に、開裂したペプチドのｍ／ｚ比（ｍ／ｚ＝１７０４）は幾つかの場合にはＰＥＧシグナルに帰因することがあるが、ｍ／ｚ比が１７２０の特徴的イオン$(M-NH_2)^+$は容易に検出することができる。

【０１１１】

１．２．４　ペプチドタグ付きハイブリダイゼーションプローブを用いる、パラフィン除去したラット脳切片におけるプロエンケファリンｍＲＮＡ分析

ｉｎ　ｓｉｔｕハイブリダイゼーションは、上記のようにパラフィンを除去したラット脳切片上でペプチドタグ付きハイブリダイゼーションプローブを用いて行った。

【０１１２】

次に、ＭＡＬＤＩ分析を異なる組織部位で行い、生成されるスペクトルを分析した。

【０１１３】

図７は、２箇所の異なる脳切片部位における得られたスペクトル（図７ＡおよびＢ）並びに非ペプチドタグ付きハイブリダイゼーションプローブを用いてハイブリダイズしたコントロール脳切片の分析によって得られたコントロールスペクトルを示す。

【０１１４】

コントロールスペクトルは、非ペプチドタグ付きハイブリダイゼーションプローブを用いるときには、開裂ペプチド（ｍ／ｚ＝１７０４）または特徴的イオン（ｍ／ｚ＝１７２０）のｍ／ｚ比でシグナルが検出されないことを示している。

【０１１５】

対照的に、プロエンケファリンｍＲＮＡ発現は、図７Ａの脳切片部位で明確に確認することができるが（ｍ／ｚ＝１７０４および１７２０のピーク参照）、プロエンケファリンｍＲＮＡ発現は、第二の異なる部位では検出することができない（図７Ｂ、ｍ／ｚ＝１７０４または１７２０にはピークがない）。

【０１１６】

プロエンケファリンのｍＲＮＡ発現は、このように組織切片の異なるスポットで明確に観察することができるので、組織切片プロエンケファリンｍＲＮＡ発現のイメージを更に再構築することができる。

【０１１７】

１．２．５　光開裂可能なリンカーを有するｄＵＴＰ－ペプチド接合体の合成およびマルチペプチドタグ付きプロエンケファリンプローブ　（４００ｂｐ）を生成するための上記修飾ペプチドタグ付きｄＵＴＰの使用

材料および方法の節に記載の合成プロトコルを用いて、図８Ａに示した式を有する修飾ペプチドタグ付きｄＵＴＰヌクレオチドを合成した。

【０１１８】

修飾ペプチドタグ付きｄＵＴＰをＭＡＬＤＩ分析によって分析し、ペプチドタグの検出を確認した。図８Ｂに示されるように、ペプチドタグは予想される１１６３．２３　Ｍ＋Ｈ$^+$に明確に検出される。

【０１１９】

このような修飾ペプチドタグ付きｄＵＴＰを用いることによって、プロエンケファリンについてのマルチペプチドタグ付き核酸ハイブリダイゼーションプローブが、ｄＡＴＰ、ｄＣＴＰ、ｄＧＴＰおよび図８Ａに示される修飾ペプチドタグ付きｄＵＴＰの存在下で簡単なＲＴ－ＰＣＲ増幅を用いて極めて簡単に生成した。この方法により、特異的ハイブリダイゼーションプローブを、任意の他のターゲットｍＲＮＡ配列について容易に合成することができた。

【０１２０】

修飾ペプチドタグ付きｄＵＴＰは容易にプロエンケファリンプローブに組込まれることを確認するため、修飾ペプチドタグ付きｄＵＴＰまたはタグなしの非修飾ｄＵＴＰを用いて増幅したプロエンケファリンの精製したＰＣＲ生成物をＭＡＬＤＩ分析を用いて分析した。

【０１２１】

図８で示されるように、ペプチドタグは修飾ペプチドタグ付きｄＵＴＰで増幅したプロエンケファリンの試料において予想された１１６３．０６　Ｍ＋Ｈ$^+$（図８Ｃ）で容易に検出されるが、タグを持たない非修飾ｄＵＴＰで増幅したプロエンケファリンの試料では、予想されたＭ＋Ｈ$^+$にシグナルは検出されない（図８Ｄ）。

【０１２２】

更に、修飾ペプチドタグ付きｄＵＴＰで合成した所定のハイブリダイゼーションプローブはその配列におけるＵ塩基の数と同数のタグペプチドを有するので、修飾ペプチドタグ付きｄＵＴＰの使用によりシグナル増幅が可能である。

【０１２３】

最後に、修飾ペプチドタグ付きｄＵＴＰで合成したハイブリダイゼーションプローブを用いて組織切片におけるｍＲＮＡ発現を定量的に分析することができる。実際に、１または数個の検討したｍＲＮＡの対応するタグによって生成したシグナルを、参照ｍＲＮＡ配列（例えば、ＨＰＲＴのアクチンのようなハウスキーピング遺伝子）について得たものと比較することができる。それぞれのハイブリダイゼーションプローブにおけるＵ塩基の数は知られているので、それぞれの検討したｍＲＮＡ配列と参照ｍＲＮＡ配列との間の発現比を計算することができる。

【０１２４】

１．３　結論

これらの結果は、ペプチドタグ付きプライマーで合成したハイブリダイゼーションプローブを用いて、核酸－光開裂可能なリンカー－ペプチド接合体を使用することにより組織切片中の多数のｍＲＮＡターゲット分子の間接的同時検出が可能であることを明確に示している。

【０１２５】

更に、任意のターゲットｍＲＮＡ特異性を有するペプチドタグ付きハイブリダイゼーションプローブを一層容易に合成し、タグシグナルの増幅により検出閾値を下げ、および組織切片における定量的ｍＲＮＡ発現分析を行うことができる修飾ペプチドタグ付きｄＵＴＰを合成した。

【０１２６】

従って、ＭＡＬＤＩ－ＭＳ分析を用いる組織切片における幾つかの異なるｍＲＮＡターゲット分子の間接的同時マッピングが低い検出閾値で可能であり、マルチプレックス分析および定量分析さえも可能である。

【０１２７】

実施例２

ＭＡＬＤＩ－ＭＳによる組織切片におけるタンパク質の間接検出のための光開裂可能なリ

ンカーを有する抗体－ペプチド接合体の使用

光開裂可能なリンカーを有する抗体－ペプチド接合体を合成し、ＭＡＬＤＩ－ＭＳを用いる組織切片における特異的なタンパク質を間接検出するその能力について試験した。

【０１２８】

２．１　抗体－ペプチド接合体の合成

ウサギ抗体の枠組構造領域に特異的なヤギ抗体（ヤギ抗ウサギ抗体）を用いて、ＵＶ－ＭＡＬＤＩレーザー波長で光開裂可能なリンカーを有する抗体－ペプチド接合体を合成した。この方法では、この接合体は、一次ウサギ抗体が利用可能な任意のペプチド、タンパク質、抗原またはハプテンの間接分析に対する二次抗体として適している。更に、下記の合成プロトコルは任意のペプチド、タンパク質、抗原、またはハプテンなどの任意の特異性を有する任意の抗体に応用可能であるので、抗体が別の種類の抗体または任意の所望なペプチド、タンパク質、抗原もしくはハプテンに特異的である他の接合体を下記のプロトコルを用いて容易に合成することができる点に留意しなければならない。

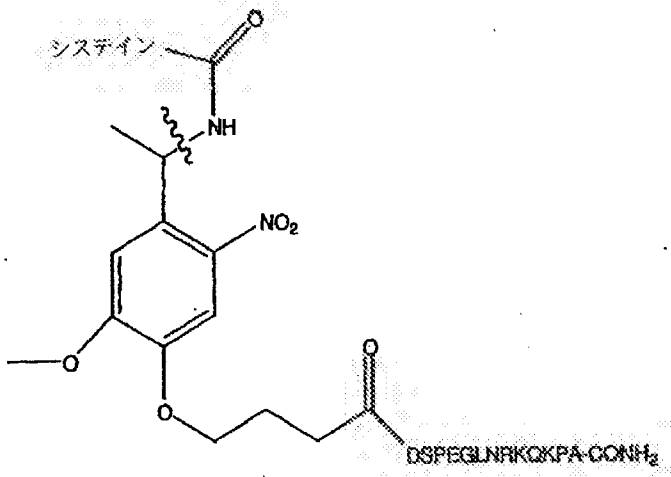
【０１２９】

タグ分子分子としては、式ＤＳＰＥＧＬＮＲＫＱＫＰＡ（配列番号：１１）を示すペプチドを用いた。しかしながら、この場合にも、下記のプロトコルはタグ分子として所望なことがある任意の他のペプチドに応用可能である。

【０１３０】

ペプチドは固相で合成して光開裂可能なリンカーに直接カップリングし、式：

【化１５】



のペプチド－光開裂可能な化合物を生成した。

【０１３１】

抗体－ペプチド接合体は、下記の反応工程を用いて合成した。

１．　光開裂可能なリンカーを用いる固相ペプチド合成（古典的Ｆｍｏｃ法）

２．　水／アセトニトリル＋０．５％　ＴＦＡを用いるＲＰ－ＨＰＬＣ（Ｃ１８）によるペプチドの精製

３．　ＭＢＳ　０．５ｍｇをＤＭＦ　３００μｌに溶解

４．　ＩｇＧ（ヤギ抗ウサギ）４ｍｇをＰＢＳ　２ｍｌに溶解

５．　生成物３および４をゆっくり混合。反応は、２０℃で３０分間撹拌下で起こる

６．　ＰＤ１０を用いて５０ｍＭリン酸緩衝液　ｐＨ　６により塩を除去

７．　ペプチド１ｍｇをＤＭＦ　３００μｌに溶解した後、ＰＢＳ　１ｍｌを加えた。

８．　このペプチド溶液を活性抗体溶液に添加すると、反応が暗所にて２０℃にて３時間撹拌下で起こる

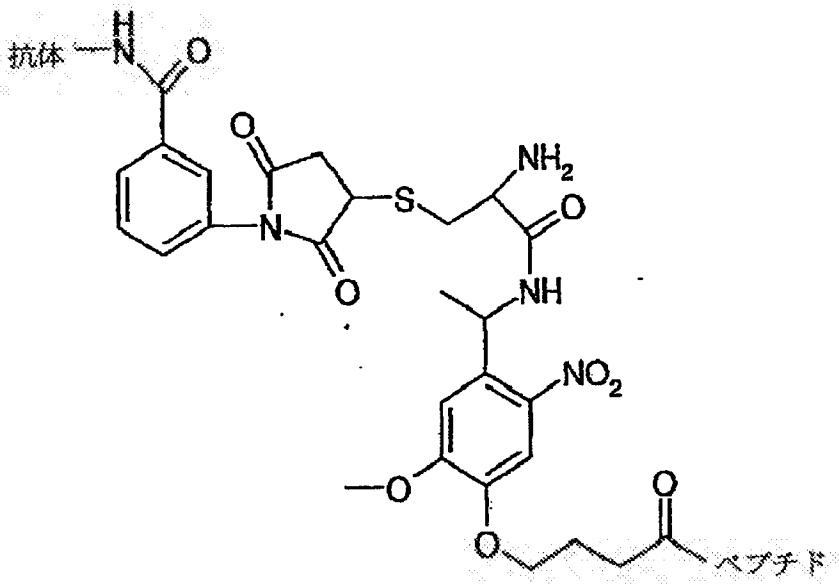
９．　ＰＢＳを一晩透析によって除去。

【０１３２】

光開裂可能なリンカーを用いる抗体－ペプチド接合体の合成の主反応を、図９に示す。

得られた抗体－ペプチド接合体は、下式：

【化１６】



を有し、図１０に省略した形態で示されている。

【０１３３】

２．２　得られた抗体－ペプチド接合体の溶液中でのＭＡＬＤＩ分析

次に、光開裂可能なリンカーを用いて得られた抗体－ペプチド接合体を、最初に溶液中で試験して、タグペプチドのＵＶ－ＭＡＬＤＩレーザーによって誘発される光解離によって効率的に開裂される能力を確かめた。

【０１３４】

結果は図１０に示されており、光開裂可能なリンカーＰＣに連結したままのペプチドタグ（ペプチド＋ＰＣ，　ｍ／ｚ＝１７０３）は、ＭＡＬＤＩ脱離／イオン化工程の後に容易に検出されることを明らかに示している。

【０１３５】

更に、タグペプチド、光開裂可能なリンカー、および最初は抗体に属するシステインを含んでなる迅速断片化による断片（図１１参照）（ペプチド＋ＰＣ＋システイン，　ｍ／ｚ＝１８２２）も検出され、得られた抗体－ペプチド接合体は、組織切片の他の型の質量分析法（例えば、ＳＩＭＳまたはＤＥＳＩ）による分析にも適していることを明確に示している。

【０１３６】

２．３　得られた抗体－ペプチド接合体を用いるラット脳組織切片のＭＡＬＤＩ直接分析

光開裂可能なリンカーを用いて得られた抗体－ペプチド接合体を、次にラット脳組織切片における膜タンパク質カルボキシペプチダーゼＤ　（１８０　ｋＤａ）の間接ＵＶ－ＭＡＬＤＩ分析について試験した。

【０１３７】

ラット脳組織切片を、最初にカルボキシペプチダーゼＤに特異的な一次ウサギ抗体とハイブリダイズした。第二工程では、光開裂可能なリンカーを有する抗体－ペプチド接合体を加えて、組織切片上に存在する一次抗体とハイブリダイズさせた。

【０１３８】

生成する着色組織切片を、次にＵＶ－ＭＡＬＤＩを用いて分析した。

【０１３９】

典型的なスペクトルを図１２に示す。これは、光開裂可能なリンカーＰＣに連結したままのペプチドタグ（ペプチド＋ＰＣ，ｍ／ｚ＝１７０３．２３）は、間接ＵＶ－ＭＡＬＤＩ組織切片分析を用いて容易に検出されることを明らかに示している。タグペプチドに対応するもう一つの断片（ペプチド＋ＰＣ＋Ｏ，ｍ／ｚ＝１６８６．４３）も、容易に検出

される。

【０１４０】

本発明による方法を用いて、ラット脳組織切片を走査し、それぞれのスペクトルデータを保存し、カルボキシペプチダーゼＤの発現マップのイメージを構築した。結果を図１３に示されており、光開裂可能なリンカーを有する抗体－ペプチド接合体を用いる本発明による方法を用いて発現マップを構築することができることを明らかに示している。

【０１４１】

更に、発現マップを構築する本発明による方法の信頼性を更に明らかにするため、ラット脳組織切片をカルボキシペプチダーゼＤに特異的なウサギ一次抗体で染色し、古典的な市販の二次抗体上に固定した後にＴ－クロロナフトールを色原体として用いて顕色した。得られた染色はカルボキシペプチダーゼＤのＭＡＬＤＩ検出と共存し（データは示さず）、従って、本発明による方法とＭＡＬＤＩイメージングの信頼性は古典的方法より感度が高いことを示している。

【０１４２】

これらの結果は、光開裂可能なリンカーを用いて合成した抗体－ペプチド接合体は組織切片におけるタンパク質の間接検出並びに発現マップの構築を行うことができることを示している。

【０１４３】

実施例３

組織切片に存在する生体分子の間接質量分析法による分析に、光解離の代わりに迅速断片化を用いることができることの証明

迅速断片化は、通常は有害であるが避け難い現象と考えられている。

【０１４４】

本発明の場合には、本発明者らは、（Ａ－Ｘ）ｎ－Ｂ接合体であって、リンカーＸが迅速断片化によって開裂されるものを用いて、組織切片における少なくとも１種類のターゲット分子マップを決定する方法を行うのに、この必要な現象を利用することができることを見出した。

【０１４５】

３．１　リンカーがＵＶ－ＭＡＬＤＩレーザーの波長で光開裂可能な接合体を用いる組織切片に存在する生体分子の間接質量分析法による分析のための迅速断片化

このコンセプトはＳＩＭＳまたはＤＥＳＩ質量分析法のような他の質量分析法の技術に置き換えることができるが、本発明者らは、ＭＡＬＤＩ質量分析法と、リンカーがＵＶ－ＭＡＬＤＩレーザーの波長で光開裂可能な接合体を用いるこのコンセプトの使用が可能であることを最初に明らかにした。

【０１４６】

実際に、オリゴペプチド接合体または抗体－ペプチド接合体を、光開裂可能なリンカー（ＰＣ）と共に用いると、予想された（ペプチド＋ＰＣ）断片の他に、ペプチドタグを含んでなる他の断片が観察され、これらの断片は接合体の迅速断片化による断片に対応している。

【０１４７】

更に正確には、３種類の異なるオリゴペプチド接合体を光開裂可能なリンカーと共に用いるときには（これらの接合体の更に精確な定義については、実施例１を参照）、タグペプチド１、ペプチド２またはペプチド３に対応する断片が、図４Ｃで明確に示されているように、予想された断片（ペプチド１、２または３＋ＰＣ）の他に観察される。例えば、ペプチド２タグ分子を用いる接合体２については、光解離断片と迅速断片化による断片の両方が観察される：

予想された光解離断片＝（ペプチド２＋ＰＣ），ｍ／ｚ＝１８９２；および

迅速断片化による断片＝ペプチド２のみ，ｍ／ｚ＝１４５０。

【０１４８】

同じ現象は、図４Ｃで接合体３について観察される：

予想された光解離断片＝（ペプチド３＋ＰＣ），ｍ／ｚ＝１６０４；および
迅速断片化による断片＝ペプチド３のみ，ｍ／ｚ＝１１６２。

【０１４９】

例２に記載の抗体－ペプチド接合体を用いるときにも、予想された光解離断片と迅速断片化による断片との両方が図１１に示されるように観察される：
予想された光解離断片＝（ペプチド＋ＰＣ），ｍ／ｚ＝１７０２；および
迅速断片化による断片＝ペプチド＋ＰＣ＋システイン，ｍ／ｚ＝１８２２。

【０１５０】

３．２　ＦＩＴＣ標識抗体を用いる組織切片に存在する生体分子の間接質量分析法による分析のための迅速断片化

本発明者らは、通常の、特に市販の標識抗体を用いる組織切片に存在する生体分子の間接質量分析法による分析のために迅速断片化を用いる可能性についても分析した。

【０１５１】

実際に、通常の標識分子は一定の分子量を有しており、従ってタグ分子として用いることができる。更に、極めて広範囲の抗原特異性を有するあらゆる種類の抗体は、標識抗体として市販されている。最後に、従来のリンカーを介して標識分子を任意の抗体に結合する技術は周知の日常的技術であり、従って組織切片に存在する生体分子の間接質量分析法による分析に通常の標識抗体を用いることができることにより、本発明による方法を用いてマッピングすることができる異なる生体分子の数を確実に増やすことができる。

【０１５２】

ＦＩＴＣ標識抗体を用いて、ＣＨＣＡマトリックスを用いるＦＩＴＣ標識抗体の溶液ＭＡＬＤＩ分析を行った。

【０１５３】

結果は図１４に示されており、ＦＩＴＣ分子の特徴的シグナルが容易に検出されることを示している（ｍ／ｚ＝３６５、１８）。抗体とＦＩＴＣ分子との間のリンカーは、ＭＡＬＤＩレーザー波長で光開裂可能でないので、迅速断片化が、抗体とその標識分子との間の結合を開裂することができることを明らかに示している。

【０１５４】

これらの結果は更に組織切片で確かめなければならないが、溶液で得られた総ての上記結果はさらに組織切片上で確かめられている。従って、これらの結果は、ＭＡＬＤＩ質量分析法だけでなく、迅速断片化を含む他の質量分析法の技術、特にＳＩＭＳまたはＤＥＳＩ質量分析法、を用いる組織切片におけるターゲット分子の間接検出に、迅速断片化および通常の標識抗体を用いることができることを強く支持している。

【０１５５】

３．３　結論

得られた結果は、様々な質量分析法の技術、特にＭＡＬＤＩ、ＳＩＭＳまたはＤＥＳＩ質量分析法を用いる、組織切片におけるターゲット分子の間接検出に、通常は有害な迅速断片化の現象を利用することができることを明らかに示している。

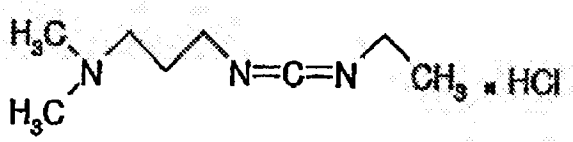
【０１５６】

実施例４

抗体とＢｒ原子との間のリンカー分子を開裂するための断片化を用いて組織切片の質量分析法による分析に用いるための抗体－Ｂｒ接合体の合成

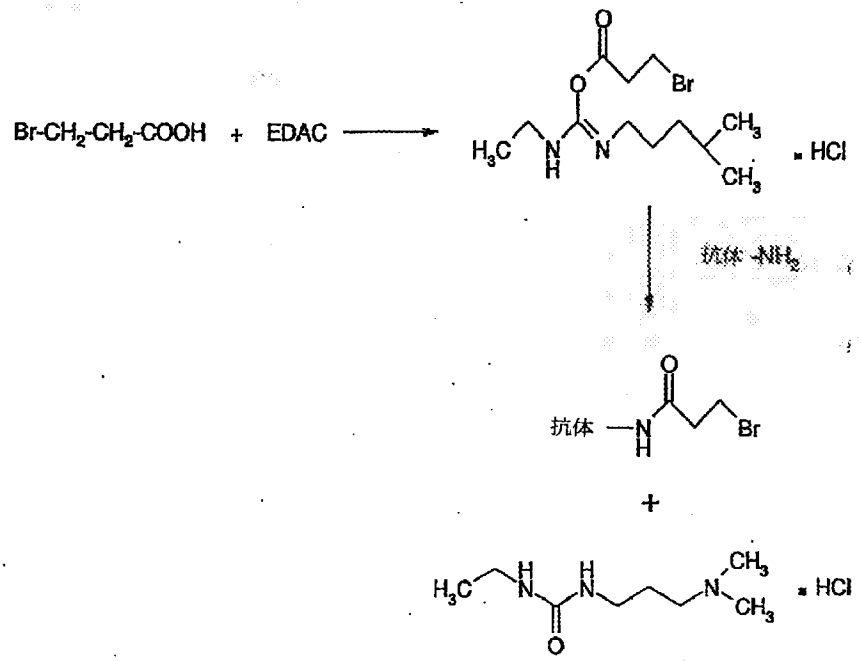
抗体－Ｂｒ接合体は、ＳＩＧＭＡから製品番号Ｅ１７６９で発売されているＥＤＡＣリンカー

【化１７】



を用い、下記のプロトコル

【化１８】



を用いて合成することができる。

【０１５７】

このようなプロトコルは任意の抗体または抗体断片を用いて適用され、これをＢｒ原子と接合させることができる。このような接合体を、断片化を用いて抗体とＢｒ原子との間のリンカーを開裂する本発明による方法で用いることができる。

【図面の簡単な説明】

【０１５８】

【図１】ＭＡＬＤＩ脱離／イオン化工程中のタグ分子放出および免疫細胞化学技術による間接検出のために特異的にデザインされた標識プローブを組み合わせるＭＡＬＤＩによるｍＲＮＡ特異的マルチプレックスイメージングの模式的原理。

【図２】異なる標識プローブ（オリゴヌクレオチド、抗体、レクチン、アプタマー）を用いる異なるターゲット（ペプチド、タンパク質、ｍＲＮＡ、糖質、薬剤）の可能な特異的イメージングの概略図。

【図３】ＭＡＬＤＩ脱離／イオン化工程中のタグ分子放出および免疫細胞化学技術による間接検出のために特異的にデザインされた標識プローブを組み合わせる、ＭＡＬＤＩによるペプチド／タンパク質マルチプレックス特異的イメージングの模式的原理。

【図４】Ａ．リニア正モードでのオリゴペプチド３、Ｂ．リニア正モードでの１００ｆｍｏｌオリゴペプチド１、Ｃ．リフレクター正モードでのオリゴペプチド１、２および３の同時の、およびＤ．金製のＭＡＬＤＩ支持プレート上のオリゴペプチド３、のペプチド残基の検出。

【図５】固体支持体上のオリゴペプチド３のマッピング。オリゴペプチド３を固体支持体上にマルチピペットで沈着させ、Ｘ文字のイメージを描いた。次に、オリゴペプチド３の存在をＭＡＬＤＩ－ＭＳ分析によってマッピングし、イメージをＩＤＬソフトウェアを用いて再構築した。生成するイメージを示す。

【図６】ペプチドタグ付きプロエンケファリンのｉｎ　ｓｉｔｕハイブリダイゼーションプローブの合成。Ａ、タグ付き（ペプチド－光開裂可能なリンカー－オリゴヌクレオチド）または通常の（光開裂可能なリンカーおよびペプチド残基のない）フォワード（５’－ＣＡＧ－ＧＡＣ－ＴＣＣ－ＣＣＡ－ＡＡＧ－ＧＡＧ－ＡＡＣ－ＡＧＧ－Ａ－３’）およびリバース（５’－ＧＡ－ＣＧＴ－ＡＣＣ－ＡＧＧ－ＣＧＧ－ＴＡＧ－ＣＴＧ－ＣＡＴ－ＴＴ－３’）オリゴヌクレオチドを用いて得たプロエンケファリンのＰＣＲ増幅生成物。ラ

イン１：通常のフォワードおよびリバースプライマー。ライン２：タグ付きフォワードおよび通常のリバースプライマー。ライン３：通常のフォワードおよびタグ付きリバースプライマー。ライン４：タグ付きフォワードおよびタグ付きリバースプライマー。ライン５：負コントロール：水コントロール。ライン６：分子量マーカー。Ｂ．シリカカラム上での精製後の対応する増幅生成物。ライン１：通常のフォワードおよびリバースプライマー。ライン２：タグ付きフォワードおよび通常のリバースプライマー。ライン３：通常のフォワードおよびタグ付きリバースプライマー。ライン４：タグ付きフォワードおよびタグ付きリバースプライマー。ライン５：負のコントロール：水コントロール。ライン６：分子量マーカー。Ｃ．およびＤ．タグ付きフォワードおよびリバースプライマー（Ｃ）、またはタグ付きフォワードおよび通常のリバースプライマー（Ｄ）を用いた精製ＰＣＲ増幅生成物のＭＡＬＤＩ分析。

【図７】ペプチドタグ付きプロエンケファリンｉｎ　ｓｉｔｕハイブリダイゼーションプローブのハイブリダイゼーション後のパラフィンを除去したラット脳切片の直接ＭＡＬＤＩ分析。Ａ．第一の組織切片局在。Ｂ．第二の異なる組織切片局在。Ｃ．非ペプチドタグ付きプロエンケファリンｉｎ　ｓｉｔｕハイブリダイゼーションプローブによるコントロール分析。

【図８】修飾ペプチドタグ付きｄＵＴＰを組込んでいる核酸プローブの合成およびＭＡＬＤＩ分析。Ａ．修飾ペプチドタグ付きｄＵＴＰの構造および修飾ペプチドタグ付きｄＵＴＰを組み込んでいるプロエンケファリン核酸プローブ（４００ｂｐ）のＲＴ－ＰＣＲ合成の略図。Ｂ．修飾ペプチドタグ付きｄＵＴＰのＭＡＬＤＩ分析。Ｃ．修飾ペプチドタグ付きｄＵＴＰを組込んでいるプロエンケファリン核酸プローブのＭＡＬＤＩ分析。Ｄ．ペプチドタグのない通常の非修飾ｄＵＴＰを組込んでいるプロエンケファリン核酸プローブのＭＡＬＤＩ分析。

【図９】光開裂可能なリンカーＸを用いる抗体－ペプチド接合体の合成反応。

【図１０】光開裂可能なリンカーＸを有する抗体－ペプチド接合体の構造。ＭＡＬＤＩ質量分析法によって検出された２個の主要断片（光開裂可能なリンカーの開裂を介するタグの光解離によって得られた第一の断片、および分子の退速断片化によって得られた第二の断片）を示す。

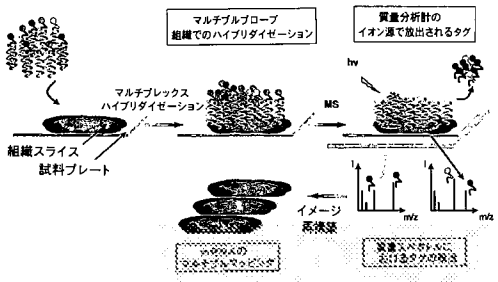
【図１１】光開裂可能なリンカーＸを用いる抗体－ペプチド接合体の溶液ＭＡＬＤＩ分析中に得られる典型的な質量スペクトル（マトリックス　シナピン酸　ＳＡ）。

【図１２】接合体を用いない場合（Ａ，ラット脳コントロール）と光開裂可能なリンカーを有する抗体－ペプチド接合体を用いる場合（Ｂ）の、ラット脳組織切片の直接分析中に得られる典型的な質量スペクトル。

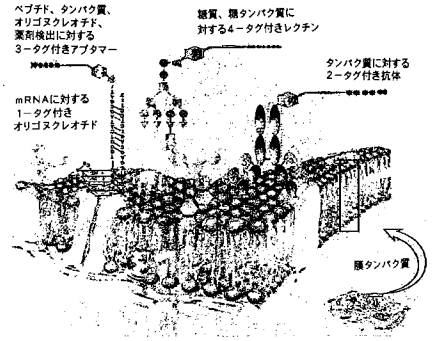
【図１３】膜タンパク質カルボキシペプチダーゼＤ（１８０ｋＤａ）に対して向けられたウサギ一次抗体と、光開裂可能なリンカーを有する抗体－ペプチド接合体とによって検出された該タンパク質の発現マップ。このイメージは、ペプチドタグに対応する検出シグナルから再構築されている。

【図１４】ＣＨＣＡマトリックスを用いるＦＩＴＣ標識抗体の溶液ＭＡＬＤＩ分析中に得られる典型的な質量スペクトル。Ａ．ＦＩＴＣ標識抗体なし（ＣＨＣＡマトリックスのみ）。Ｂ．ＦＩＴＣ標識抗体あり。

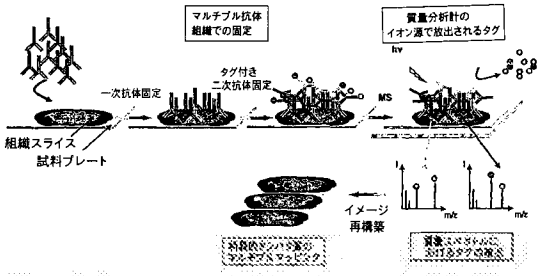
【図１】



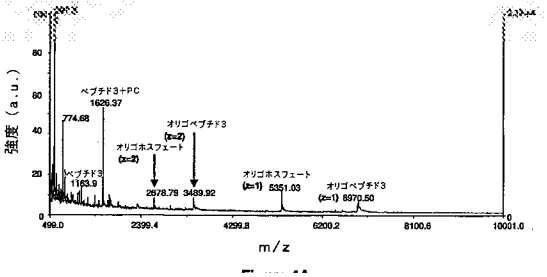
【図２】



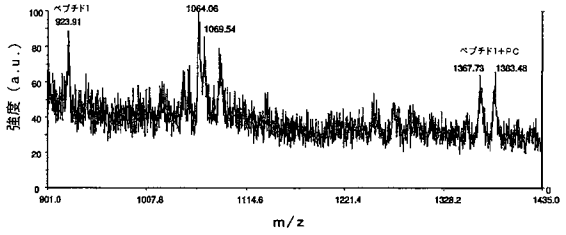
【図３】



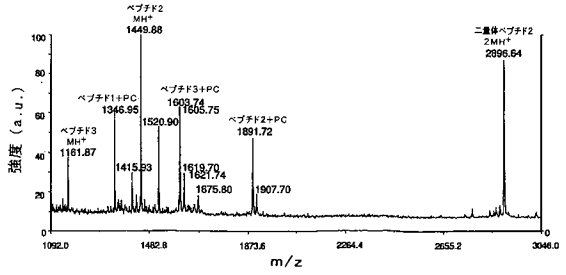
【図４Ａ】



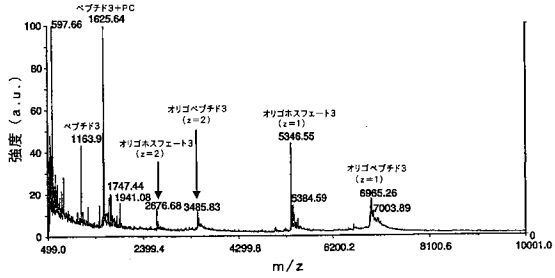
【図４Ｂ】



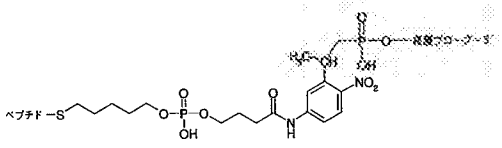
【図４Ｃ】



【図４Ｄ】



【図４Ｅ】



【図５】

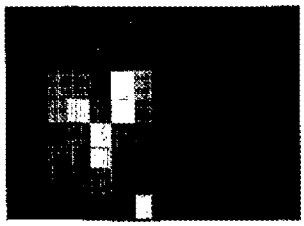
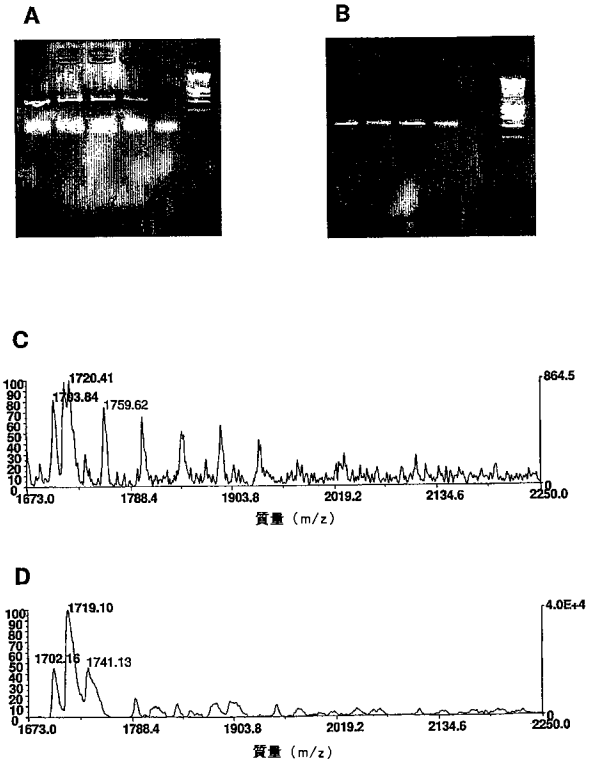
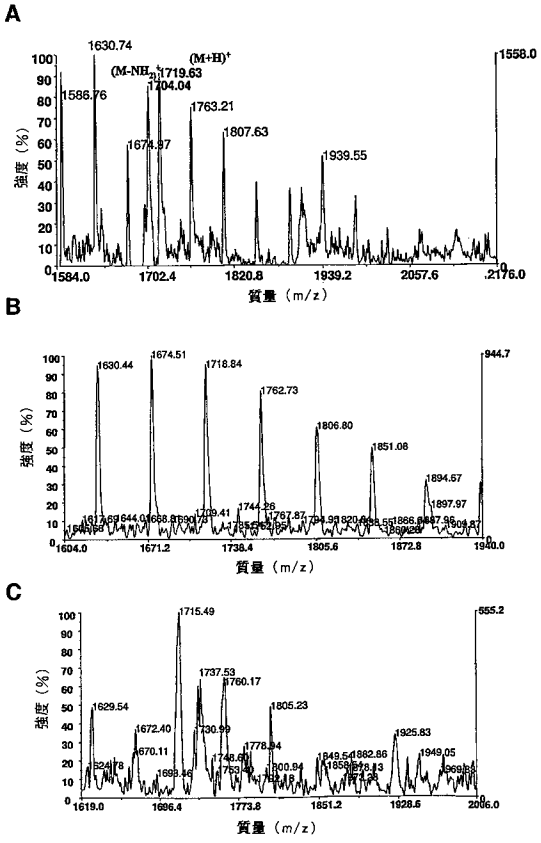


Figure 5

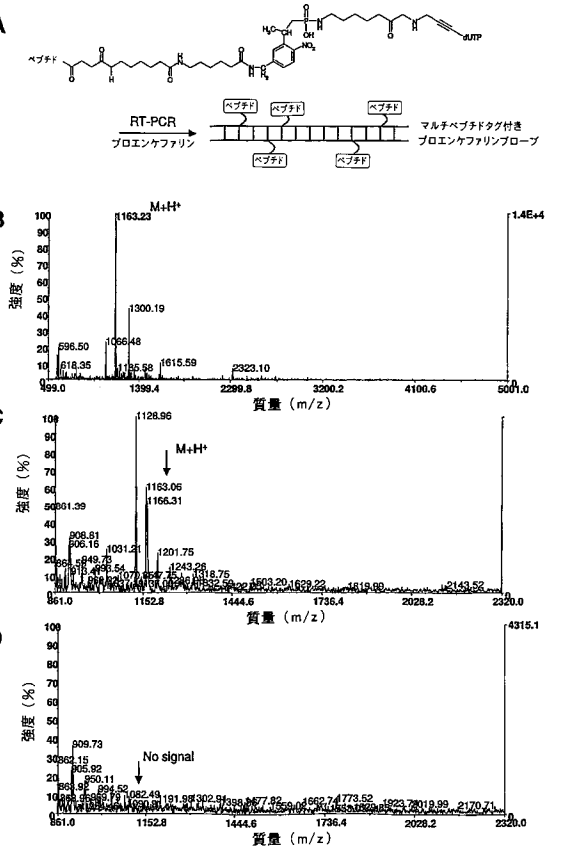
【図６】



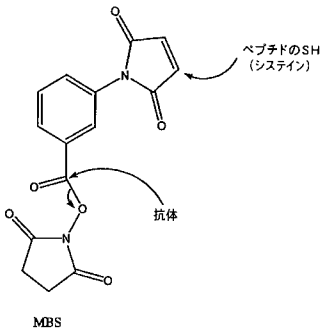
【図７】



【図８】

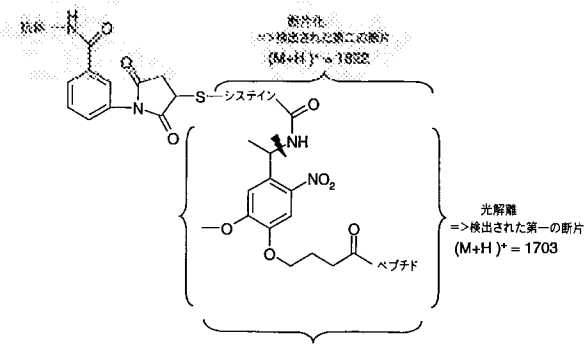


【図９】

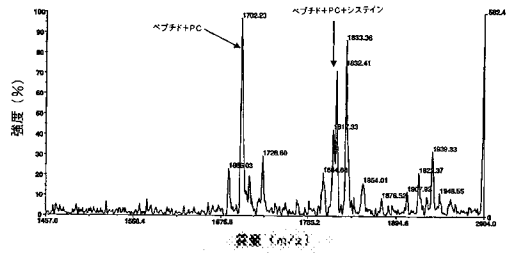


MBS

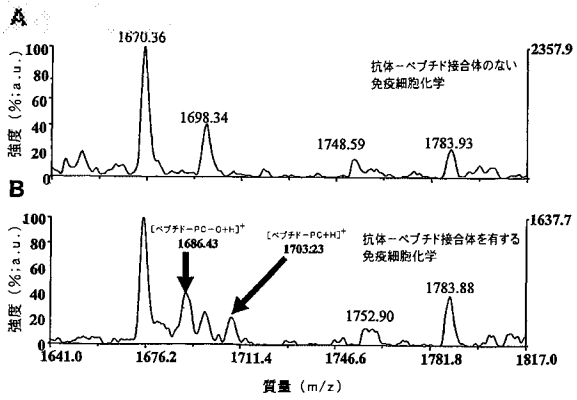
【図１０】



【図１１】



【図１２】



【図１３】

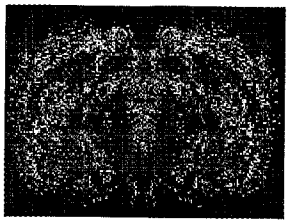
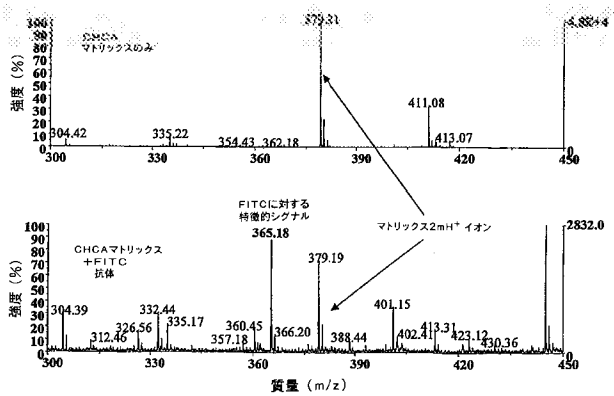


Figure 13

【図１４】



【配列表】
2008542783000001.xml

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2006/002309

A. CLASSIFICATION OF SUBJECT MATTER
INV. G01N30/72 G01N33/483 H01J49/16 G01N33/68 C12Q1/68
A61K47/48

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G01N

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, BIOSIS, EMBASE, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	WO 2005/067648 A2 (UNIV VANDERBILT [US]; LEVY SHAWN UNIV VANDERBILT [US]; LEVY SHAWN [US]) 28 July 2005 (2005-07-28) page 6 page 52 figure 4	1-12,20, 21
X	RICCARDO LEVI-SETTI, JAN M. CHABALA, KONSTANTIN GAVRILOV, RAFAEL ESPINOSA, MICHELLE M. LE BEAU: "Advances in high resolution studies of BrdU-labelled human metaphase chromosomes with a high resolution scanning ion microprobe" MICROSCOPY RESEARCH AND TECHNIQUE, vol. 36, 7 December 1998 (1998-12-07), pages 301-312, XP002413911 page 306	1-3,5,6, 20,23

-/--

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents :

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

Date of the actual completion of the international search	Date of mailing of the international search report
24 May 2007	08/10/2007

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hohwy, Morten

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2006/002309

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	OLEJNIK J ET AL: "Photocleavable peptide-DNA conjugates: synthesis and applications to DNA analysis using MALDI-MS" 1 December 1999 (1999-12-01), NUCLEIC ACIDS RESEARCH, OXFORD UNIVERSITY PRESS, SURREY, GB, PAGE(S) 4626-4631 , XP002220753 ISSN: 0305-1048 figure 2 abstract	1-23
Y	BAI XIAOPENG ET AL: "Design and synthesis of a photocleavable biotinylated nucleotide for DNA analysis by mass spectrometry." NUCLEIC ACIDS RESEARCH, vol. 32, no. 2, 2004, pages 535-541, XP002413912 ISSN: 0305-1048 figure 1 abstract	1-23
Y	BAI XIAOPENG ET AL: "Photocleavage of a 2-nitrobenzyl linker bridging a fluorophore to the 5' end of DNA." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. 21 JAN 2003, vol. 100, no. 2, 21 January 2003 (2003-01-21), pages 409-413, XP002413913 ISSN: 0027-8424 figures 1,4 abstract	1-23
A	STOECKLI M ET AL: "Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues" April 2001 (2001-04), NATURE MEDICINE, NATURE PUBLISHING GROUP, NEW YORK, NY, US, PAGE(S) 493-496 , XP002979722 ISSN: 1078-8956 the whole document	1-23
Y	US 6 635 452 B1 (MONFORTE JOSEPH A [US] ET AL) 21 October 2003 (2003-10-21) figure 9 column 16	1-23

-/--

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2006/002309

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WALTER K: "An inside attack on cancer" SCIENCE AND TECHNOLOGY REVIEW, LAWRENCE LIVERMORE NATIONAL LABORATORY, 31 August 2003 (2003-08-31), XP002414040 pages 10-11	1,2,4-7
T	US 2006/198820 A1 (MCDONALD JOHN R [CA] ET AL) 7 September 2006 (2006-09-07) paragraphs [0244], [0264], [0300]	24
A	WO 2004/051270 A (NOVARTIS AG [CH]; NOVARTIS PHARMA GMBH [AT]; AUER MANFRED [AT]; MEISNE) 17 June 2004 (2004-06-17) examples 27-32	24
T	US 2006/121535 A1 (BRUEGGEMEIER SHAWN B [US] ET AL) 8 June 2006 (2006-06-08) figure 21	24

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IB2006/002309

Box II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

1-23, 24 (partly) concerning invention 1 and 6

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2004)

International Application No. PCT/IB2006/002309

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-23

A method of carrying out imaging mass spectroscopy using cleavable tags.

---

2. claim: 24 (partly)

Compound 1 of claim 24 directed to photocleaveable compounds useful for the method of claims 1-23.

---

3. claim: Claim 24 (partly)

Compound 2 of claim 24 directed to photocleaveable compounds useful for the method of claims 1-23.

---

4. claim: Claim 24 (partly)

Compound 3 of claim 24 directed to photocleaveable compounds useful for the method of claims 1-23.

---

5. claim: Claim 24 (partly)

Compound 4 of claim 24 directed to photocleaveable compounds useful for the method of claims 1-23.

---

6. claim: Claim 24 (partly)

Compound 5 of claim 24 directed to photocleaveable compounds useful for the method of claims 1-23.

---

7. claim: Claim 24 (partly)

Compound 6 of claim 24 directed to photocleaveable compounds useful for the method of claims 1-23.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2006/002309

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2005067648 A2	28-07-2005	NONE	
US 6635452 B1	21-10-2003	US 2003022225 A1	30-01-2003
US 2006198820 A1	07-09-2006	NONE	
WO 2004051270 A	17-06-2004	AU 2003293766 A1 US 2006134691 A1	23-06-2004 22-06-2006
US 2006121535 A1	08-06-2006	NONE	

Form PCT/ISA/210 (patent family annex) (April 2005)

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LS,MW,MZ,NA,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HU,IE,IS,IT,LT,LU,LV,MC,NL,PL,PT,RO,SE,SI,SK,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BW,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KM,KN,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,LY,MA,MD,MG,MK,MN,MW,MX,MZ,NA,NG,NI,NO,NZ,OM,PG,PH,PL,PT,RO,RS,RU,SC,SD,SE,SG,SK,SL,SM,SY,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,ZA,ZM,ZW

(特許庁注：以下のものは登録商標)
1．テフロン

(71)出願人 507403012
ウロジャンテック、ソシエテ、アノニム
ＥＵＲＯＧＥＮＴＥＣ ＳＡ
ベルギー国ベー・４１０２、セラン、リュ、ボワ、サン・ジャン、５、リエジュ、シアンス、パーク
(74)代理人 100075812
弁理士 吉武 賢次
(74)代理人 100091487
弁理士 中村 行孝
(74)代理人 100094640
弁理士 紺野 昭男
(74)代理人 100107342
弁理士 横田 修孝
(74)代理人 100137497
弁理士 大森 未知子
(72)発明者 レミ、ルメール
フランス国スシー、アン、ブリ、ビス、リュ、セビニェ、１７
(72)発明者 イザベル、フルニエ
フランス国ブルジェール、リュ、デュ、８、メ、１９４５、１０
(72)発明者 ミッシェル、サルゼ
フランス国ブルジェール、リュ、デュ、８、メ、１９４５、１０
(72)発明者 ミッシェル、デシャン
ベルギー国９８００、ベルビエ、リュ、ド、ラ、シテ、７８
(72)発明者 エドモン、ジャン・クロード、タベ
フランス国モランジ、リュ、デ、イロンデル、２３
(72)発明者 ゴットフリード、プロエス
ベルギー国ベー・４７５０、ブッゲンバッハ、ツム、バルケルスタル、１２
(72)発明者 イボ、ルドロフ
ベルギー国ベー・４７００、ウペン、オペレ、ロッテルガッセ、１４
(72)発明者 マルク、ルメートル
ベルギー国ベー・４４３２、アリュール、クロ、デルメ、３３
Ｆターム(参考) 2G041 CA01 DA16 EA12 FA10 FA11 FA12

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 January 2007 (04.01.2007)

PCT

(10) International Publication Number
WO 2007/000669 A2

(51) International Patent Classification:
G01N 33/68 (2006.01) *C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/IB2006/002309

(22) International Filing Date: 7 June 2006 (07.06.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/687,848 7 June 2005 (07.06.2005) US

(71) Applicants *(for all designated States except US)*: **CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)** [FR/FR]; 3, rue Michel-Ange, F-75016 Paris (FR). **UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE** [FR/FR]; F-59650 Villeneuve D'Ascq (FR). **EUROGENTEC SA** [BE/BE]; Liège Science Park, Rue Bois Saint-Jean 5, B-4102 Seraing (BE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LEMAIRE, Rémi** [FR/FR]; 17 bis rue Sévigné, F-94370 Sucy en Brie (FR). **FOURNIER, Isabelle** [FR/FR]; 10 rue du 8 mai 1945, F-59830 Bourghelles (FR). **SALZET, Michel** [FR/FR]; 10 rue du 8 mai 1945, F-59830 Bourghelles (FR). **DESCHAMPS, Michel** [BE/BE]; 78 rue de la Cité, F-9800 Verviers (BE). **TABET, Edmond, Jean-Claude** [FR/FR]; 23 rue des Hirondelles, F-91420 Morangis (FR). **PROESS, Gottfried** [BE/BE]; Zum Walkerstal 12, B-4750 Butgenbach (BE). **RUDLOFF, Ivo** [BE/BE]; Obere Rottergasse 14, B-4700 Eupen (BE). **LEMAITRE, Marc** [BE/BE]; Clos del'Mé 33, B-4432 Alleur (BE).

(74) Agents: **SCHRIMPF, Robert,** et al.; Cabinet Regimbeau, 20, rue de Chazelles, F-75847 Paris Cedex 17 (FR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: USE OF CONJUGATES WITH LINKERS CLEAVABLE BY PHOTODISSOCIATION OR FRAGMENTATION FOR MASS SPECTROMETRY ANALYSIS OF TISSUE SECTIONS

(57) Abstract: The invention concerns a method for determining at least one target molecule map in a tissue section, using at least one (A-X)n-B conjugate, wherein A is a tag molecule of known molecular weight, X is a linker that is cleaved during sample desorption/ionization, n is an integer of at least l,and B is a binding molecule that binds specifically to said target molecule. When using MALDI mass spectrometry, said linker molecule X may be cleaved by photodissociation during sample laser irradiation if photocleavable at the wavelength of said MALDI laser. Alternatively, when using UV- MALDI, IR-MALDI, SIMS or DESI mass spectrometry, said linker molecule X may be cleaved by fragmentation during sample desorption/ionization.

WO 2007/000669 A2

Use of conjugates with linkers cleavable by photodissociation or fragmentation for mass spectrometry analysis of tissue sections

The invention concerns a method for determining at least one target molecule map in a tissue section, using at least one $(A\text{-}X)_n\text{-}B$ conjugate, wherein A is a tag molecule of known molecular weight, X is a linker that is cleaved during sample desorption/ionization, n is an integer of at least 1,and B is a binding molecule that binds specifically to said target molecule. When using MALDI mass spectrometry, said linker molecule X may be cleaved by photodissociation during sample laser irradiation if photocleavable at the wavelength of said MALDI laser. Alternatively, when using UV-MALDI, IR-MALDI, SIMS or DESI mass spectrometry, said linker molecule X may be cleaved by fragmentation during sample desorption/ionization.

Recently, transcriptome and proteome studies have led to the identification of many proteins implicated in a wide diversity of diseases, such as several kinds of cancers.

However, most of these results have been obtained on purified extracted nucleic acid or protein samples, which do not generate information on the tissue localisation of the incriminated proteins, although this kind of information is crucial for the understanding of physiological processes. Another shortcoming of most of the available data is that few studies have simultaneously analyzed the mRNA and protein expression of molecules of interest, which is nevertheless important to clarify the expression pattern of a particular protein. It would thus be very useful to have combined data of both mRNA and protein expression maps.

Current technologies for the tissue expression mapping of mRNA usually resort to nucleic acid probes coupled to radioactive, fluorescent or chemoluminescent markers after in situ hybridization (ISH). For protein expression mapping in tissue sections, usual technologies include immunohistochemistry and immunofluorescence.

The main drawback of all these technologies of tissue section expression mapping is that the number of target molecules that can be analyzed simultaneously is limited, even with fluorescent probes or antibodies, since no more than 3 or 4 distinct target molecules can be studied in the same experiment.

Mass spectrometry, on the other hand, allows for the simultaneous multiplex analysis of complex mixtures of biomolecules, depending on their molecular weight. In particular, Matrix Assisted Laser Desorption/Ionization (MALDI) mass spectrometry has become a powerful tool in the field of biological researches and is used for the detection, identification and characterization of nucleic acids, peptides and proteins from complex mixtures.

In particular, Olejnik and collaborators describe the synthesis and characterization of photocleavable peptide-DNA conjugates along with their use as photocleavable mass marker (PCMM) hybridization probes for the detection of immobilized synthetic target DNAs by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Olejnik et al. Nucleic Acids Res. 1999 Dec 1;27(23):4626-31).

Patent application WO 98/26095 describes the synthesis and use of mass-labeled compounds to specifically interact with biomolecular targets. Patent application WO 00/68434 describes a method of detecting multiple analytes in a sample in a single assay, based on encoding target molecules with signals followed by decoding of the encoded signal.

However, no document in the prior art describes a method for the MALDI mass spectrometry (MALDI-MS) analysis of nucleic acids comprised in tissue sections.

The inventors have set up a new method for the detection of biomolecules, in particular mRNA, in tissue sections by MALDI-MS, using conjugates composed of a moiety that binds specifically to the target molecule(s), a moiety of known molecular weight (the "tag" moiety), and a photocleavable linker that is directly cut by the MALDI laser during the ionization process (see Figure 1). This new method allows for the easy and precise indirect mapping of mRNA in tissue sections, and the use of tag

moieties of widely dispersed molecular weights makes it possible to analyze a high number of distinct biomolecules simultaneously.

In addition, when this new method is used to map tissue expression of mRNA, a common picture of mRNA and corresponding protein expression can be obtained in two consecutive tissue sections. Indeed, several publications have shown that MALDI-MS could become an efficient tool for direct analysis of peptides and proteins in tissue sections (Caprioli, R.M.; Farmer, T.B.; Gile, J. *Anal. Chem.* **1997**, *69*, 4751-4760; Stoeckli, M.; Farmer, T.B.; Caprioli, R.M. *Nat. Med.* **2001**, *7*, 493-496; Chaurand, P.; Schwartz,. S.A.; Caprioli, R.M. *Anal. Chem.* **2004**, 87A-93A).

Finally, the inventors have found that such a method is also transposable to mass spectrometry in general provided that the photocleavable linker is replaced by a linker which is cleaved by fragmentation during sample desorption/ionization. In this case, MALDI (UV-MALDI, or IR-MALDI), SIMS (Second Ion Mass spectrometry) or DESI (Desorption Electrospray Ionization) mass spectrometry may be used instead of MALDI, and preferably UV-MALDI, only. In addition, the inventors have found that conjugates with photocleavable linkers suitable for UV-MALDI analysis may also be used with other mass spectrometry technologies, since the chosen photocleavable linkers are also cleaved by fragmentation during sample desorption/ionization. The particular conjugates set up by the inventors may thus be used with any mass spectrometry technology suitable for tissue section analysis, notably with MALDI (UV-MALDI, or IR-MALDI), SIMS (Second Ion Mass spectrometry) or DESI (Desorption Electrospray Ionization) mass spectrometry.

The invention thus concerns a method for determining at least one target molecule map in a tissue section, comprising :

a) hybridizing said tissue section with at least one $(A\text{-}X)_n\text{-}B$ conjugate, wherein

- A is a tag molecule of known molecular weight,
- X is a linker that is cleaved during sample desorption/ionization,
- n is an integer of at least 1,
- B is a binding molecule that binds specifically to said target molecule, and

- each distinct B molecule is linked to a distinct A tag molecule;

b) scanning the tissue section surface and analyzing each adjacent spot with a mass spectrometer, wherein said linker X is cleaved during sample ionization, and wherein the resulting data of each spot is saved; and

c) analyzing the obtained data in the molecular mass window(s) of each distinct tag molecule to create as many maps of the tissue section as the number of distinct studied target molecules.

According to the invention, a "tissue section" preferably has the following properties: it may be frozen or paraffin-embedded, its thickness is preferably in the order of a mammalian cell diameter, thus comprised between 5 and 20 μm. In the case of a frozen section that was obtained from a frozen tissue using a cryostat, OCT (optimal cutting temperature polymer) is preferably used only to fix the tissue but the frozen tissue is not embedded in OCT, so that tissue sections were not brought into contact with OCT. The tissue section may then be transferred on a MALDI plate composed of any material suitable for further MALDI analysis, including metals, inorganic or organic materials, such as gold, steel, glass fiber, glass, nylon 6/6, silicon, plastic, polyethylene, polypropylene, polyamide, polyvinylidenedifluoride or a glass slice of any thickness coated with conductive metal keeping transparency properties such as nickel or ITO.

According to the invention, a "target molecule" means a molecule of interest that is capable to specifically bind to another molecule, which is referred to as a "binding molecule". Such tandem target/binding molecules may display any chemical structure, providing that they are capable to generate a specific hybridization in a tissue section. A large variety of tandem target/binding molecules, as well as tandem binding/target molecules, are comprised in the scope of the present invention, including nucleic acids/nucleic acids, nucleic acids/peptides, nucleic acids/proteins, nucleic acids/antibodies, peptides/peptides, peptides/proteins, peptides/antibodies, proteins/proteins (in particular ligands/receptors), proteins/sugars, antigens/antibodies, haptens/antibodies, organic compounds/receptor (see some example in Figure 2).

In particular, target nucleic acid sequences may be specifically detected using single stranded nucleic acid probes with a nucleic sequence that is complementary to the single-stranded target nucleic acid or to one of the strands of a double-stranded target nucleic acid (see Figure 1). In the case of a target mRNA molecule, a nucleic acid probe complementary to the mRNA sequence may be used as binding molecule. Such a nucleic acid probe preferably has a nucleotide length comprised between 250 and 550, more preferably 300 to 500, 350 to 450, most preferably around 400.

Particular nucleic acid sequences may be detected using proteins or protein fragments (peptides) that specifically bind to the target sequence, such as for instance transcription factors or antibodies or antibody fragments that are specific for particular DNA sequences. For instance, auto-immune anti-DNA antibodies may be used.

In the case of peptides and proteins, all peptidic ligand/peptidic receptor tandem molecules that are known in the art are enclosed in the scope of the present invention. Such peptidic ligand/peptidic receptor tandem molecules includes peptidic antigens/antibodies or antibody fragments, as well as any hormone/hormone receptor, cytokine/cytokine receptor tandem, chemokine/chemokine receptor, aptamer/peptide, aptamer/protein. Membrane sugars that are implicated in cell migration and their proteic receptors are also in the scope of the present invention.

Also, antigens of any kind (such as nucleic acids, haptens, peptides or proteins) and their specific antibodies are included in the tandem target/binding and binding/target molecules according to the invention. In particular, the use of an antibody or antibody fragment as binding molecule in a conjugate allows for a novel kind of immunocytochemistry in tissue sections using MALDI-MS analysis as revelation technology thanks to the antibody photocleavably linked tag molecule.

Organic compounds may also be mapped using methods according to the invention. In particular, the in vivo distribution of administered organic drugs may be monitored using methods according to the invention.

In a particular embodiment of the above described method according to the invention, each target molecule is independently chosen in the group constituted of nucleic acids, in particular mRNA molecules, peptides, proteins, in particular receptors and ligands, antibodies, antigens, haptens, and organic compounds. In a preferred

embodiment, at least one target molecule is a mRNA molecule. In another preferred embodiment, at least one target molecule is selected from a peptide, a protein, an antigen or a hapten.

In a particular embodiment of any above described method according to the invention, each binding B molecule that binds specifically to a target molecule is independently chosen in the group constituted of nucleic acids, particularly oligonucleotides, peptides, proteins, in particular receptors and ligands, antibodies, antigens, haptens, aptamers and organic compounds. In a preferred embodiment, when at least a target molecule is a mRNA molecule, each B molecule that binds specifically to a target nucleic acid is a nucleic acid probe with a sequence that is complementary to said target mRNA sequence. In another preferred embodiment, when at least a target molecule is a peptide, protein, antigen or hapten, each B molecule that binds specifically to a target peptide, protein, antigen or hapten is an antibody directed against said peptide, protein, antigen or hapten. In particular, when an antibody or antibody fragment is used as binding molecule B, the target molecule against which said antibody or fragment thereof is specifically directed may be the frame-work region of a particular antibody subclass, such as IgM, IgD, IgG, or IgE. Alternatively, when an antibody or antibody fragment is used as binding molecule B, the target molecule against which said antibody or fragment thereof is specifically directed may be the frame-work region of antibodies generated in a particular species, such as for instance rabbit, mouse, rat, goat, hamster, sheep and human. In these two cases, a 2 steps indirect detection may be performed using a primary non modified antibody of a specific subclass directed against any desirable target molecule, said primary antibody being then recognized by an antibody-tag conjugate according to the invention (see Example 2 and Figure 3). Such a conjugate is highly advantageous since it permits to use it with any primary antibody, of any antigenic specificity, provided that said primary antibody belongs to the subclass recognized by the antibody of the conjugate.

According to the invention, a "tag molecule" refers to a molecule of known molecular weight that is detectable by mass spectrometry using MALDI (UV/IR), SIMS and DESI. A MALDI mass spectrometer usually allows for the detection of compounds in a wide range of m/z ratios. For instance, MALDI-Time of Flight (TOF) analyzers can

detect compounds of m/z ratios up to 1 000 000. A suitable tag molecule also needs not to interfere with the specific binding of the target and the binding molecules. Preferably, a suitable tag molecule thus has limited steric volume to avoid steric hindrance of the binding of the target and binding molecules. For both above described reasons, suitable tag molecules thus preferably have a m/z ratio inferior to 10 000. Particular examples of suitable tag molecules include peptides, nucleic acids, sugars, polymers, lipids, and organic compounds. Among these, labelling molecules that are used to reveal a complex formation between a target molecule and a binding molecule, may also be used as tag molecules, including fluorochromes such as for instance usual fluorescein isothyocanate (FITC), R-phycoerythrin (PE), Cy3, Cy5, Cy7, PerCP, Allophycocyanin (APC), Texas Red, TRITC, PE-Cy5 conjugates, PE-Cy7 conjugates, or APC-Cy7 conjugates; enzymes such as alkaline phosphatase or peroxydase; biotin; gold; or all MALDI matrices or even all compounds that can be analyzed in pure laser desorption mode and for example pre-ionized molecules. When a mass spectrometry technology suitable for the detection of very small molecules is used, atoms may also be used as tag molecule.

All tag molecules described above may be used in the invention, and in any case, depending on the mass spectrometry technology used, a person even moderately skilled in the art of mass spectrometry analysis will know what type of tag molecule to choose to optimize the detection step. For instance, when MALDI mass spectrometry is used, tag molecules preferably have a m/z ratio <5000. Alternatively, when SIMS is used, tag molecules preferably have a m/z ratio <500, and when DESI is used, tag molecules preferably have a m/z ratio <5000.

In a particular embodiment of any above described method according to the invention, each A tag molecule is chosen in the group constituted of peptides, nucleic acids, sugars, and organic compounds. In a preferred embodiment, at least one A tag molecule is a peptide.

In a preferred embodiment of an above described method according to the invention, target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, and A tag molecule(s) are peptides. In a preferred embodiment of an above described method

according to the invention, target molecule(s) are peptides, proteins (including antibodies), antigens, or haptens, binding molecules are antibodies or antibody fragments and tag molecules are peptides. In another preferred embodiment of an above described method according to the invention, target molecule(s) are peptides, proteins (including antibodies), antigens, or haptens, binding molecules are antibodies or antibody fragments and tag molecules are usual antibody labelling molecules such as fluorochromes, including for instance usual fluorescein isothyocanate (FITC), R-phycoerythrin (PE), Cy3, Cy5, Cy7, PerCP, allophycocyanin (APC), Texas Red, TRITC, PE-Cy5 conjugates, PE-Cy7 conjugates, or APC-Cy7 conjugates ; enzymes such as alkaline phosphatase or peroxydase; or biotin or gold; or all MALDI matrices or even all compounds that can be analyzed in pure laser desorption mode and for example pre-ionized molecules.

According to the invention, "hybridizing said tissue section with at least one $(A-X)_n$-B conjugate" refers to a reaction in which the tissue section and the conjugate(s) are brought into contact in such conditions that the binding B molecule of the conjugate is able to bind specifically to its target molecule in the tissue section. Depending on the nature of the binding and the target molecule, well-known hybridization protocols are available to the ones skilled in the art. Indeed, in the case when the binding molecule is a nucleic acid probe, in situ hybridization in tissue section has been well documented and is thus easily available to a man skilled in the art. Antibody staining in tissue sections is also a routine technology which is easily available to a man skilled in the art.

Using a method according to the invention, it is thus possible to obtain a map of at least one target molecule in a tissue section. By a target molecule "map" in a tissue section is meant a two dimensional representation of the expression of said target molecule in said tissue section. This two dimensional representation is obtained by scanning the tissue section surface with the MALDI analyzer at a defined spot density, performing MALDI analysis on each successive spot and storing both the obtained data and the coordinates of each spot. The higher the spot density, the more precise is the resulting map. The diameter of a MALDI laser is generally between 50-200 μm depending on the focalisation of the system, so that two adjacent irradiation spots are

preferably separated of the laser beam diameter (i.e. 50-200 μm). To allow for the acquisition of precise target molecule map, adjacent spots are preferably separated of at most 300 μm, at most 200 μm, more preferably at most 100 μm, at most 80 μm, at most 60 μm, at most 50 μm, at most 40 μm, most preferably of the diameter of the MALDI laser.

Each spot data is then analyzed in the molecular window of the tag molecule and the signal intensity of the tag molecule is reported at the spot coordinates. Such image reconstruction may be performed automatically using any suitable image reconstruction software known in the art or commercially available. Examples of suitable softwares are the IDL (Interactive Data Language) software, which is commercialized by RSI (RSI Corporate Headquarters. 4990 Pearl East Circle. Boulder, CO 80301), flexImaging (Bruker Daltonics, Bremmen, DE), MIT (M. Stoeckli, Novartis, Bâle, Switzerland).

In a method according to the invention for determining at least one target molecule map in a tissue section, several distinct target molecules can be mapped simultaneously. Indeed, it is sufficient to use several conjugates with distinct tag A molecules, thus displaying distinct molecular weights, to allow for the detection of several distinct target molecules. Using tag molecules with widely dispersed molecular weights, it is thus possible using any above described method according to the invention to map simultaneously the expression of many distinct target molecules in the same tissue section. In particular embodiments, at least 2, at least 3, at least 5, at least 8, at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 75, or at least 100 distinct target molecules may be mapped simultaneously. In particular, in the case of mRNA target molecules, at least 2, at least 3, at least 5, at least 8, at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 75, or at least 100 distinct target mRNA molecules may be mapped simultaneously in the same tissue section.

In a particular embodiment of any above described method in which target molecule(s) are mRNA molecule(s), said method may further comprise a step d) consisting in analyzing the obtained data in the molecular mass windows of each mRNA molecule corresponding protein to create each corresponding protein expression map of the tissue section.

In a first preferred embodiment, MALDI mass spectrometry is used and the X linker molecule is photocleavable at the wavelength of a MALDI laser. The invention then concerns a method for determining at least one target molecule map in a tissue section, comprising:

a) hybridizing said tissue section with at least one (A-X)n-B conjugate, wherein

- A is a tag molecule of known molecular weight,
- X is a linker that is photocleavable at the wavelength of a MALDI laser,
- n is an integer of at least 1,
- B is a binding molecule that binds specifically to said target molecule, and
- each distinct B molecule is linked to a distinct A tag molecule;

b) scanning the tissue section surface and analyzing each adjacent spot with a MALDI mass spectrometer, wherein the MALDI laser is used both to release the tag molecule A and to induce the sample ionization, and wherein the resulting data of each spot is saved; and

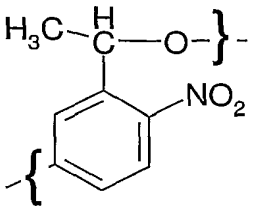
c) analyzing the obtained data in the molecular mass window(s) of each distinct tag molecule to create as many maps of the tissue section as the number of distinct studied target molecules.

According to the invention, a "linker that is photocleavable at the wavelength of a MALDI laser" refers to a chemical compound that separates two other chemical moieties, and that may be cleaved at least at one site under exposition to a MALDI laser. Most MALDI lasers usually have an ultraviolet (UV: inferior to 500 nm) wavelength, usually between 300 and 500 nm. For instance, many UV-MALDI analyzers have a pulsed nitrogen laser with a wavelength of 337 nm. Thus, a linker according to the invention is, in a particular embodiment, efficiently cleaved at least at one site under exposition to a wavelength of 250 to 500 nm, preferably a wavelength of 320 to 360 nm or 320 to 350 nm, more preferably a wavelength of 337 nm, so that the UV-MALDI laser acts both to cleave the linker and to ionize the sample. Other MALDI analyser display an infrared (IR: superior to 770 nm) laser. For instance, a Nd:YAG laser (wavelength = 1060 nm), Er:YAG laser (wavelength = 2940 nm), a mid-infrared

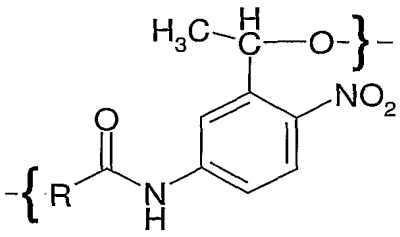
optical parametric oscillator (OPO) (wavelength = 2940 nm) or a TEA-CO_2 laser (wavelength = 10600 nm) may be used as IR-MALDI lasers. In other particular embodiments, a linker according to the invention is thus efficiently cleaved at least one site under exposition to a wavelength of 1000 to 1100 nm, preferably a wavelength of 1060 nm, or a wavelength of 2900 to 3000 nm, preferably a wavelength of 2940 nm, or a wavelength of 10500 to 10700 nm, preferably a wavelength of 10600 nm, so that respectively a Nd:YAG, a Er:YAG , or a TEA-CO_2 IR-MALDI laser acts both to cleave the linker and to ionize the sample.

In the case of UV-MALDI lasers, several linkers have been described (Olejnik et al. Nucleic Acids Res. 1999 Dec 1;27(23):4626-31; Bai X, et al. Nucleic Acids Res. 2004 Jan 26;32(2):535-41; Wenzel T et al. Nucleosides Nucleotides Nucleic Acids. 2003 May-Aug;22(5-8):1579-81), and several reagents useful to introduce photocleavable linkers in organic molecules are commercially available, notably for instance from AmberGen™ (1106 Commonwealth Avenue. Boston, MA 02215, USA), Link Technologies (3 Mallard Way, Strathcycle Business Park, Bellshill, Lanarkshire ML4 3BF, Scotland), Integrated DNA Technologies (1710 Commercial Park ° Coralville, IA 52241, USA), Glen Research (22825 Davis Drive, Sterling, Virginia, 20164, USA), Eurogentec (EUROGENTEC s.a. Headquarters. LIEGE Science Park. Rue Bois Saint Jean 5. 4102 SERAING. BELGIUM).

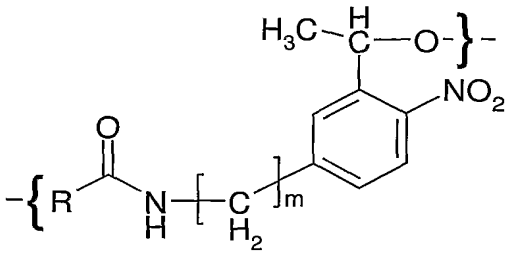
In a particular embodiment of any above described method according to the invention in which MALDI mass spectrometry and a linker molecule X that is photocleavable at the wavelength of a MALDI laser are used, the X linker molecule comprises a moiety chosen in the group constituted of:



,

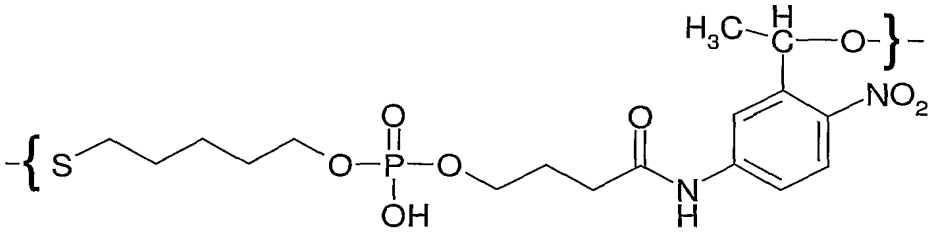


,



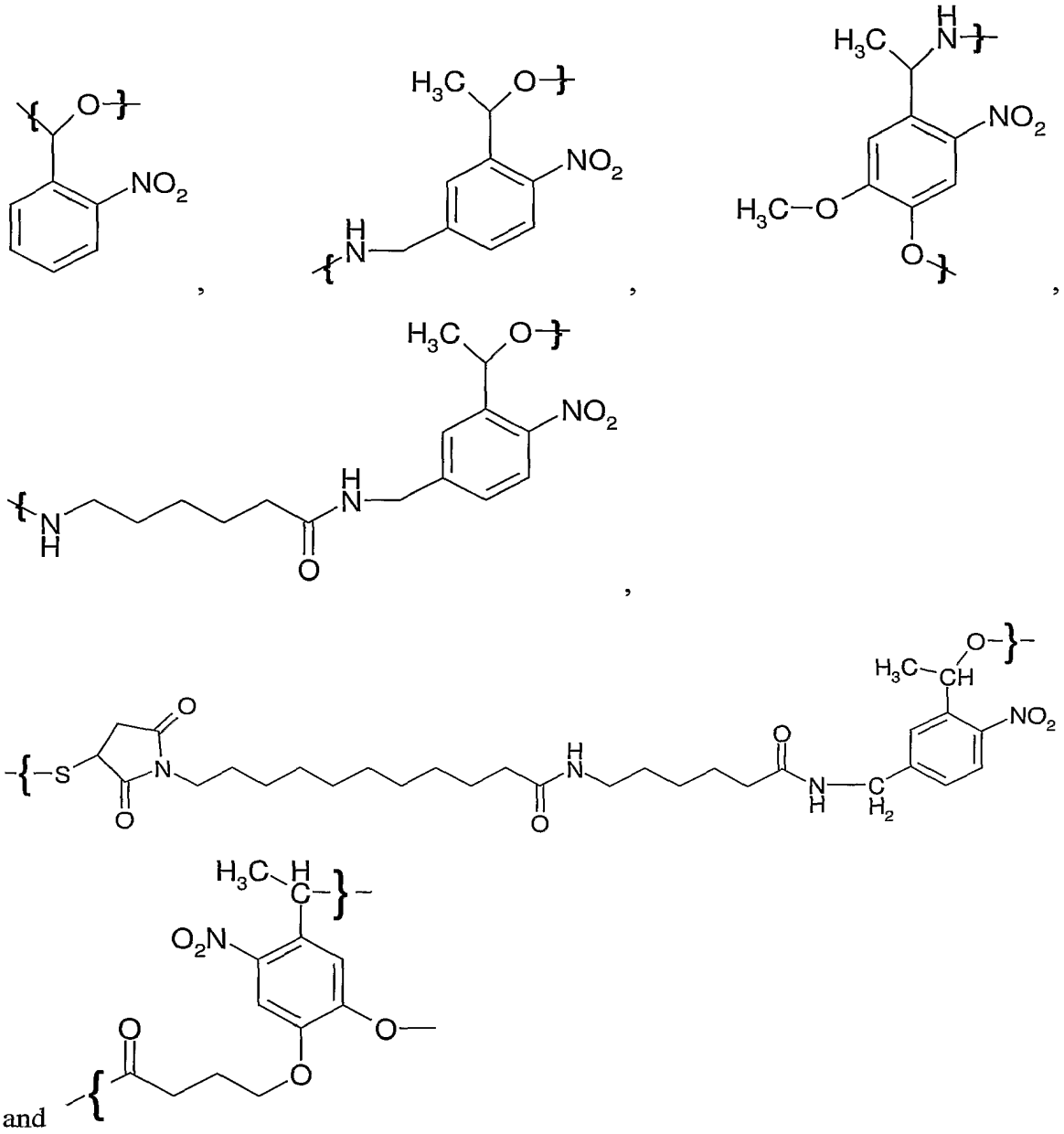
,

and



, wherein R is a C1-C6 alkyl group and m is an integer comprised between 1 and 4.

Alternatively, said linker molecule X that is photocleavable at the wavelength of a MALDI laser are used, the X linker molecule comprises a moiety chosen in the group constituted of:



and

Linker molecules X linking a tag molecule A and a binding molecule B may be obtained between A and B using commercially available linkers comprising such moieties. For instance, linkers suitable for preparing conjugates with photocleavable linkers for use in a method according to the invention may be obtained from Glen Research Corporation (22825 Davis Drive, Sterling, VA 20164, USA), including linkers named :

PC Linker (ref. 10-4920-02) ,

PC Spacer Phosphoramidite (ref. 10-4913-02)

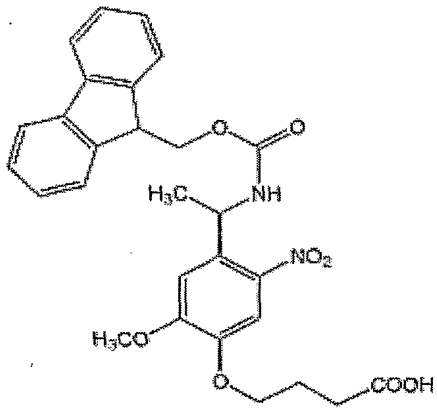
, and

PC Amino Modifier (ref. 10-4906-02)

,

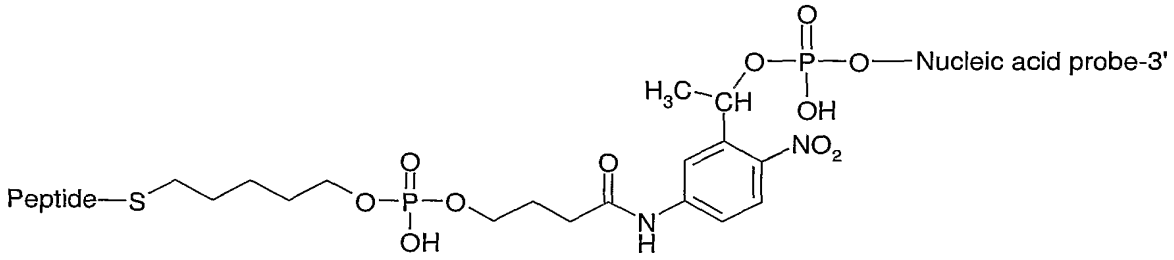
or from Pierce (Pierce Biotechnology, Inc. Customer Service Department P.O. Box 117 Rockford, IL. 61105 U.S.A), for instance linker Sulfo-KMUS (ref. #21111)

;

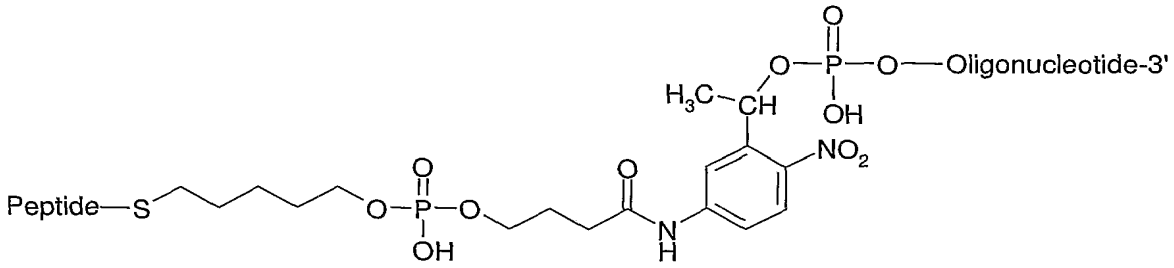
or from Merck Biosciences Ltd (Boulevard Industrial Park, Padge Road. Beeston, Nottingham NG9 2JR, United Kingdom) trade mark Novabiochem®, for instance linker  (ref. 01-60-0042);

or from Solulink under reference B1001, named Chromalink Biotin 354S, which is cleavable at a wavelength of 354 nm:

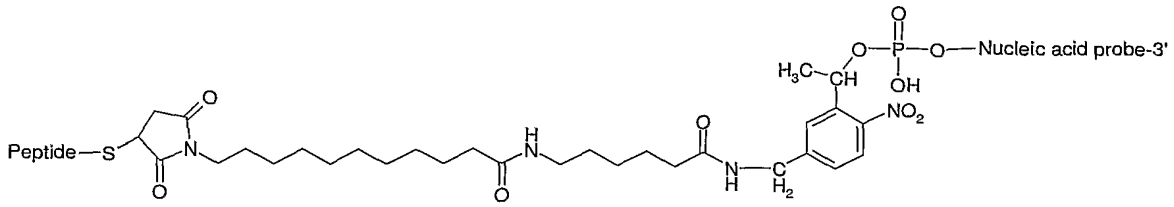
In a preferred embodiment where target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, and A tag molecule(s) are peptides, n is 1 is the $(A\text{-}X)_n\text{-}B$ conjugate(s) , and the resulting (A-X)-B conjugate(s) have the following structure:



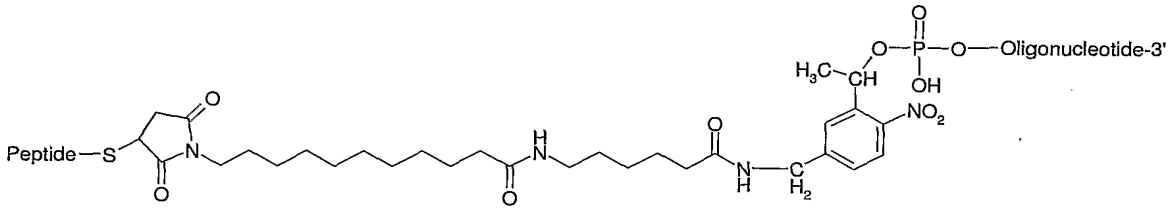
Such conjugates may for instance be obtained by amplification of a desired nucleic sequence using a forward (5') primer of the following structure:



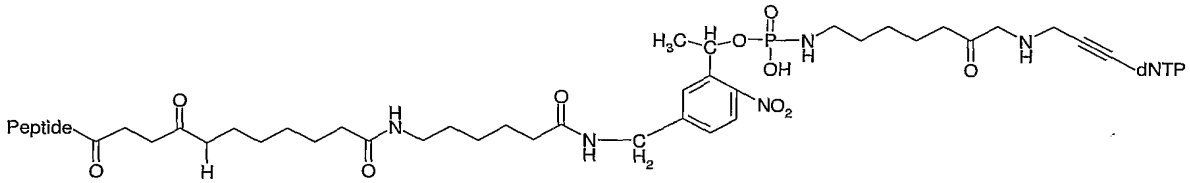
In another preferred embodiment where target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, and A tag molecule(s) are peptides, n is 1 is the $(A\text{-}X)_n\text{-}B$ conjugate(s) , and the resulting (A-X)-B conjugate(s) have the following structure:



Such conjugates may for instance be obtained by amplification of a desired nucleic sequence using a forward (5') primer of the following structure:



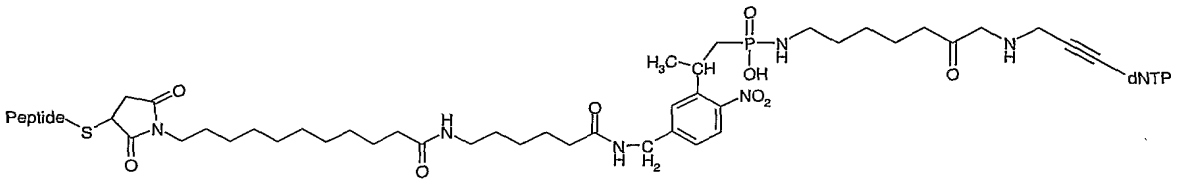
Alternatively, in another preferred embodiment where target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, and A tag molecule(s) are peptides, n is superior to 1 and the nucleic acid probe(s) comprise at least one modified base of the following structure:



,

wherein "dNTP" means any triphosphate deoxynucleotide, including dATP, dCTP, dGTP, dTTP, and dUTP. In a preferred embodiment, a modified dUTP base is used.

In another preferred embodiment where target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, and A tag molecule(s) are peptides, n is superior to 1 and the nucleic acid probe(s) comprise at least one modified base of the following structure:



,

wherein "dNTP" means any triphosphate deoxynucleotide, including dATP, dCTP, dGTP, dTTP, and dUTP. In a preferred embodiment, a modified dUTP base is used.

With the use of such a modified peptide-tagged dUTP, peptide-tagged nucleic acid hybridization probes can be very simply generated using a simple PCR amplification in the presence of dATP, dCTP, dGTP and the modified peptide-tagged dUTP. This way, a specific hybridization probe can be easily synthesized for any target mRNA sequence.

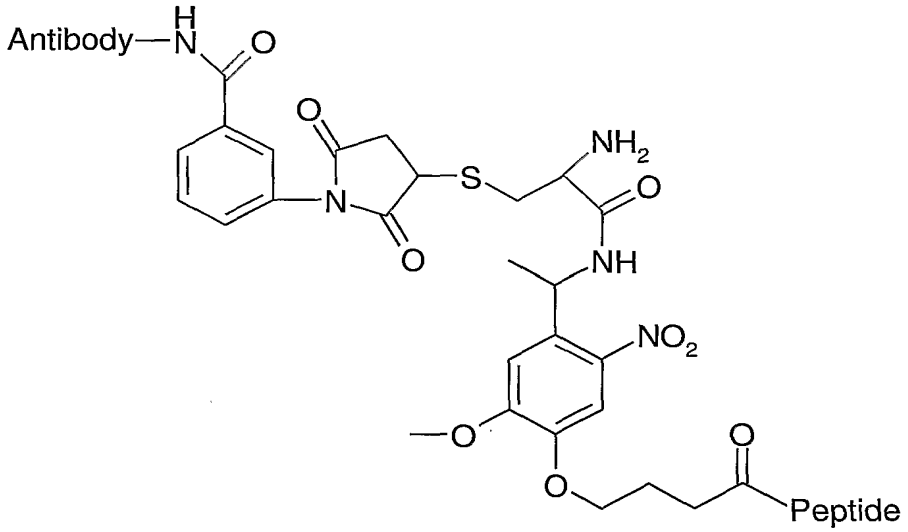
Moreover, it allows for a signal amplification since a given hybridization probe synthesized with the modified peptide-tagged dUTP will carry as many tag peptides as the number of U bases in its sequence.

Finally, the use of hybridization probes synthesized with the modified peptide-tagged dUTP allows for a quantitative analysis of mRNA expression in tissue sections. Indeed, the signal generated by the corresponding tag(s) of one or several studied mRNA(s) can be compared with that obtained for a reference mRNA sequence (for instance a house-keeping gene such as actin of HPRT). As the number of U bases in each hybridization probe is known, the expression ratio between each studied mRNA sequence and the reference mRNA sequence can be calculated.

In a method where target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, and A tag molecule(s) are peptides, using conjugates described above, several distinct target mRNA molecules can be mapped simultaneously. Indeed, it is sufficient to use several conjugates with distinct nucleic acid probes and distinct tag peptides displaying distinct molecular weights, to allow for the detection of several distinct target molecules. Using tag peptides with widely dispersed molecular weights, it is thus possible using any above described conjugate to map simultaneously the expression of many distinct target mRNA molecules in the same tissue section. In particular embodiments, at least 2, at least 3, at least 5, at least 8, at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 75, or at least 100 distinct target mRNA molecules may be mapped simultaneously.

In addition, stored spectrum data also display direct peptide/protein analysis of the tissue section, so that the tissue section analysis method of target mRNA molecules using the conjugates described above may further comprise a last step d) consisting in analyzing the obtained data in the molecular mass windows of each mRNA molecule corresponding protein to create each corresponding protein expression map of the tissue section.

In a preferred embodiment where target molecule(s) are peptide(s), protein(s) (including antibodies) or hapten(s), binding B molecule(s) are antibodies directed against said target molecules, A tag molecule(s) are peptides, n is 1 and the (A-X)-B conjugate(s) have the following structure:



In any method according to the invention in which MALDI mass spectrometry is used, the matrix that is used for MALDI-MS analysis may be any classical MALDI matrix (e.g. CHCA, 2,5- DHB, SA, HABA, 3-HPA...), or 2,4-DNPH.

By "matrix" is meant any material that, when mixed with the analyte, generates crystalline matrix-embedded analyte molecules that are successfully desorbed by laser irradiation and ionized from the solid phase crystals into the gaseous or vapour phase and accelerated as molecular ions. Commonly used MALDI-MS matrices are generally small, acidic chemicals absorbing at the laser wavelength, including nicotinic acid, cinnamic acid, 2,5-dihydroxybenzoic acid (2,5-DHB), α-cyano-4-hydroxycinnamic acid

(CHCA), 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid or SA), 3-methoxy-4-hydroxycinnamic acid (ferulic acid), 3,4-dihydroxycinnamic acid (caffeic acid), 2-(4-hydroxyphenylazo)benzoic acid (HABA), 3-hydroxy picolinic acid (HPA), 2,4,6-trihydroxy acetophenone (THAP) and 2-amino-4-methyl-5-nitropyridine. Protocols for the preparation of these matrices are well-known in the art, and most of these matrices are commercially available. Current commonly used matrices for peptide/protein analysis include α–cyano-4-hydroxycinnamic acid (CHCA), 2,5-dihydroxybenzoic acid (2,5-DHB) and sinapinic acid (SA). DNPH is 2,4-Dinitrophenylhydrazine and is used for aldehydes and ketones detection.

In a second preferred embodiment of the invention, a linker molecule X that is cleaved by fragmentation during sample desorption/ionization is used, and the analysis is then not restricted to MALDI mass spectrometry, but generally transposable to any mass spectrometry technology suitable for tissue section analysis. In particular, UV-MALDI, IR-MALDI, SIMS or DESI mass spectrometry may be used.

Indeed, indirect detection of compounds necessitates a release of the tag molecule at the very beginning of the desorption and ion production in the source of the instrument. When using MALDI mass spectrometry, which uses a laser for ion desorption/ionization via energy absorption by the matrix molecules, several ways to induce the tag molecule release can be thought of. The first, most elegant, approach is to use photodissociation of the tag molecule by the MALDI laser irradiation (see Figure 1 and Figure 3). This means a very specific release but also necessarily more complex structures and synthesis.

However, fast fragmentation can also be considered as a way of releasing the tag molecule. This phenomenon is generally present in mass spectrometry technologies, and its use is thus compatible with MALDI and with other mass spectrometry sources. Notably, it can be taken advantage of this fragmentation phenomenon with other mass spectrometry technologies such as SIMS (Secondary Ion Mass Spectrometry), where the source is a primary ion beam, or with a recently described technology named DESI (Desorption Electrospray Ionisation). Moreover, as for MALDI/MS, these two technologies can be used for the direct analysis of tissue sections (Touboul, D.; Kolmer,

F.; Niehis, E.; Brunelle, A.; Laprevote, O.. Journal of the American Society for Mass Spectrometry, (JASMS) 2005, 15, 1608-1618) (Cooks, R.G.M.; Ouyang, Z. Takats, Z.; Wiseman, J.M.; Science, 17 March 2006, Vol 311, 1566-1570).

Thus, the invention concept of using conjugates of a specific binding molecules linked to a tag molecule via a linker molecule can be transposed to other mass spectrometry technologies than MALDI, such as SIMS or DESI, provided that a linker that is cleaved by fast fragmentation is used, resulting in a fast fragmentation release of the tag molecule. Considering this phenomenon, it is possible to design other conjugates without more simple linkers, since they do not need to be photocleavable but only to be cleaved by fast fragmentation, which can be easily obtained. In fact, all sorts of linkers may be cleaved by fragmentation, and the capacity of any linker may be easily tested

In addition, linkers that were described above as photocleavable at the wavelength of a MALDI laser were found by the inventor to be cleaved also by fast fragmentation. Alternatively, such linkers may thus be used in conjugates for any method according to the invention described above, including when other mass spectrometry technologies than MALDI are used, notably for SIMS or DESI mass spectrometry.

However, it must be clearly understood that the (A-X)n-B conjugates usable by taking advantage of the fast fragmentation phenomenon are not limited to the particular suitable linkers described above that may be cleaved by photodissociation and fragmentation. Indeed, the capacity to be cleaved by fast fragmentation of any linker that would permit an easy linkage between a A tag molecule and a B binding molecule is easily tested by simply analyzing the synthesized conjugate in solution and monitoring the presence of a cleaved A tag molecule. This way, any potential linker can be easily tested and approved or refused for use in the present invention. In addition, any person skilled in the art of mass spectrometry will know, for the great majority of potential linkers, if their chemical structure will or not permit there cleavage by fragmentation, so that the easy to perform test described above will even not be necessary in most cases.

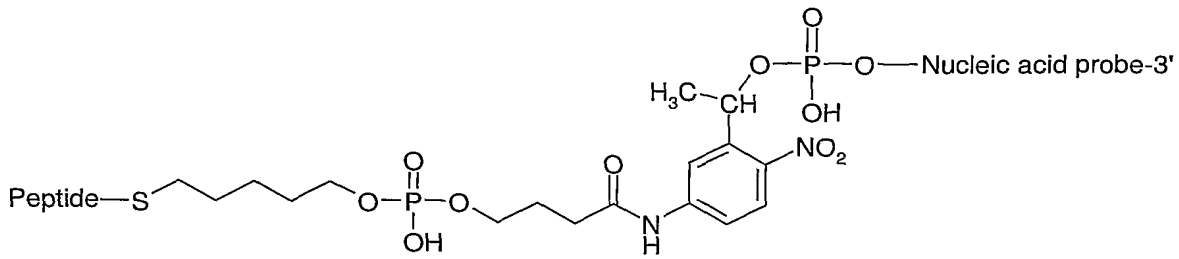
In particular, conventional, and even commercial, labelled antibodies may be used as conjugates when fast fragmentation is used to cleave the linker. For instance, fluorescent conjugated antibodies may be used, the fluorescent molecule being used as

tag molecule. Notably, FITC-conjugated, PE-conjugated, PerCP-conjugated, APC-conjugated, Cy3-conjugated, Cy5-conjugated, Cy7-conjugated, Texas red-conjugated, TRITC-conjugated, PE-Cy5-conjugated, PE-Cy7-conjugated,or APC-Cy7-conjugated antibodies may be used. Alternatively, antibodies conjugated to other molecules such as enzymes alkaline phosphatase or peroxydase may also be used, the enzyme being used as tag molecule. This possibility has the crucial advantage that all sorts of antibodies, with a very wide range of antigenic specificities, are commercially available as labelled antibodies. In addition, technologies to attach a labelling molecule via conventional linkers to any antibody are well-known routine technologies. This way, indirect mass spectrometry analysis of biomolecules present in tissue sections can be performed for a very large number of distinct biomolecules using the methods according to the invention.

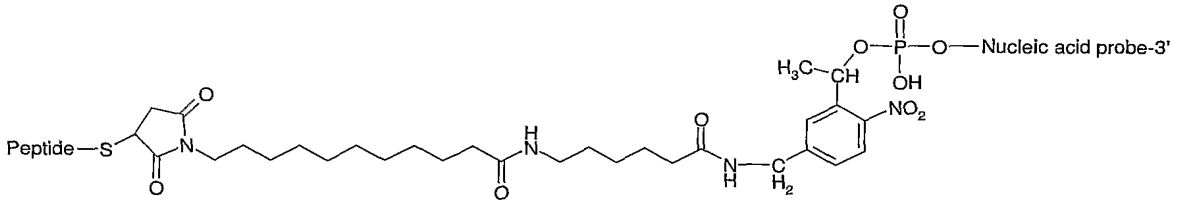
Thus, in a preferred embodiment, when fast fragmentation is used, target molecule(s) are peptides, proteins (including antibodies), antigens, or haptens, binding molecules are antibodies or antibody fragments, n is 1 and tag molecules are fluorochromes, enzymes, biotin or gold. Atoms surch as Br may also be conjugated to antibodies, in particular when SIMS mass spectrometry is used (see Example 4). In this case, a linker named EDAC commercially available from Sigma under referenceE1769 may be used to conjugate the antibody to the Br atom (see Example 4).

In addition, when such a linker molecule X that is cleaved by fast fragmentation is used, potentially any mass spectrometry technology, provided that it is suitable for tissue sections analysis (which is immediately apparent for any skilled person), may be used to implement the method according to the invention. In particular, several additional mass spectrometry technologies than UV-MALDI or IR-MALDI may be used, such as SIMS or DESI mass spectrometry. Thus, in a preferred embodiment of a method in which a linker molecule X that is cleaved by fast fragmentation is used, UV-MALDI, IR-MALDI, SIMS or DESI mass spectrometry is used

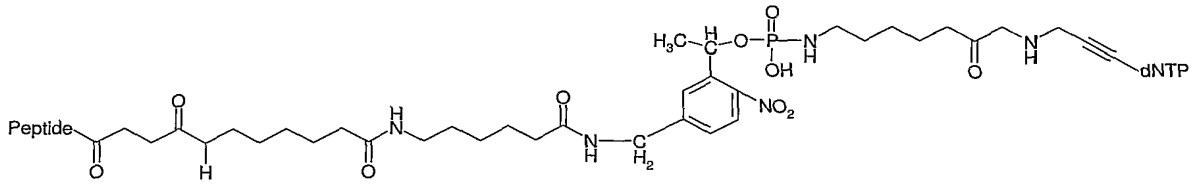
The invention further concerns a conjugate suitable for use in any method according to the invention, wherein said conjugate is selected from:


Peptide—S
O—P—O
OH
O
O
N
H
H3C—CH
O—P—O—Nucleic acid probe-3'
OH
NO2

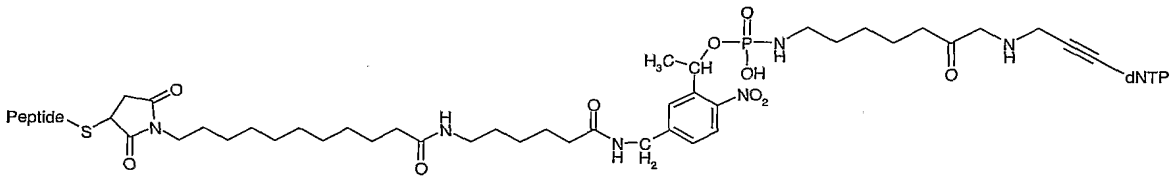
,

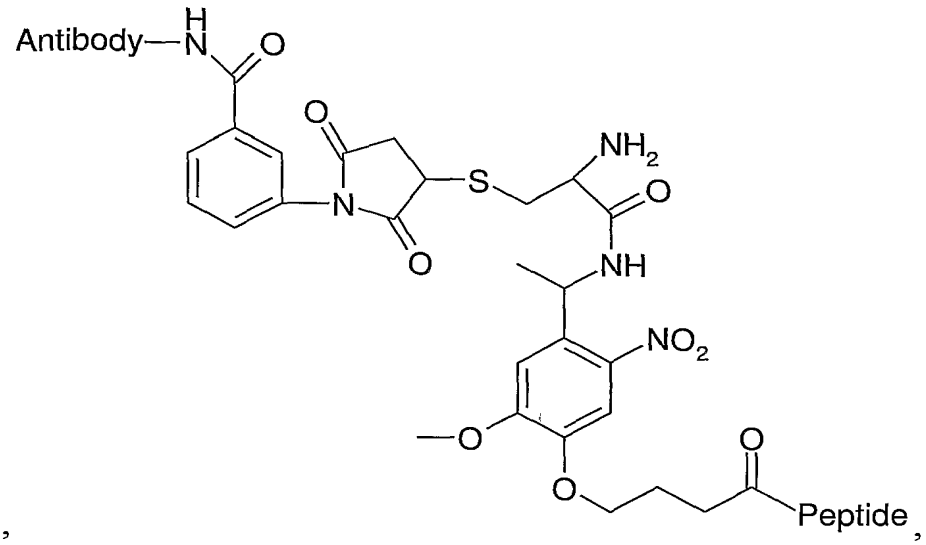

Peptide—S
O
N
O
H
N
O
N
H
C
H2
H3C—CH
O—P—O—Nucleic acid probe-3'
OH
NO2

,

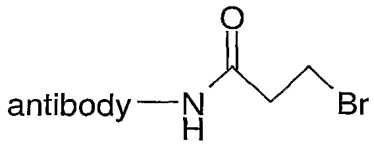

Peptide
O
H
N
H
H3C—C—O—P—N
H
OH
O
N
H
dNTP
NO2

,


Peptide
S
O
N
O
H
N
O
N
H
C
H2
O—P—N
H
OH
O
N
H
dNTP
NO2



, or



Each of the above conjugates is individually preferred as a conjugate according to the invention

Having generally described this invention, a further understanding of characteristics and advantages of the invention can be obtained by reference to certain specific examples and figures which are provided herein for purposes of illustration only and are not intended to be limiting unless otherwise specified.

DESCRIPTION OF THE DRAWINGS

Figure 1. Schematic principle of mRNA specific multiplex imaging by MALDI combining specifically designed labeled probes for indirect detection by tag molecule release during MALDI desorption/ionization step and ISH technology.

Figure 2. General scheme of possible specific imaging of different targets (peptides, proteins, mRNA, sugars, drugs) using different labeled probes (oligonucleotides, antibodies, lectins, aptamers).

Figure 3. Schematic principle of peptides/proteins multiplex specific imaging by MALDI combining specifically designed labeled probes for indirect detection by tag released during MALDI desorption/ionization step and immunocytochemistry technique.

Figure 4. Detection of the peptide moiety of **A.** oligo-peptide 3 in linear positive mode. **B.** 100 fmol of oligo-peptide 1 in linear positive mode **C.** oligo-peptides 1, 2 and 3 simultaneously in reflector positive mode, and **D.** oligo-peptide 3 on a MALDI support plate made of gold.

Figure 5. Mapping of oligo-peptide 3 on a solid support. Oligo-peptide 3 was deposited with a multipipette on a solid support to draw the image of a X letter. The presence of oligo-peptide 3 was then mapped by MALDI-MS analysis and the image was reconstructed using IDL software. The resulting image is displayed

Figure 6. Synthesis of a peptide-tagged proenkephalin in situ hybridization probe **A.** PCR amplification products of proenkephalin obtained using either tagged (peptide-photocleavable linker-oligonucleotide) or conventional (without the photocleavable linker and the peptide moiety) forward (5'-CAG-GAC-TCC-CCA-AAG-GAG-AAC-AGG-A-3') and reward (5'-GA-CGT-ACC-AGG-CGG-TAG-CTG-CAT-TT-3') oligonucleotides. Line 1: conventional forward and reward primers. Line 2: tagged forward and conventional reward primers. Line 3: conventional forward and tagged reward primers. Line 4: tagged forward and tagged reward primers. Line 5: negative control: water control. Line 6: molecular weight markers. **B.** Corresponding amplification products after purification on silica columns. Line 1: conventional forward and reward primers. Line 2: tagged forward and conventional reward primers. Line 3: conventional forward and tagged reward primers. Line 4: tagged forward and tagged reward primers. Line 5: negative control: water control. Line 6: molecular weight markers. **C. and D.** MALDI analysis of purified PCR amplification products with tagged forward and reward primers (**C.**) or tagged forward and conventional reward primers (**D.**).

Figure 7. Direct MALDI analysis of a deparaffined rat brain section after hybridization of a peptide-tagged proenkephalin *in situ* hybridization probe. **A.** First tissue section localisation. **B.** Second distinct tissue section localisation. **C.** Control analysis with a non peptide-tagged proenkephalin *in situ* hybridization probe.

Figure 8. Synthesis and MALDI analysis of a nucleic probe incorporating a modified, peptide-tagged dUTP. **A.** Structure of the modified, peptide-tagged dUTP and scheme of the RT-PCR synthesis of a proenkephalin nucleic probe (400 bp) incorporating the modified, peptide-tagged dUTP. **B.** MALDI analysis of the modified, peptide-tagged dUTP. **C.** MALDI analysis of the proenkephalin nucleic probe incorporating the modified, peptide-tagged dUTP. **D.** MALDI analysis of a proenkephalin nucleic probe incorporating a normal non modified dUTP, without peptide tag.

Figure 9. Reactions for the synthesis of an antibody-peptide conjugate with a photocleavable linker X.

Figure 10. Structure of an antibody-peptide conjugate with a photocleavable linker X. The 2 major fragments detected by MALDI mass spectrometry (1st fragment obtained by photodissociation of the Tag via cleavage of the photocleavable linker, and 2nd fragment obtained by fast fragmentation of the molecule) are shown.

Figure 11. Typical mass spectrum obtained during in solution MALDI analysis of an antibody-peptide conjugate with a photocleavable linker X (matrix sinapinic acid SA).

Figure 12. Typical mass spectrum obtained during direct analysis of a rat brain tissue section using no conjugate (**A**, rat brain control) versus using an antibody-peptide conjugate with a photocleavable linker (**B**).

Figure 13. Expression map of membrane protein carboxypeptidase D (180 kDa) detected with a rabbit primary antibody directed against this protein and the antibody-peptide conjugate with a photocleavable linker. The image has been reconstructed from the detection signal corresponding to the peptide tag.

Figure 14. Typical mass spectrum obtained during in solution MALDI analysis of a FITC labelled antibody using CHCA matrix. **A.** without FITC labeled antibody (CHCA matrix alone). **B.** with FITC labeled antibody.

EXAMPLE 1

Use of oligo-peptide conjugates with a photocleavable linker for the indirect detection of mRNA is tissue sections by MALDI-MS

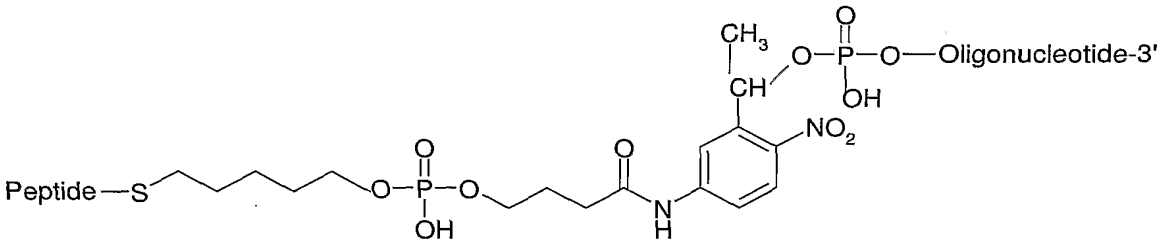
Oligo-peptide conjugates with a photocleavable linker were tested for their capacity to allow indirect detection of specific mRNA in tissue sections by MALDI-MS.

1.1 Materials and Methods

1.1.1 Oligo-peptide conjugates

Structure

Oligo-peptide conjugates that were studied displayed the following structure:



Five oligo-peptide conjugates with distinct peptide moieties were synthesized. The peptide amino acid sequence, monoisotopic (M_{mono}) and average (M_{avg}) molecular weights, and the nucleic sequence of the oligonucleotide of each conjugate are displayed in the following Table 2.

Table 2. Features of the studied oligo-peptide conjugates

Oligo-peptide	Peptide sequence (N to C-terminal)	M_{mono} (u.m.a)	M_{avg} (u.m.a)	Oligonucleotide sequence (5'-3')
1	GRALGVFVC (SEQ ID NO :1)	918,50	919,14	CACGTACAGGATGTACAG (SEQ ID NO :6)
2	RPKPQQFFGLMC (SEQ ID NO :2)	1448,73	1449,79	TCGAGAGGTACATCGTG (SEQ ID NO :7)
3	RPPGFSPFRC (SEQ ID NO :3)	1160,58	1161,37	AAGCGGTACGAGTAGCA (SEQ ID NO :8)
Forward	RPPGFSPFRC (SEQ ID NO :4)	1718,00	1719,00	GACGTACCAGGCGGTAGCTGCATTT (SEQ ID NO :9)
Reverse	RPPGFSPFRC (SEQ ID NO :5)	1718,00	1719,00	CAGGACTCCCCAAAGGAGAACAGGA (SEQ ID NO :10)

Synthesis

Oligo-peptide conjugates were synthesized using the following protocol:

The peptide is synthesized on Symphony (Protein Technologies Inc) and purified on a Delta-Pak C18 15µm 100A column (Waters).

The oligonucleotide is synthesized from 3' to 5' on Expedite (ABI). The amine function with photocleavable linker is added in 5' before cleavage and deprotection. These steps are performed using a NH_4OH 28% solution during 24 hours in the dark. The amino oligonucleotide is then purified on a Delta-Pak C18 15µm 300A column (Waters). The amino function of the oligonucleotide is coupled to a heterobifunctional reagent comprising a maleimide function. The maleimido oligonucleotide is solubilized in water and added to an 1.2 equivalent of peptide in solution. The mixture is let under stirring for 16 hours.

The oligo-peptide conjugate is then purified on a Delta-Pak C18 15µm 300A column (Waters) and characterized by mass spectrometry.

1.1.2 Preparation of samples before MALDI-MS

Several commonly used matrices were indifferently used for MALDI-MS analysis : α-cyano-4-hydroxycinnamic acid (CHCA), 3-hydroxy picolinic acid (HPA), and sinapinic acid (SA).

For CHCA, 10mg of the matrix was dissolved in 1mL of acetonitrile/water (2:1, v/v, 0.1% TFA/H_2O). For SA, 20mg of matrix was dissolved in the same solvent.

For SA, 20 mg of matrix was dissolved in the same solvent.

For HPA, 50 mg/ml of the matrix was dissolved in water.

In some cases, an additive (ammonium citrate or acetate) was added to the matrix.

Before MALDI-MS analysis, 1µL of sample solution and 1µL of matrix solution were mixed on the MALDI plate according to the procedure of the dried-droplet preparation (Karas, M.; Hillenkamp, F.; *Anal.Chem.* **1998**, *60*, 2299-2301).

Three distinct MALDI plate materials have been tested: stainless steel, gold and Teflon.

1.1.3 MALDI-MS analysis

MALDI-TOF mass spectra were performed on a Voyager-DE STR mass spectrometer (Applied Biosystems, Framingham, MA, USA) with delayed extraction (DE) and operating with a pulsed nitrogen laser emitting at 337nm.

Analysis in linear mode

For inox MALDI plates, acquisition parameters were set to: acceleration voltage: 20 kV ; 1st grid voltage: 94% ; guide-wire voltage: 0.05% ; extraction delay time: 100-250 ns.

For gold MALDI plates, acquisition parameters were set to: acceleration voltage: 25 kV ; 1st grid voltage: 96% ; guide-wire voltage: 0.05% ; extraction delay time: 600 ns.

Analysis in reflector mode:

Acceleration voltage: 20 kV, 1st grid voltage: 70%, guide-wire voltage: 0.05%, extraction delay time: 200ns.

1.1.4 Software for image reconstruction

For image reconstruction, software flexImaging (Bruker daltonics, Bremmen, DE) was used.

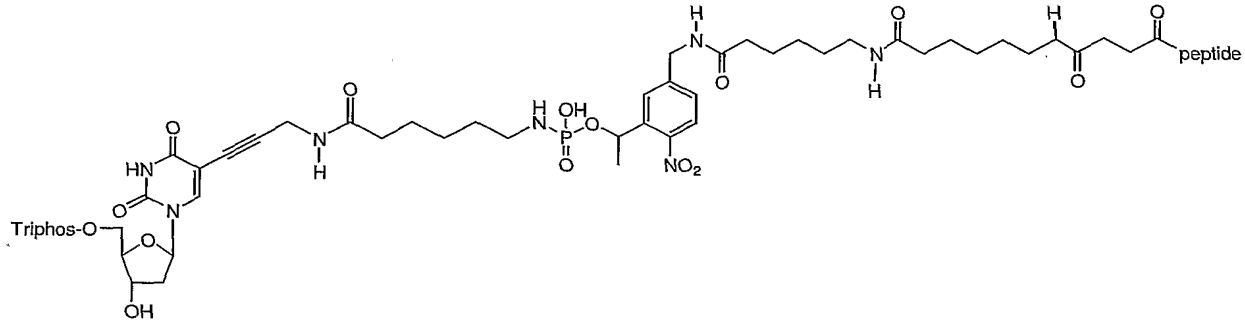
1.1.5 PCR amplifications using oligo-peptide conjugates

PCR amplifications of proenkephalin were obtained using either tagged (peptide-photocleavable linker-oligonucleotide) or conventional (without the photocleavable linker and the peptide moiety) forward (5'-CAG-GAC-TCC-CCA-AAG-GAG-AAC-AGG-A-3', SEQ ID NO:9, see Table 2) and reward (5'-GA-CGT-ACC-AGG-CGG-TAG-CTG-CAT-TT-3', SEQ ID NO:10, see Table 2) primers.

Amplification products were then analyzed by electrophoresis in 2% agarose gel.

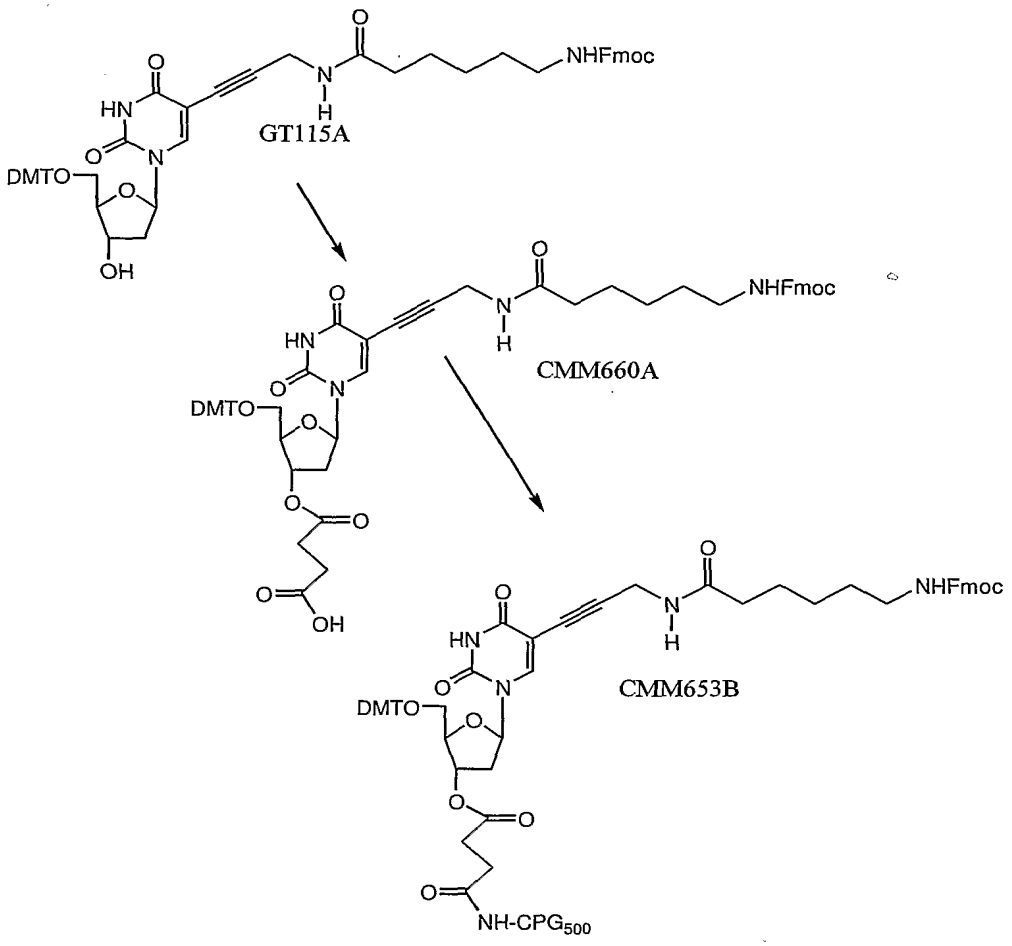
1.1.6 Synthesis of a dUTP-peptide conjugate with a photocleavable linker

A modified peptide-tagged dUTP with the following formula was synthesized:



For the synthesis, the following synthesis scheme was followed:

Scheme 1. Triphosphate from Fmoc-aminopropargylcaproyldU



In order to prepare this triphosphate a Fmoc protected CPG resin was required. Since none of the precursor was available, then the succinylate was prepared from GT115A (100mg). The sample was relatively pure but contained a small amount (by TLC) of a higher running non-tritylated compound (originates from the Sonogashira reaction and does not interfere with subsequent reactions and was not visible in the nmr spectra of the sample). Since it was not possible to purify the succinate, the reaction was modified slightly. It is normal to add 2 equivalents of succinic anhydride to the reaction to get quantitative yield but if this is not removed completely the amino residues of the cpg resin can become blocked during functionalisation. Therefore, 1.5 equivalents were used since the exact purity of the product is undetermined. The reaction did not go to

completion (from TLC this was more than 50% by comparing the intensity of the components on the TLC by UV(254nm) and the intensity of the DMT cation on treatment with HCl fumes. Since the non-succinylated product will not react, the resin was functionalised using this mixture. The resin was prepared but the loading is very low, 5.4μmolg^{-1} (180mg).

The resin was detritylated using 2% TCA/DCM washed with DCM and the process repeated until no orange colour due to the DMT cation was observed.

This was then dried (suction under argon) and the resin soaked in pyr/DMF 1:3 (0.4ml) for 5 minutes before a solution of 0.1M Eckstein's reagent in dioxane was added (0.1ml). The reaction was allowed to stand for 15 minutes after which time the resin was washed (dioxane, MeCN) and dried (suction under argon).

The resin was then soaked in a solution on 0.5M *bis*-(tributylammonium) pyrophosphate in anhydrous DMF and tri-*n*-butylamine for 20 minutes and the resin washed (DMF, MeCN) and dried (suction under argon).

The product was oxidised (iodine/water/pyridine/THF for 30 minutes), washed (MeCN) dried (suction under argon).

The Fmoc protecting group was removed (20% piperidine/DMF, 0.5ml, 20mins) and the resin washed thoroughly, (DMF, MeCN) and dried (suction under argon). This was then washed with DCI and a solution of DCI/photolabile amino linker CEP (1:1, 0.5ml) was added and the reaction allowed to stand for 20 minutes. The solution was removed and the resin washed (MeCN) and dried (suction under argon). A mixture of cap A/cap B (1:1, 0.5ml) was added and the resin soaked for 5 minutes before removing the capping reagents and washing and drying the resin as before. The product was oxidised (I_2/THF/pyr/H_2O, 5mins) and the resin washed and dried as before. This was cleaved from the resin with cNH_4OH at room temperature for 30 mins, then purified by anion exchange HPLC on a Dionex NucleoPac100 HPLC column using the following solvent system Buffer A:0.1M NH_4Cl with 10% acetonitrile; Buffer B: 1M NH_4Cl with 10% acetonitrile; flow rate 2.5 mL/min. using 6Triphos.mth. This gave 3 fractions (A:-7mins, B:-7.9mins and C:-10.3mins). All 3 fractions were lyophilized over night before being desalted by reverse phase HPLC Buffer A:Water; Buffer B: acetonitrile; flow rate 4 mL/min. The 3 fractions were again lyophilized overnight before being suspended in

200ul of water. M.S. showed that CMM661A pk 1 was definitely not the triphosphate but it could be either CMM661pk 2 or 3 (very similar M.S. profiles). (CMM662A was formed from CMM661A pk 2 and CMM663A was formed from CMM661A pk 3).

Both samples were then used in the subsequent reaction. Bicarbonate buffer (10ul) and the maleimide NHS ester (50ul) were added to each sample and the reactions agitated overnight. The samples were diluted with milliQ water (500ul) and filtered. The samples were purified by RP-HPLC, buffer A:0.1M TEAA, buffer B:MeCN, flow rate 4mL/min. using MeCN50.mth and the coupling of the peptide was carried out on these fractions.

1.1.7 Synthesis of a proenkephalin nucleic probe (400 bp) incorporating the modified, peptide-tagged dUTP

A proenkephalin nucleic probe (400 bp) incorporating the previously synthesized modified, peptide-tagged dUTP (see paragraph 1.1.6) was synthesized by RT-PCR using the conventional forward and reward primers described in paragraph 1.1.5.

1.2 Results

1.2.1 Detection of one or multiple oligo-peptide conjugates in various modes of analysis

The three oligo-peptides conjugates were analyzed in various analysis modes, either separately or simultaneously (Figure 4).

The analysis of separate oligo-peptide conjugates in linear positive mode (Figure 4A), linear negative mode and reflector positive mode showed oligo-peptide conjugates are readily detectable in various MALDI analysis modes.

Sensitivity tests were also performed to assess the detection threshold in the various MALDI analysis modes. The observed thresholds were of 100 fmol with a signal/noise ratio of 8 in linear positive mode (Figure 4B), and 1 pmol with a signal/noise ratio of 5 in reflector positive mode.

Oligo-peptides 1, 2 and 3 were further analyzed simultaneously in the linear positive mode, linear negative mode and reflector positive mode (Figure 4C). In Figure 4C, for each oligo-peptide conjugate, two peaks of distinct m/z ratios are observed. The

higher m/z ratio peaks (m/z 1346.95, 1603.74 and 1891.72 respectively for peptides 1, 3 and 2) correspond to the photocleaved peptides at the expected cleavage site (containing the photocleavable linker). The lower m/z ratio peaks (m/z 1161.87 et 1449.88 respectively for peptides peptides 3 and 2) appear to correspond to a subsequent cleavage in gaseous phase leading to the formation of a protonated ion of peptide only (without the photocleavable linker).

In any case, these results show that three distinct oligo-peptides can be easily detected simultaneously.

The influence of the MALDI plate material on resolution, sensitivity and signal/noise ratio was investigated using stainless steel, gold and Teflon MALDI plates, with or without an additive composed of ammonium citrate or acetate.

Results show that cleavage of the photocleavable linker is possible no matter which material is used, with or without coating with ammonium citrate or acetate.

Moreover, it appears that the signal/noise ratio and signals intensity are higher on gold MALDI plates (Figure 4D).

1.2.2 Mapping of an oligo-peptide conjugate on a solid support

To evaluate the possibility to map an oligo-peptide conjugate on a surface, the image of an X letter made of oligo-peptide 3 was drawn using a multipipette on a MALDI plate.

The corresponding area was then analyzed, spot by spot, by MALDI-MS and the repartition of the ion m/z 1626.37 corresponding to the photocleaved fragment of oligo-peptide 3 under its sodium cationized form MNa^+ was reconstructed using image reconstruction software.

Figure 5 shows the obtained reconstructed image, demonstrating the possibility to obtain a simple image from direct cleavage.

1.2.3 Synthesis of a peptide-tagged proenkephalin in situ hybridization probe using oligo-peptide conjugates as primers

To confirm the possibility to synthesize larger hybridization probes using an oligo-peptide conjugate as forward and/or reward primer, PCR amplifications were carried out using either tagged (peptide-photocleavable linker-oligonucleotide) or conventional (without the photocleavable linker and the peptide moiety) forward (5'-

CAG-GAC-TCC-CCA-AAG-GAG-AAC-AGG-A-3') and reverse (5'-GA-CGT-ACC-AGG-CGG-TAG-CTG-CAT-TT-3') primers.

The resulting amplification products are displayed in Figure 6A. The obtained results show that the use of forward and/or reward peptide-tagged primers does not inhibit the amplification and results in a proenkephlain in situ hybridization probe of around 400 bp.

After silica column purification, non amplified primer dimers were removed and pure PCR amplification products were obtained (Figure 6B).

MALDI analysis of these purified PCR amplification products was carried out. Results obtained with both tagged forward and reverse primers (Figure 6C) and with a tagged forward and a conventional reward primers (Figure 6D) show that, despite the presence of significant amounts of PEG, the peptide tag is cleaved and can be detected. Indeed, although the m/z ratio of the cleaved peptide (m/z = 1704) may be in some cases attributed to a PEG signal, the characteristic ion $(M\text{-}NH_2)^+$ with a m/z ratio of 1720 can be easily detected.

1.2.4 Proenkephalin mRNA analysis in a deparaffined rat brain section using a peptide-tagged hybridization probe

In situ hybridization was performed on a deparaffined rat brain section using a peptide-tagged hybridization probe, as described before.

MALDI analysis was then performed on different tissue localizations and the resulting spectra analyzed.

Figure 7 shows the obtained spectra in two distinct brain section localizations (Figure 7A and B), as well as a control spectrum obtained by analysis of a control brain section which was hybridized with a non peptide-tagged hybridization probe.

The control spectrum shows that no signal is detected at the m/z ratios of the cleaved peptide (m/z = 1704) or the characteristic ion (m/z = 1720) when using a non peptide-tagged hybridization probe.

In contrast, proenkephalin mRNA expression can be clearly identified in the Figure 7A brain section localisation (see peaks at m/z = 1704 and 1720), while no proenkephalin mRNA expression can be detected at the second distinct localization (Figure 7B, no peak at m/z = 1704 or 1720).

The mRNA expression of proenkephalin can thus be clearly monitored at distinct spots of the tissue section, thus allowing to further reconstruct an image of the tissue section proenkephalin mRNA expression.

1.2.5 Synthesis of a dUTP-peptide conjugate with a photocleavable linker and use of such a modified, peptide-tagged dUTP to generate a multi-peptide-tagged proenkephalin probe (400 bp).

With the synthesis protocol described in the Material and Methods section, a modified peptide-tagged dUTP nucleotide has thus been synthesized, with the formula displayed on Figure 8A.

The modified peptide-tagged dUTP was analyzed by MALDI analysis to confirm the detection of the peptide tag. As shown on Figure 8B, the peptide tag is clearly detected at the expected 1163.23 $M+H^+$.

With the use of such a modified peptide-tagged dUTP, a multi-peptide-tagged nucleic acid hybridization probe for proenkephalin has been very simply generated using a simple RT-PCR amplification in the presence of dATP, dCTP, dGTP and the modified peptide-tagged dUTP, as displayed on Figure 8A This way, a specific hybridization probe could be easily synthesized for any other target mRNA sequence.

To confirm that the modified peptide-tagged dUTP was readily incorporated into the proenkephalin probe, the purified PCR product of proenkephalin amplified with either the modified peptide-tagged dUTP or a non modified dUTP with no tag was analyzed using MALDI analysis.

As shown on Figure 8, the peptide tag is readily detected in the sample of proenkephalin amplified with the modified peptide-tagged dUTP at the expected 1163.06 $M+H^+$ (Figure 8C), whereas no signal at the expected $M+H^+$ is detected in the sample of proenkephalin amplified with a non modified dUTP with no tag (Figure 8D).

Moreover, the use of a modified, peptide-tagged dUTP allows for a signal amplification since a given hybridization probe synthesized with the modified peptide-tagged dUTP will carry as many tag peptides as the number of U bases in its sequence.

Finally, the use of hybridization probes synthesized with the modified peptide-tagged dUTP allows for a quantitative analysis of mRNA expression in tissue sections. Indeed, the signal generated by the corresponding tag(s) of one or several studied mRNA(s) can be compared with that obtained for a reference mRNA sequence (for

instance a house-keeping gene such as actin of HPRT). As the number of U bases in each hybridization probe is known, the expression ratio between each studied mRNA sequence and the reference mRNA sequence can be calculated.

1.3 Conclusion

These results clearly demonstrate the possibility to use nucleic acid-photocleavable linker-peptide conjugates for the indirect simultaneous detection of multiple mRNA target molecules in tissue sections, using hybridization probes synthesized with peptide-tagged primers.

Moreover, a modified peptide-tagged dUTP has been synthesized that allows for an easier synthesis of peptide-tagged hybridization probes with any target mRNA specificity, an amplified tag signal, thus lowering the detection threshold, and the possibility to perform a quantitative mRNA expression analysis in tissue sections.

Thus, the indirect simultaneous mapping of several distinct mRNA target molecules in tissue sections using MALDI-MS analysis is now possible, with low detection threshold, possible multiplex analysis, and even possible quantitative analysis.

EXAMPLE 2

Use of an antibody-peptide conjugate with a photocleavable linker for the indirect detection of a protein is tissue sections by MALDI-MS

An antibody-peptide conjugate with a photocleavable linker was synthesized and tested for its capacity to allow indirect detection of a specific protein in tissue sections using MALDI-MS.

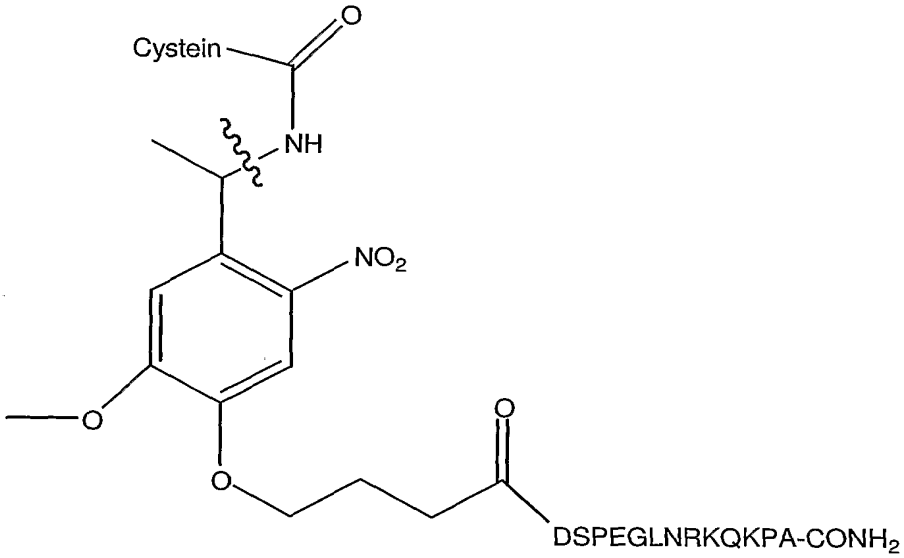
2.1 Synthesis of the antibody-peptide conjugate

A goat antibody specific for the framework region of rabbit antibodies (goat anti-rabbit antibody) has been used to synthesize an antibody-peptide conjugate with a photocleavable linker at UV-MALDI laser wavelength. This way, the conjugate is suitable as secondary antibody for the indirect analysis of any peptide, protein, antigen or hapten for which a primary rabbit antibody is available. In addition, it has to be noted that the protocol of synthesis described below is applicable to any antibody of any

specificity, including any peptide, protein, antigen or hapten, so that other conjugates in which the antibody is specific for another type of antibody or for any desirable peptide, protein, antigen or hapten may be easily synthesized using the protocol described below.

As tag molecule, a peptide was used displaying the following formula: DSPEGLNRKQKPA (SEQ ID NO: 11). However, in this case also, the protocol described below is applicable to any other peptide that might be desirable a tag molecule.

The peptide was synthesized in solid phase and directly coupled to the photocleavable linker, resulting in a peptide-phocleavable compound of formula:

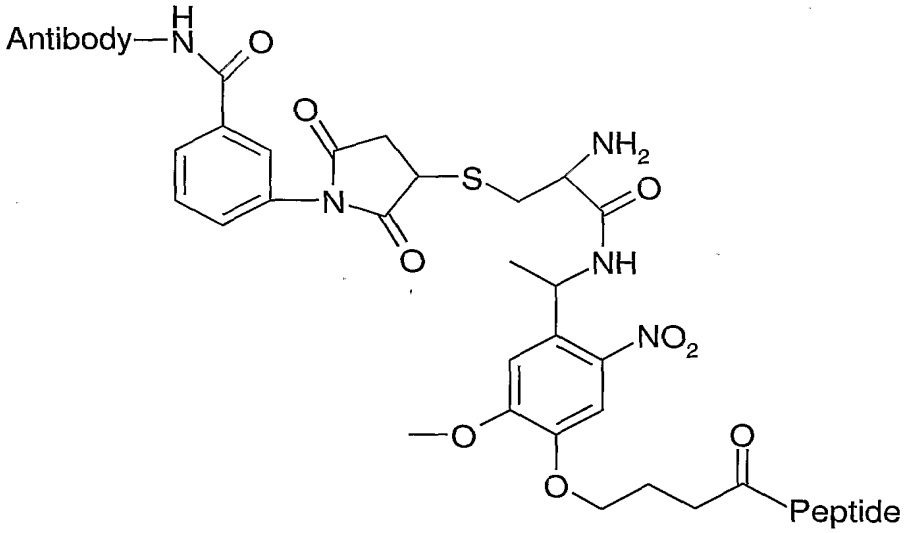


The antibody-peptide conjugate was synthesized using the following reaction steps:

1. solide phase peptide synthesis with photocleavablelinker (classical Fmoc strategy)
2. Purification of peptide using RP-HPLC (C18) using water/acetonitrile +0.5% TFA
3. 0,5 mg of MBS dissolved in 300 µl of DMF
4. 4 mg IgG (goat anti rabbit) was dissolved in 2 ml of PBS

5. Mix slowly product 3 et 4. Reaction occurs under agitation during 30 min at 20°C
6. salts were removed using PD10 with phosphate buffer 50 mM pH 6
7. 1 mg of peptide was dissolved in 300 µl of DMF then 1 ml de PBS was added.
8. Add this peptide solution on active antibody solution; the reaction occurs under agitation during 3h at 20°C in the dark.
9. PBS was removed by dialysis overnight.

The main reactions for the synthesis of the anti-body-peptide conjugate with a photocleavable linker are shown on Figure 9. The obtained antibody-peptide conjugate has the following formula:



which is shown in abbreviated form in Figure 10.

2.2 MALDI analysis of the obtained antibody-peptide conjugate in solution

The obtained antibody-peptide conjugate with a photocleavable linker has then been tested in solution first, to verify the ability of the tag peptide to be efficiently cleaved by photodissociation induced by the UV-MALDI laser.

The results are shown on Figure 10, and clearly indicate that the peptide tag, still linked to the photocleavable linker PC (peptide + PC, m/z = 1703), is easily detected after MALDI desorption/ionization process.

In addition, a fast fragmentation fragment comprising the tag peptide, the photocleavable linker, and a cystein initially belonging to the antibody (see Figure 11) (peptide + PC + cystein, m/z = 1822) is also detected, clearly demonstrating that the obtained antibody-peptide conjugate would also be suitable for other type of mass spectrometry (such as SIMS or DESI) analysis of a tissue section.

2.3 MALDI direct analysis of rat brain tissue sections using the obtained antibody-peptide conjugate

The obtained antibody-peptide conjugate with a photocleavable linker has then been tested for indirect UV-MALDI analysis of membrane protein carboxypeptidase D (180 kDa)in rat brain tissue sections.

Rat brain tissue sections were first hybridized with a primary rabbit antibody specific for carboxypeptidase D. In a second step, the antibody-peptide conjugate with a photocleavable linker was added to hybridize with primary antibodies present on the tissue section.

The resulting stained tissue sections were then analyzed using UV-MALDI.

A typical spectrum is displayed on Figure 12. It clearly indicates that the peptide tag, still linked to the photocleavable linker PC (peptide + PC, m/z = 1703.23), is easily detected using indirect UV-MALDI tissue section analysis. Another fragment corresponding to the tag peptide (peptide + PC+O, m/z = 1686.43) is also easily detected.

Using the method according to the invention, the rat brain tissue section was scanned, each spectrum data being stored, and the image of the expression map of carboxypeptidase D was constructed. The result is displayed in Figure 13, and clearly shows that it is possible to construct an expression map using the method according to the invention with an antibody-peptide conjugate with a photocleavable linker.

In addition, to further demonstrate the reliability of the method according to the invention to construct expression maps, the rat brain tissue section stained with rabbit primary antibodies specific for carboxypeptidase D where developed using T-chloronaphtol as chromogen after fixation on a classical commercial secondary antibody. The obtained staining colocalized with the MALDI detection of

carboxypeptidase D (data not shown), thus demonstrating the reliability of the method according to the invention and MALDI imaging shows to be even more sensitive than classical revelation.

These results demonstrate that the synthesized antibody-peptide conjugate with a photocleavable linker permits the indirect detection of a protein in a tissue section, as well as the construction of an expression map.

EXAMPLE 3

Demonstration of the possibility to use fast fragmentation instead of photodissociation for indirect mass spectrometry analysis of biomolecules present in tissue sections

Fast fragmentation is usually considered as a detrimental but inevitable phenomenon.

In the present case, the inventors have found that it is possible to take advantage of this necessary phenomenon to implement a method for determining at least one target molecule map in a tissue section using (A-X)n-B conjugates in which the linker X is cleaved by fast fragmentation.

3.1 Fast fragmentation for indirect mass spectrometry analysis of biomolecules present in tissue sections using conjugates in which the linker is photocleavable at the wavelength of a UV-MALDI laser

Although this concept is transposable to other mass spectrometry technologies, such as SIMS or DESI mass spectrometry, the inventors have first proven the possible use of this concept using MALDI mass spectrometry and conjugates in which the linker is photocleavable at the wavelength of a UV-MALDI laser..

Indeed, when using either oligo-peptide conjugates or antibody-peptide conjugates with photocleavable linkers (PC), in addition to expected (peptide + PC) fragments, other fragments comprising the peptide tag are observed, these fragments corresponding to fast fragmentation fragments of the conjugates.

More precisely, when using 3 distinct oligo-peptide conjugates with photocleavable linkers (see Example 1 for a more precise definition of these conjugates), fragments corresponding to the tag peptide 1, peptide 2 or peptide 3 are observed in addition to expected fragments (peptide 1, 2 or 3 +PC), as is clearly shown on Figure 4C. For instance, for conjugate 2 using a peptide 2 tag molecule, both a photodissociation fragment and a fast fragmentation fragment are observed:

- expected photodissociation fragment = (peptide 2 + PC), m/z = 1892; and
- fast fragmentation fragment = peptide 2 only, m/z = 1450.

The same phenomenon is observable for conjugate 3 on Figure 4C:

- expected photodissociation fragment = (peptide 3 + PC), m/z = 1604 ; and
- fast fragmentation fragement = peptide 3 only, m/z = 1162.

When using the antibody-peptide conjugate described in Example 2 also, both an expected photodissociation fragment and a fast fragmentation fragment are observed as shown in Figure 11:

- expected photodissociation fragment = (peptide + PC), m/z = 1702; and
- fast fragmentation fragment = peptide + PC + cystein, m/z = 1822.

3.2 Fast fragmentation for indirect mass spectrometry analysis of biomolecules present in tissue sections using a FITC labelled antibody

The inventors have also analyzed the possibility to use fast fragmentation for indirect mass spectrometry analysis of biomolecules present in tissue sections using conventional, and notably commercial, labelled antibodies.

Indeed, conventional labelling molecules have definite molecular weights and may thus be used as tag molecules. In addition, all sorts of antibody, with a very wide range of antigenic specificities, are commercially available as labelled antibodies. Finally, technologies to attach a labelling molecule via conventional linkers to any antibody are well-known routine technologies, thus ensuring that the possibility to use conventionally labelled antibodies for indirect mass spectrometry analysis of biomolecules present in tissue sections would permit to significantly enlarge the number of distinct biomolecules that may be mapped using the method according to the invention.

A FITC labelled antibody has been used for in solution MALDI analysis of a FITC labelled antibody using CHCA matrix.

Results are displayed in Figure 14 and show that a characteristic signal for FITC molecule is easily detected (m/z = 365.18). Since the linker between the antibody and the FITC molecule is not photocleavable at the MALDI laser wavelength, it clearly demonstrate that fast fragmentation permits the cleavage of the linkage between the antibody and its labelling molecule.

Although these results have yet to be confirmed on tissue section, all preceding results obtained in solution have been further confirmed on tissue sections. These results thus highly support the possibility to use fast fragmentation and conventional labelled antibodies for indirect detection of target molecules in tissue section using MALDI mass spectrometry, but also other mass spectrometry technologies involving fast fragmentation, and notably SIMS or DESI mass spectrometry.

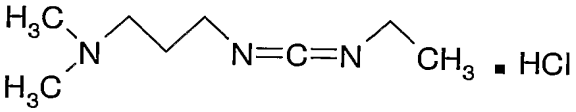
3.3 Conclusion

The obtained results clearly demonstrate the possibility to take advantage of the usually detrimental phenomenon of fast fragmentation for the indirect detection of target molecules in tissue section using various mass spectrometry technologies, in particular MALDI, SIMS or DESI mass spectrometry.

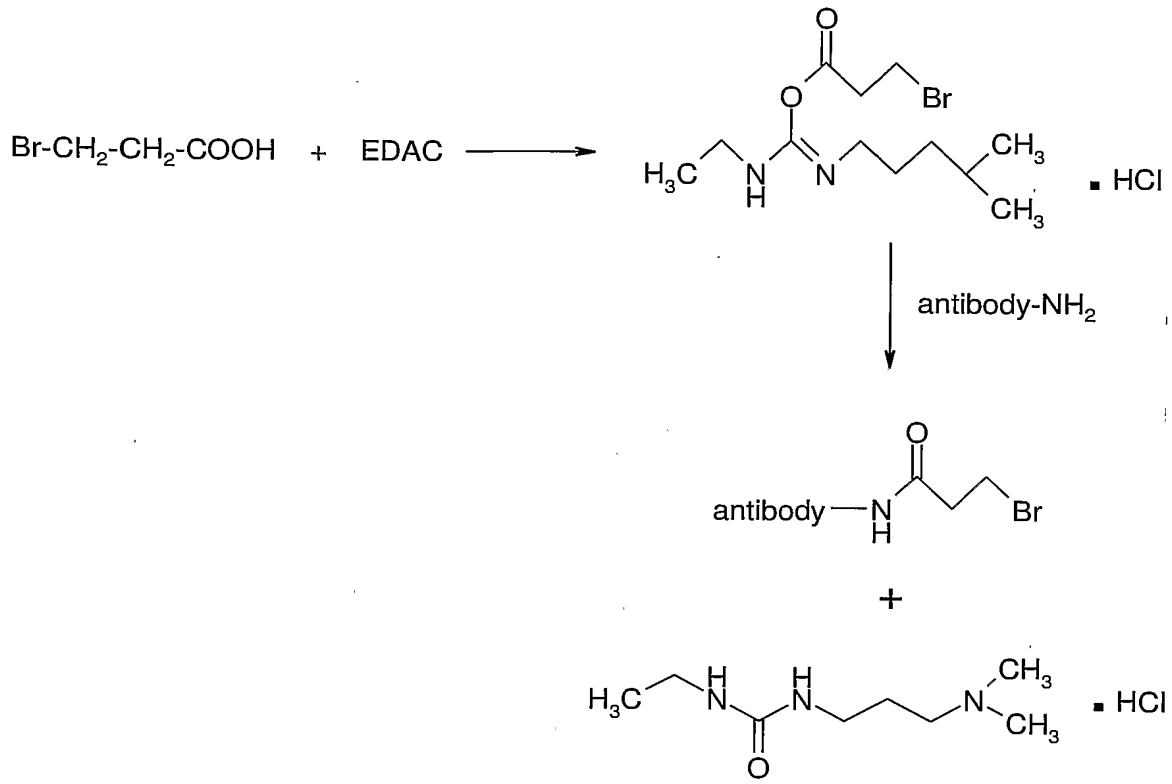
EXAMPLE 4

Synthesis of an antibody-Br conjugate for use in mass spectrometry analysis of tissue sections using fragmentation to cleave the linker molecule between the antibody and the Br atom

An antibody-Br conjugate may be synthesized using EDAC linker:



available from SIGMA under reference E1769, using the following protocol:



Such a protocol may be applied using any antibody or antibody fragment, to conjugate it with a Br atom. Such conjugates may be used in a method according to the invention in which fragmentation is used to cleave the linker between the antibody and the Br atom.

CLAIMS

1. A method for determining at least one target molecule map in a tissue section, comprising :

a) hybridizing said tissue section with at least one $(A\text{-}X)_n\text{-}B$ conjugate, wherein

- A is a tag molecule of known molecular weight,
- X is a linker that is cleaved during sample desorption/ionization,
- n is an integer of at least 1,
- B is a binding molecule that binds specifically to said target molecule, and
- each distinct B molecule is linked to a distinct A tag molecule;

b) scanning the tissue section surface and analyzing each adjacent spot with a mass spectrometer, wherein said linker X is cleaved during sample desorption/ionization, and wherein the resulting data of each spot is saved; and

c) analyzing the obtained data in the molecular mass window(s) of each distinct tag molecule to create as many maps of the tissue section as the number of distinct studied target molecules.

2. The method of claim 1, wherein each target molecule is independently chosen in the group constituted of nucleic acids, in particular mRNA molecules, peptides, proteins, in particular receptors and ligands, antibodies, antigens, aptamers, haptens, and organic compounds.

3. The method of claim 2, wherein at least one target molecule is a mRNA molecule.

4. The method of claim 2, wherein at least one target molecule is a peptide, a protein, an antigen, or a hapten.

5. The method of any of claims 1-4, wherein each B molecule that binds specifically to a target molecule is independently chosen in the group constituted of nucleic acids, particularly oligonucleotides, peptides, proteins, in particular receptors and ligands, antibodies, antigens, haptens and organic compounds.

6. The method of claim 5, wherein each B molecule that binds specifically to a target nucleic acid is a nucleic acid probe with a sequence that is complementary to said target mRNA sequence.

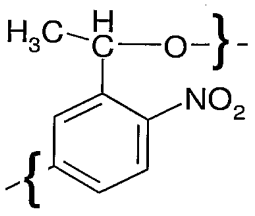
7. The method of claim 5, wherein each B molecule that binds specifically to a target peptide, protein, antigen, or hapten is an antibody directed against said peptide, protein, antigen or hapten.

8. The method of any of claims 1-7, wherein each A tag molecule is chosen in the group constituted of peptides, nucleic acids, sugars, polymers, lipids and organic compounds.

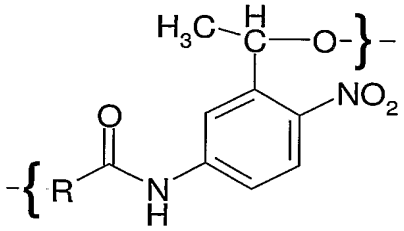
9. The method of claim 8, wherein at least one A tag molecule is a peptide.

10. The method of any of claims 1-9, wherein MALDI mass spectrometry is used and the X linker molecule is photocleavable at the wavelength of a MALDI laser.

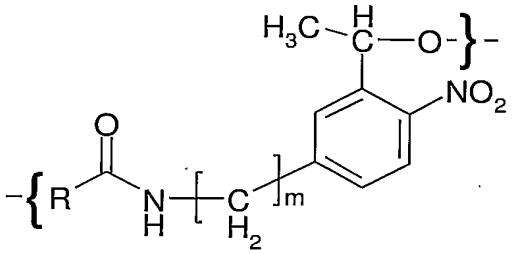
11. The method of claim 10, wherein the X linker molecule comprises a moiety chosen in the group constituted of :



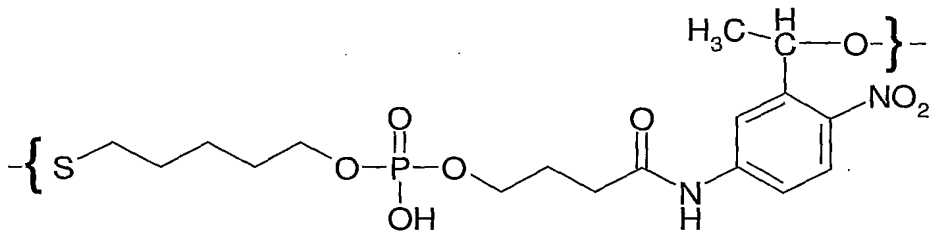
,



,

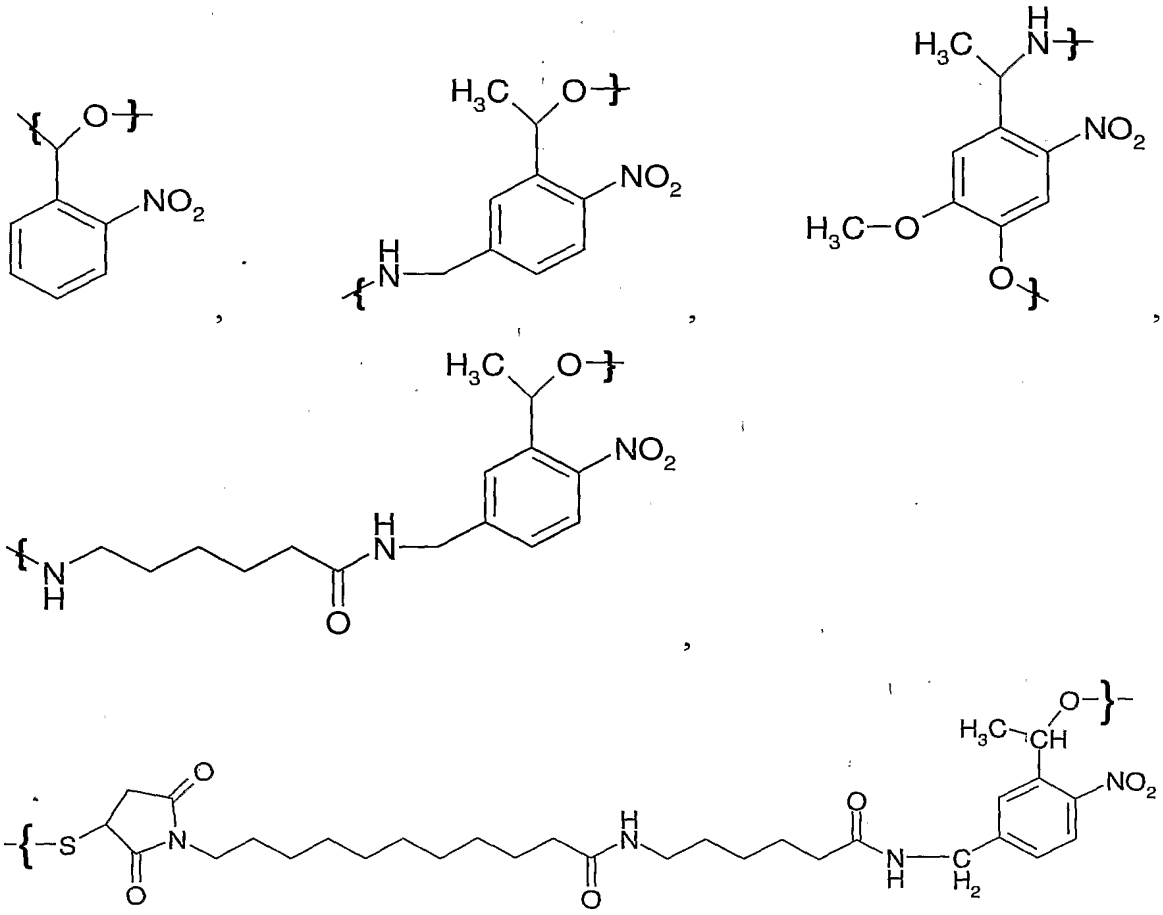


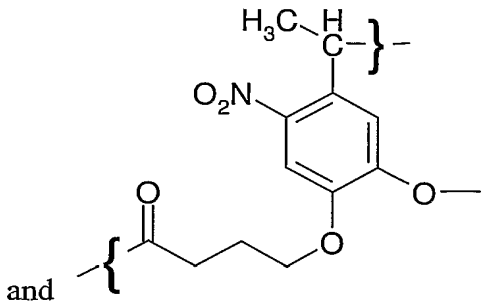
, and



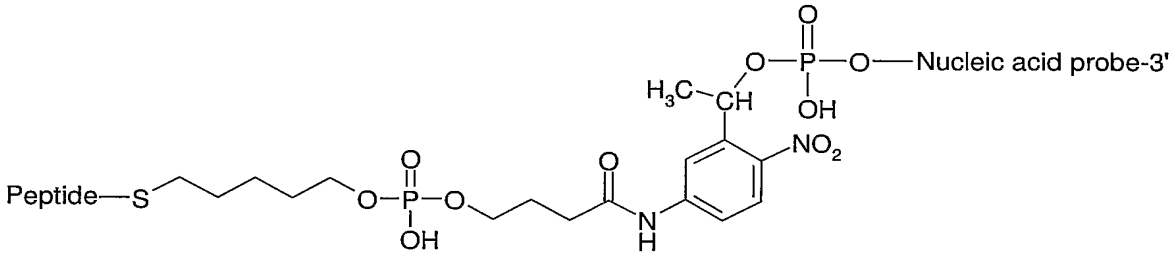
, wherein R is a C1-C6 alkyl group and m is an integer comprised between 1 and 4.

12. The method of claim 10, wherein the X linker molecule comprises a moiety chosen in the group constituted of :

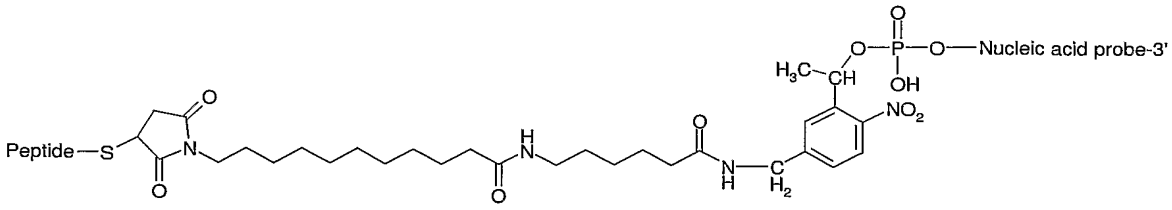




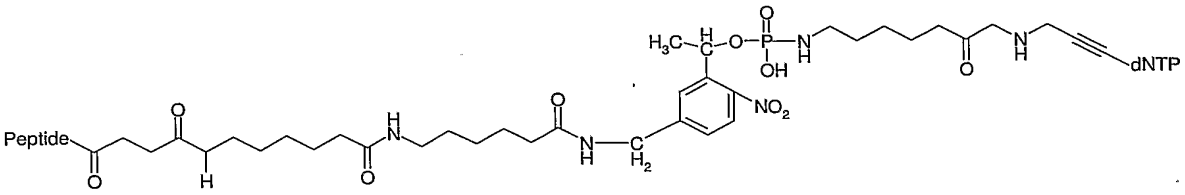
13. The method of claim 10, wherein target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, A tag molecule(s) are peptides, n is 1 and the (A-X)-B conjugate(s) have the following structure:



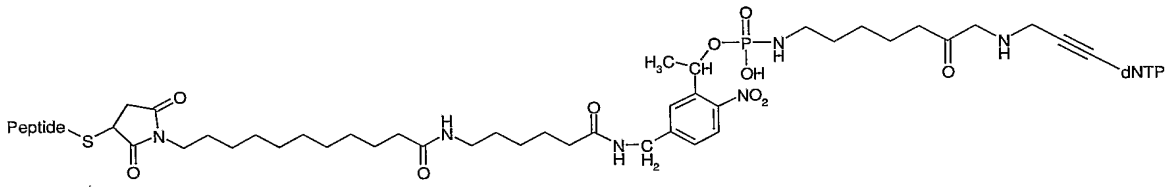
14. The method of claim 10, wherein target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, A tag molecule(s) are peptides, n is 1 and the (A-X)-B conjugate(s) have the following structure:



15. The method of claim 10, wherein target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, A tag molecule(s) are peptides, n is superior to 1 and the nucleic acid probe(s) comprise at least one modified base of the following structure:



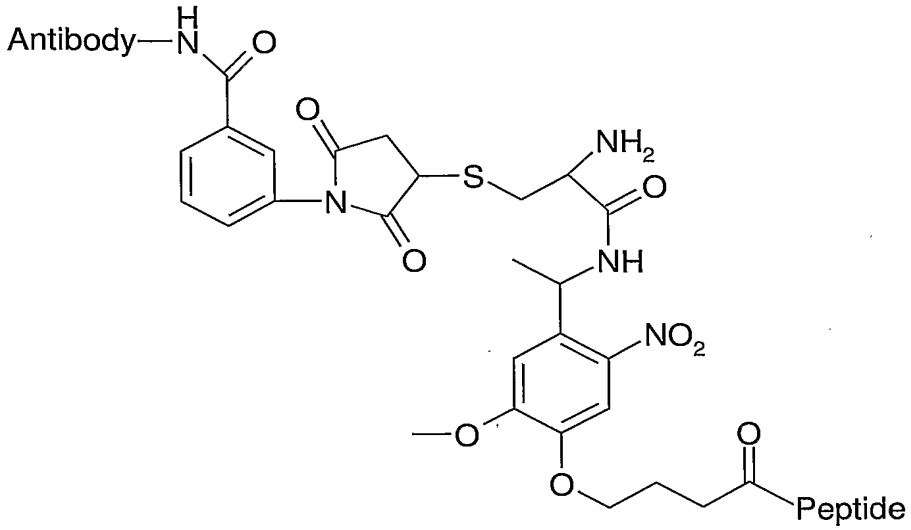
16. The method of claim 10, wherein target molecule(s) are mRNA molecules, binding B molecule(s) are nucleic acid probes with a sequence complementary to mRNA sequences, A tag molecule(s) are peptides, n is superior to 1 and the nucleic acid probe(s) comprise at least one modified base of the following structure:



17. The method of any of claims 13-16, wherein at least 5 distinct mRNA molecules are mapped on the tissue section.

18. The method of any of claims 13-17, further comprising a step d) consisting in analyzing the obtained data in the molecular mass windows of each mRNA molecule corresponding protein to create each corresponding protein expression map of the tissue section.

19. The method of claim 10, wherein target molecule(s) are peptide(s), protein(s) or hapten(s), binding B molecule(s) are antibodies directed against said target molecules, A tag molecule(s) are peptides, n is 1 and the (A-X)-B conjugate(s) have the following structure:



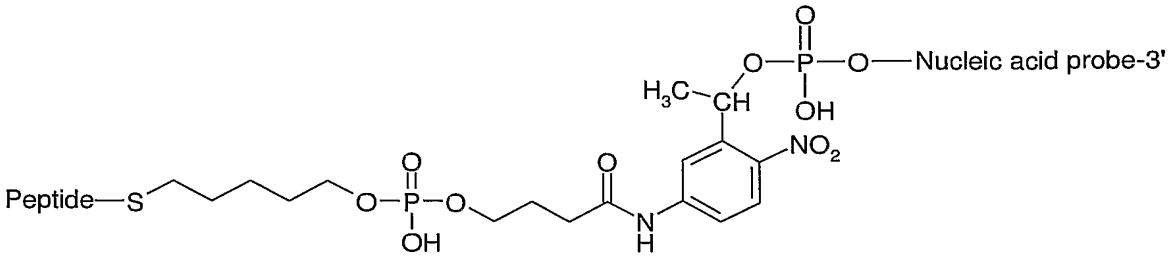
20. The method of any of claims 1-9, wherein said X linker molecule is cleaved by fragmentation during sample ionization.

21. The method of claim 20, wherein said X linker molecule is as defined in any of claims 11-12.

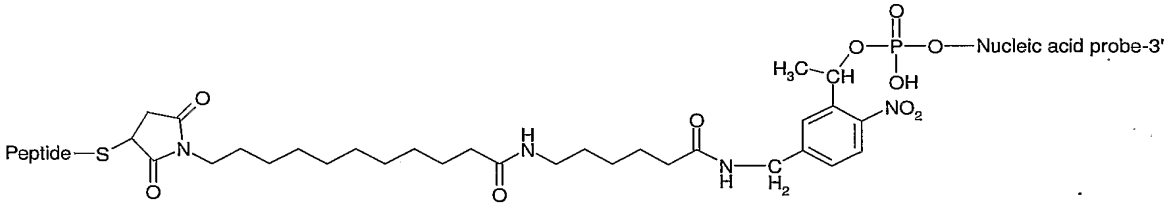
22. The method of claim 20, wherein , target molecule(s) are peptides, proteins (including antibodies), antigens, or haptens, binding molecules are antibodies or antibody fragments, n is 1 and tag molecules are fluorochromes, enzymes, biotin or gold.

23. The method of any of claims 20-22, wherein UV-MALDI, IR-MALDI, SIMS or DESI mass spectrometry is used.

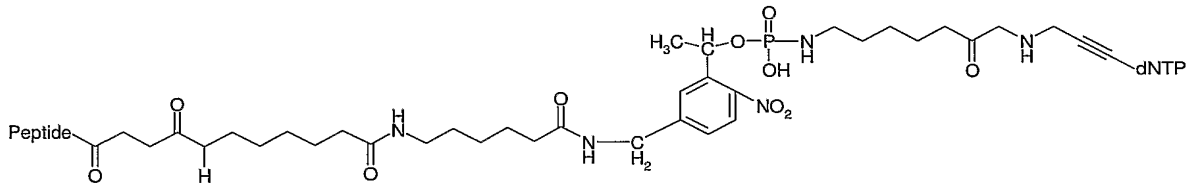
24. A conjugate suitable for use in a method according to any of claims 1, 10 or 20, wherein said conjugate is selected from:


Peptide—S
O—P—O
OH
O
N
H
O—P—O—Nucleic acid probe-3'
OH
H3C—CH
NO2

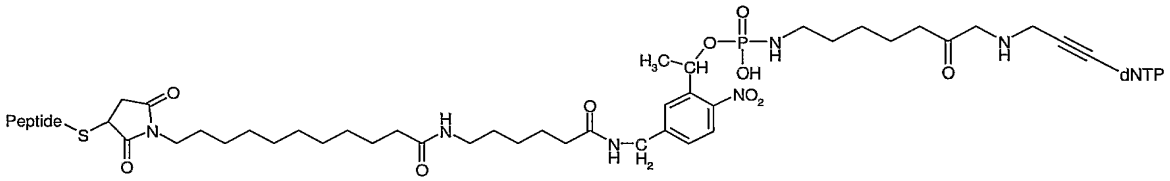
,

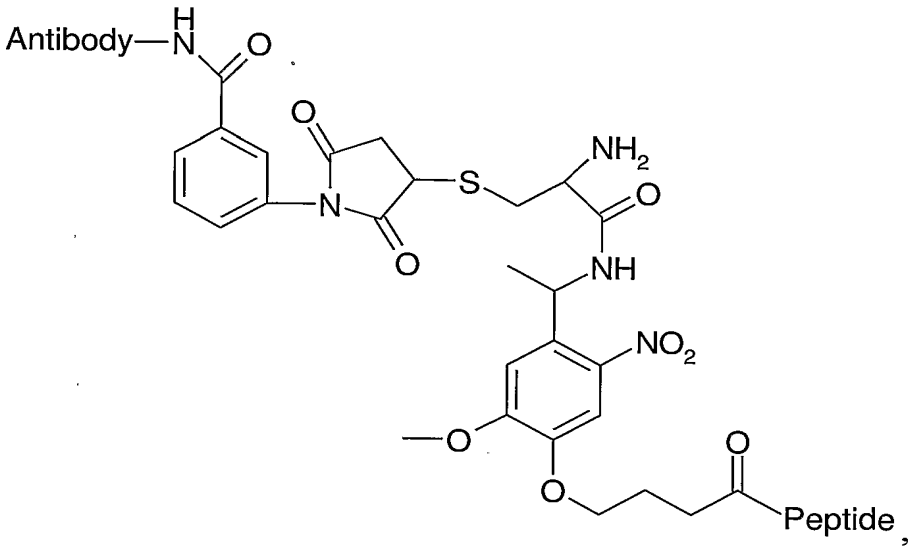

Peptide—S
O—P—O—Nucleic acid probe-3'
OH
H3C—CH
NO2

,

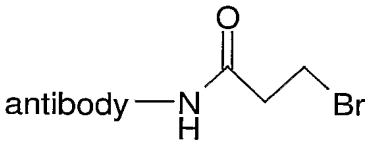

Peptide
H3C—CH
O—P—N
OH
NO2
dNTP

,


Peptide—S
H3C—CH
O—P—N
OH
NO2
dNTP



, or



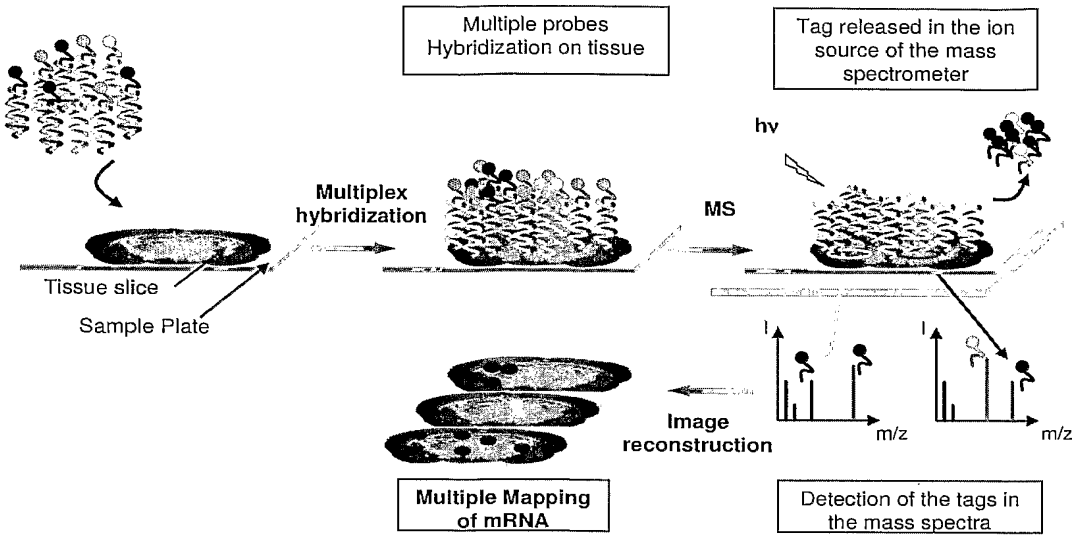


Figure 1

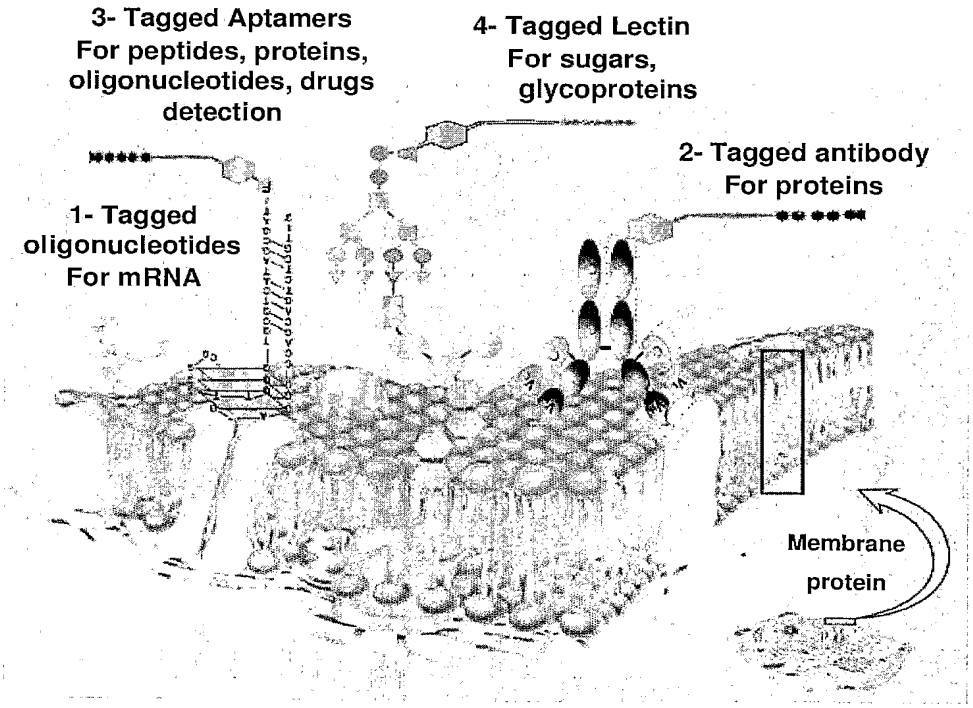


Figure 2

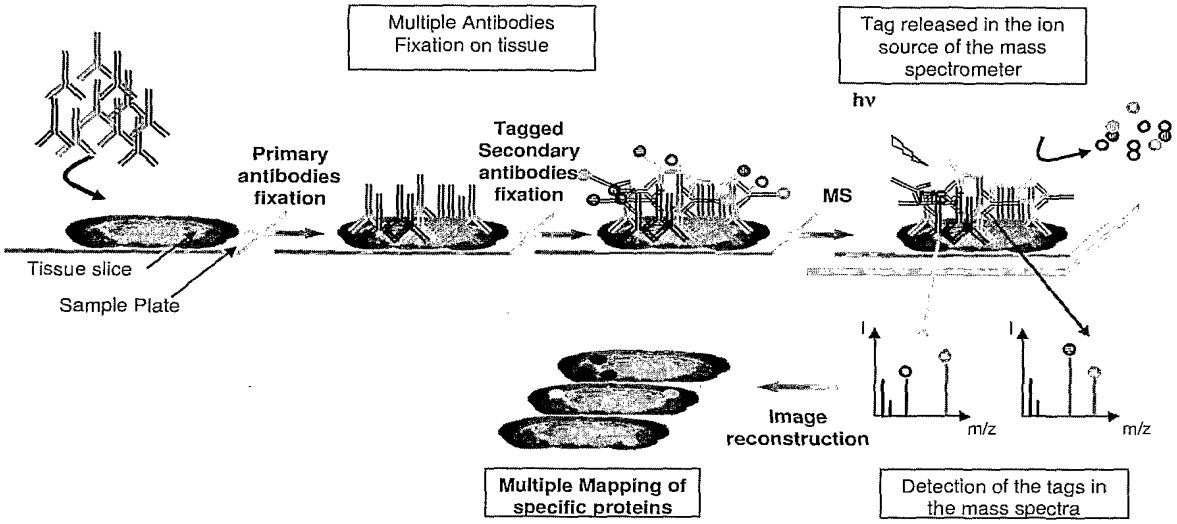

Multiple Antibodies Fixation on tissue
Tag released in the ion source of the mass spectrometer
hν
Primary antibodies fixation
Tagged Secondary antibodies fixation
MS
Tissue slice
Sample Plate
I
m/z
I
m/z
Image reconstruction
Multiple Mapping of specific proteins
Detection of the tags in the mass spectra

Figure 3

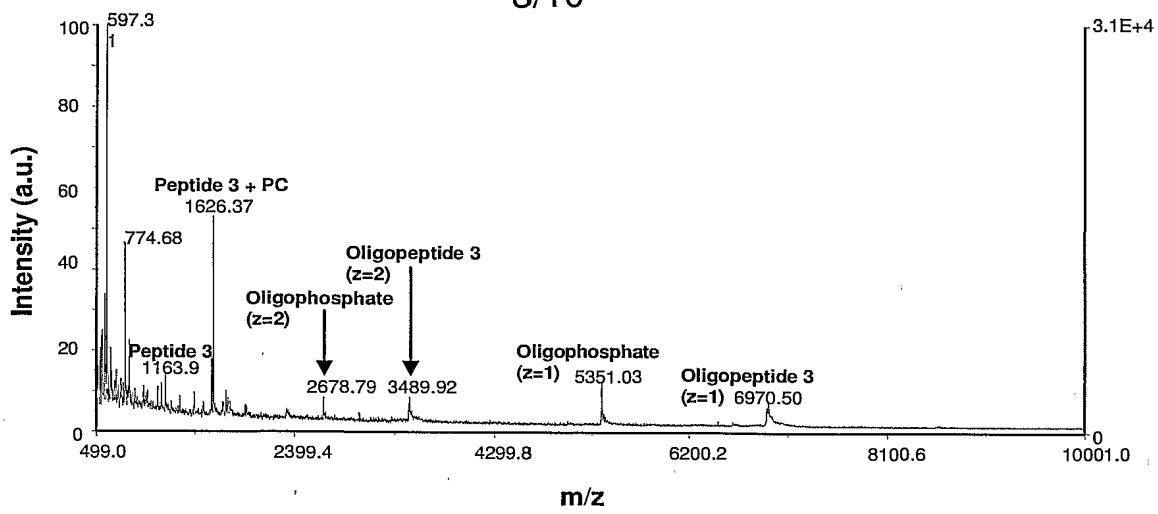


Figure 4A

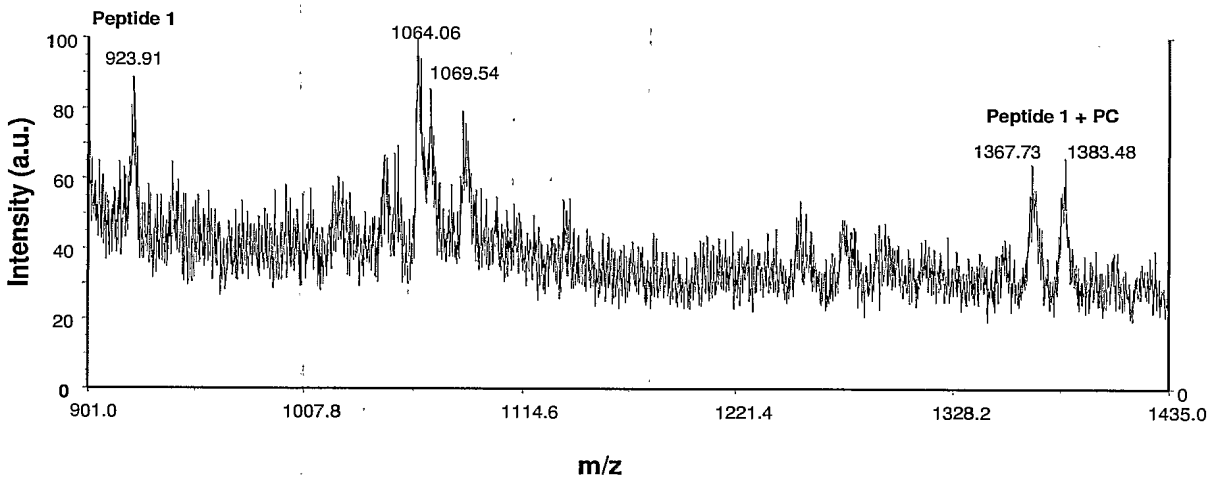


Figure 4B

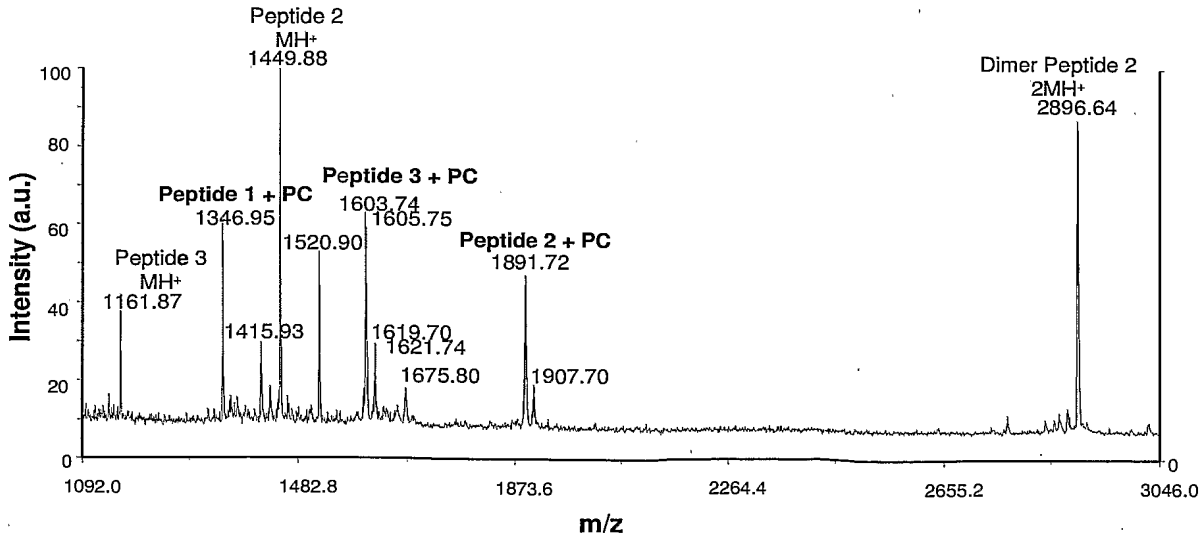


Figure 4C

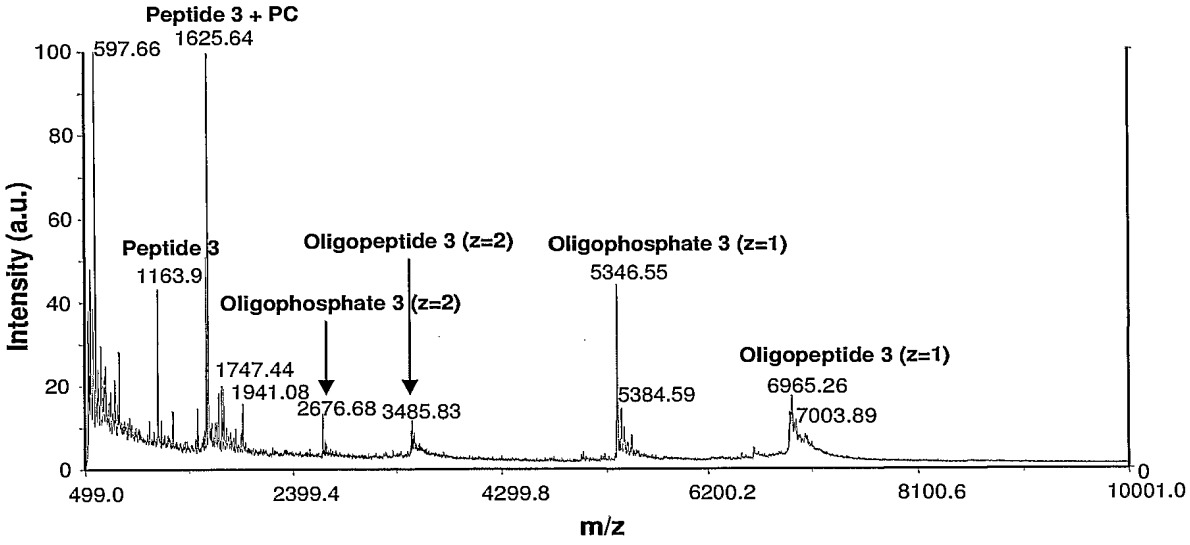


Figure 4D

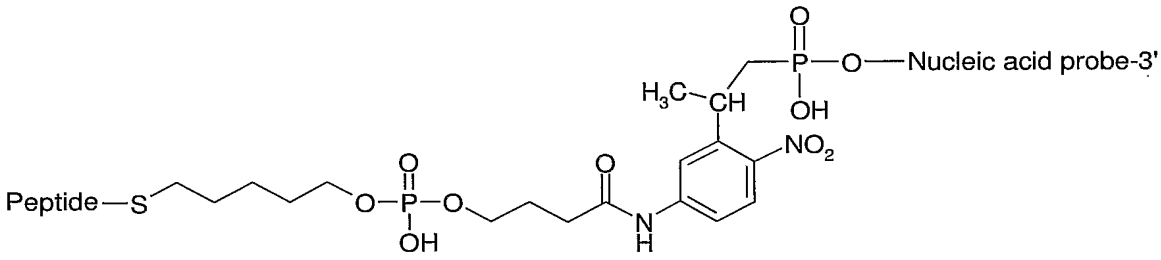


Figure 4E

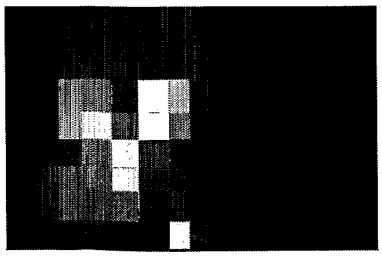


Figure 5

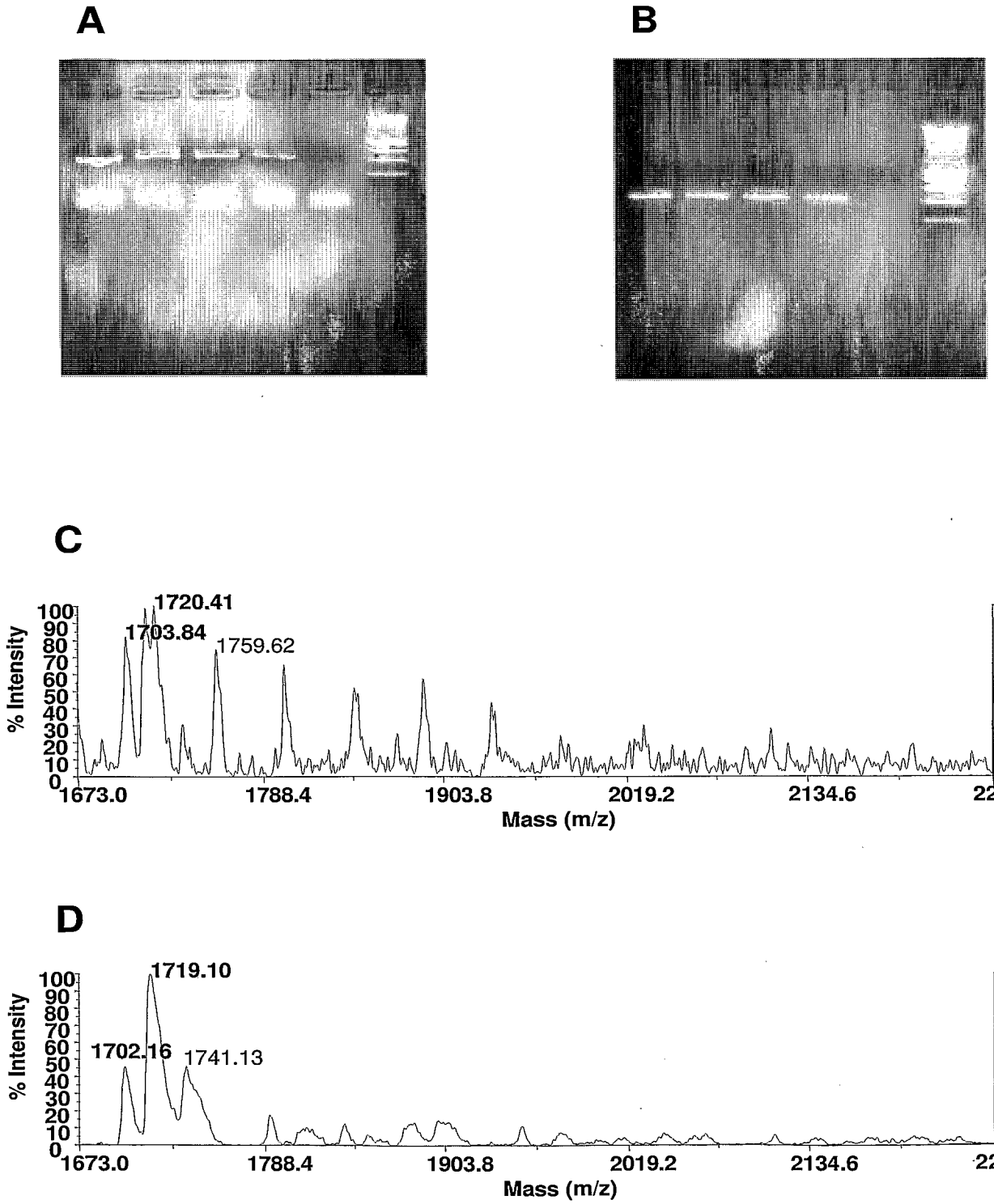
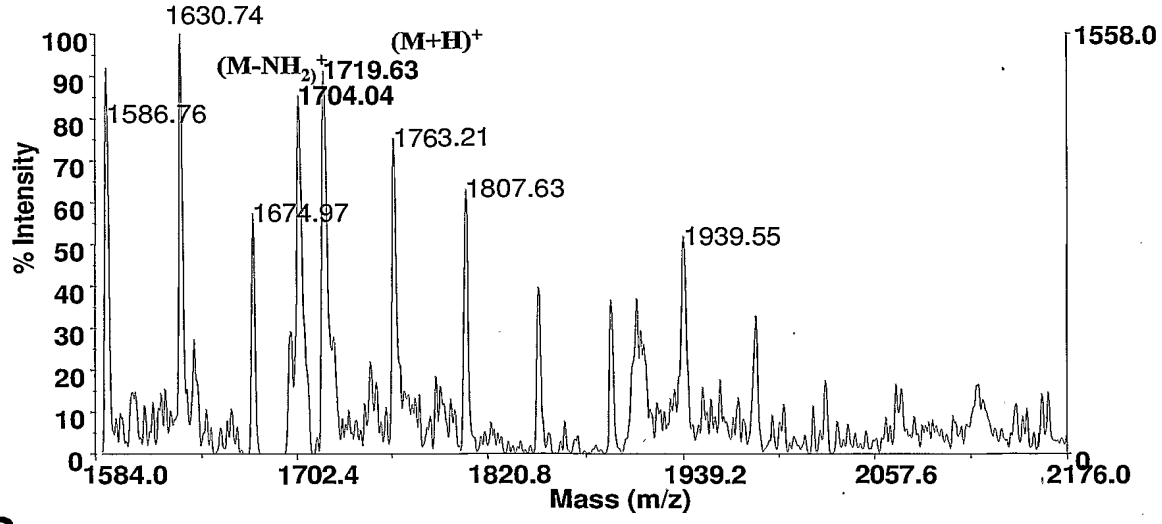
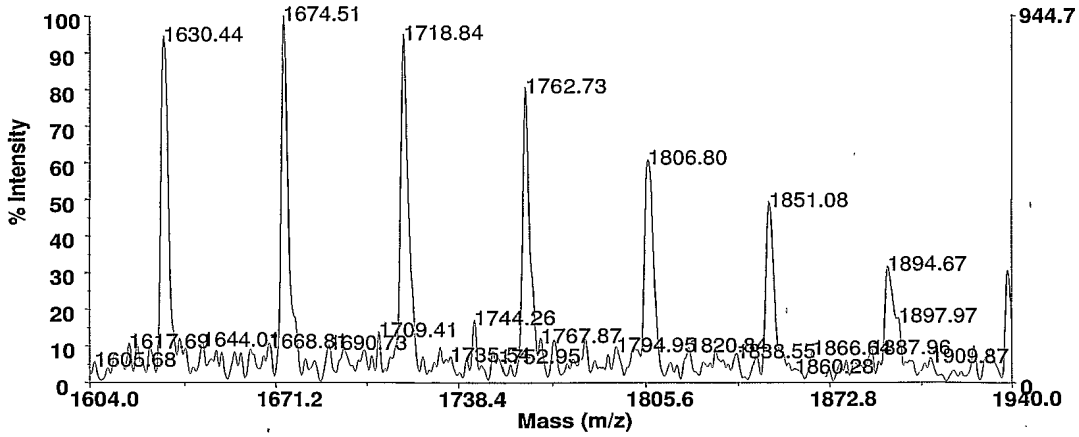

A
B
C
1720.41
1703.84
1759.62
% Intensity
100
90
80
70
60
50
40
30
20
10
0
1673.0
1788.4
1903.8
2019.2
2134.6
2250.0
Mass (m/z)
864.5
0
D
1719.10
1702.16
1741.13
% Intensity
1673.0
1788.4
1903.8
2019.2
2134.6
2250.0
Mass (m/z)
4.0E+4
0

Figure 6

A


1630.74
(M+H)+
(M-NH2)+
1719.63
1704.04
1586.76
1763.21
1807.63
1674.97
1939.55
1558.0
% Intensity
Mass (m/z)
1584.0
1702.4
1820.8
1939.2
2057.6
2176.0

B


1674.51
1630.44
1718.84
1762.73
1806.80
1851.08
1894.67
1897.97
944.7
% Intensity
Mass (m/z)
1604.0
1671.2
1738.4
1805.6
1872.8
1940.0

C

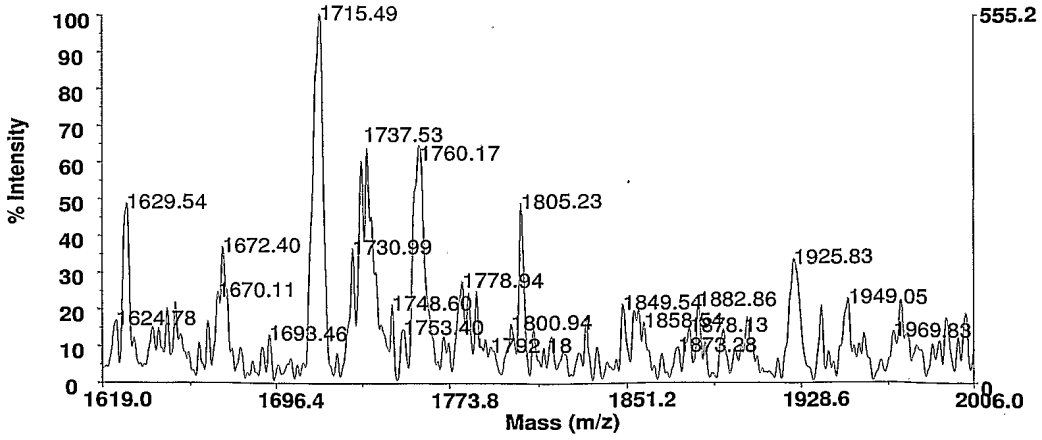

1715.49
1737.53
1760.17
1629.54
1805.23
1672.40
1730.99
1925.83
1778.94
1670.11
1949.05
1748.60
1882.86
1624.78
1693.46
1753.40
1800.94
1858.54
1878.13
1969.83
555.2
% Intensity
Mass (m/z)
1619.0
1696.4
1773.8
1851.2
1928.6
2006.0

Figure 7

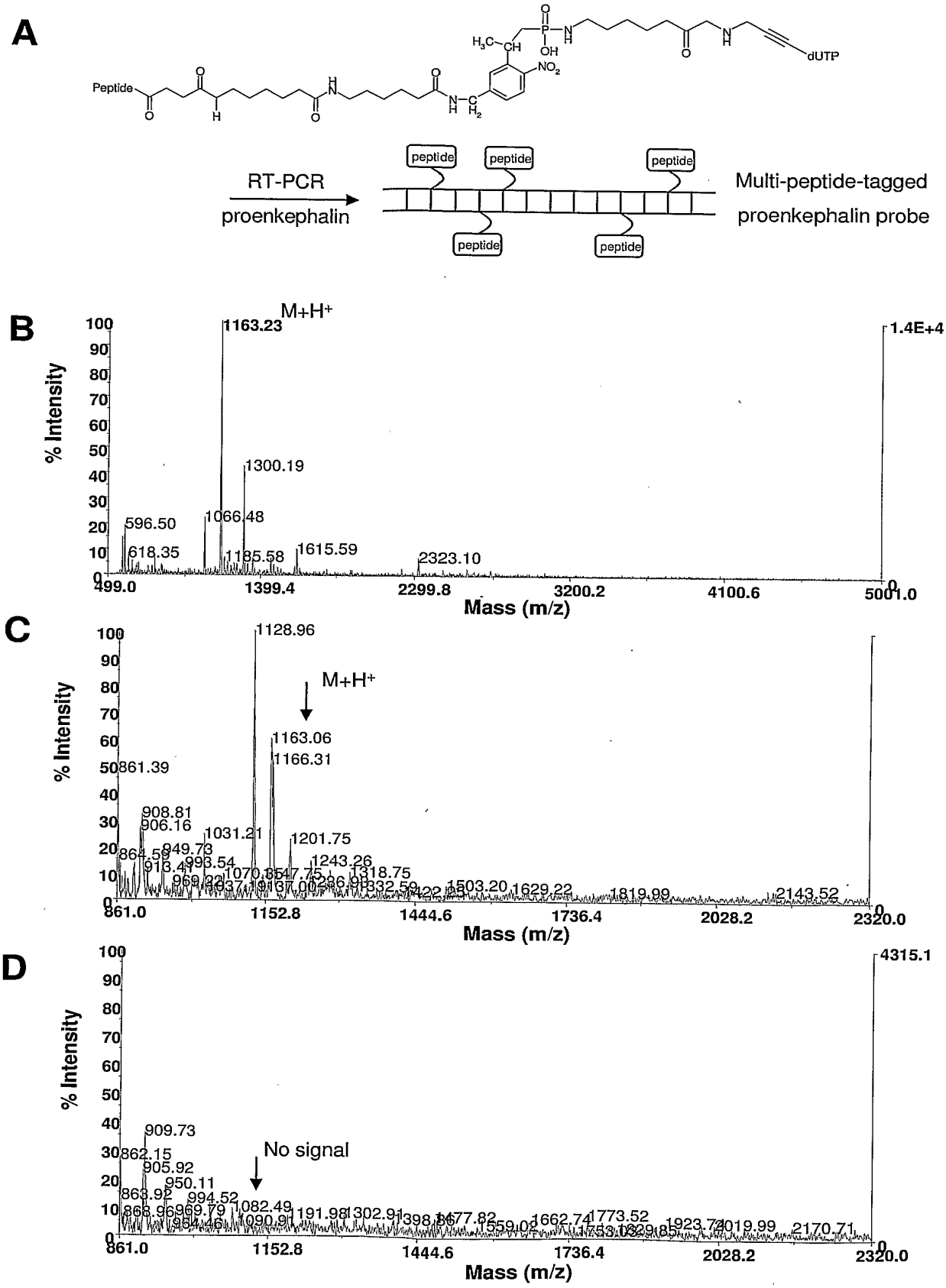

A
Peptide
dUTP
RT-PCR
proenkephalin
peptide
Multi-peptide-tagged
proenkephalin probe
B
M+H+
1163.23
1.4E+4
1300.19
596.50
1066.48
618.35
1185.58
1615.59
2323.10
% Intensity
Mass (m/z)
C
1128.96
M+H+
1163.06
1166.31
861.39
908.81
906.16
1031.21
1201.75
949.73
1243.26
1318.75
1503.20
1629.22
1819.99
2143.52
% Intensity
Mass (m/z)
D
4315.1
909.73
No signal
862.15
905.92
950.11
994.52
1082.49
1191.98
1302.91
1773.52
2019.99
2170.71
% Intensity
Mass (m/z)

Figure 8

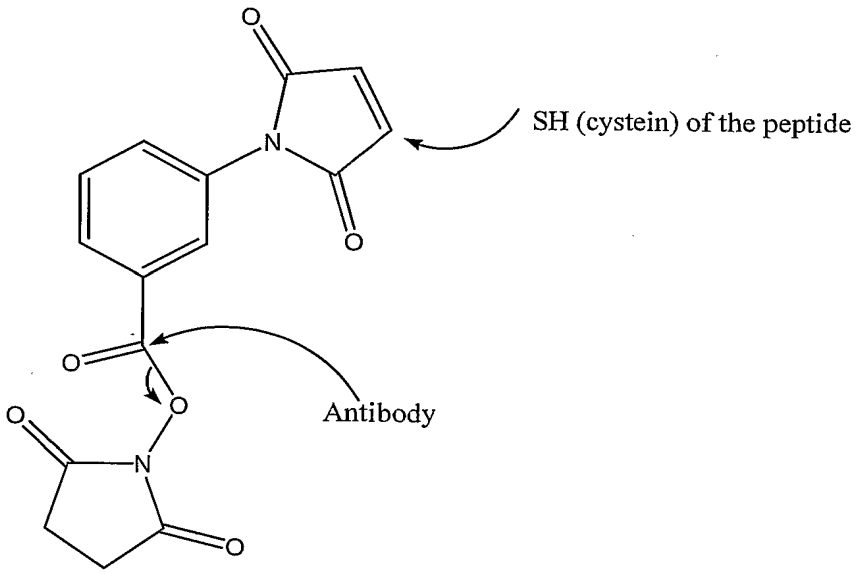


Figure 9

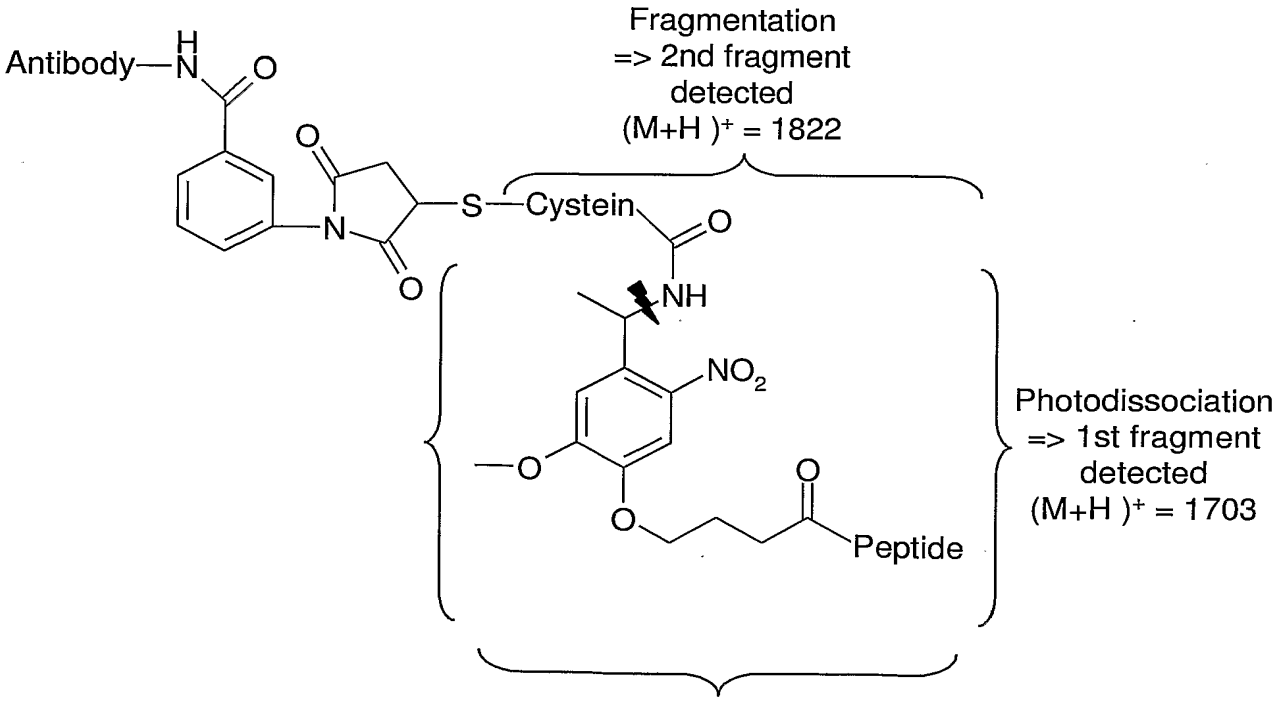


Figure 10

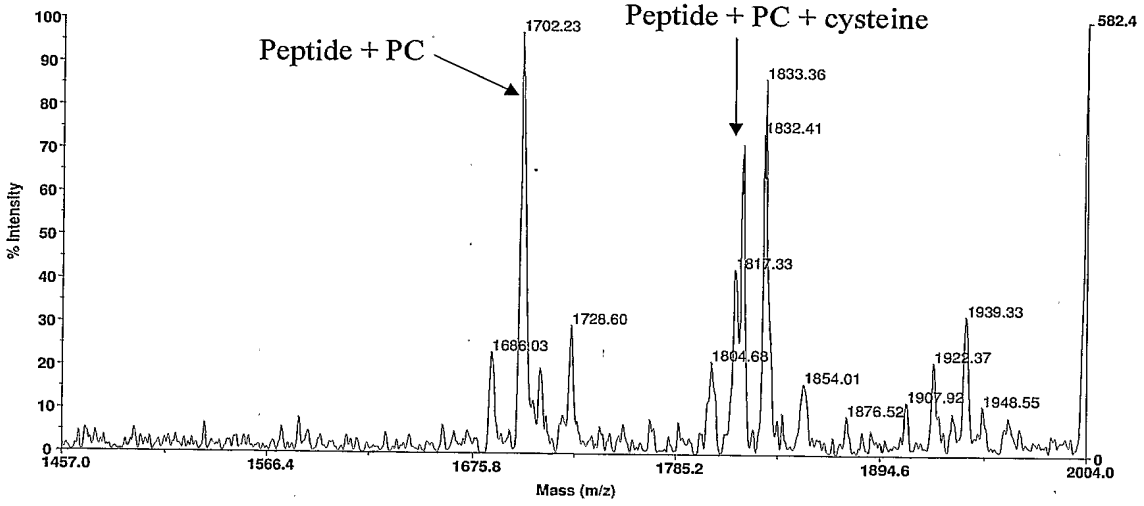


Figure 11

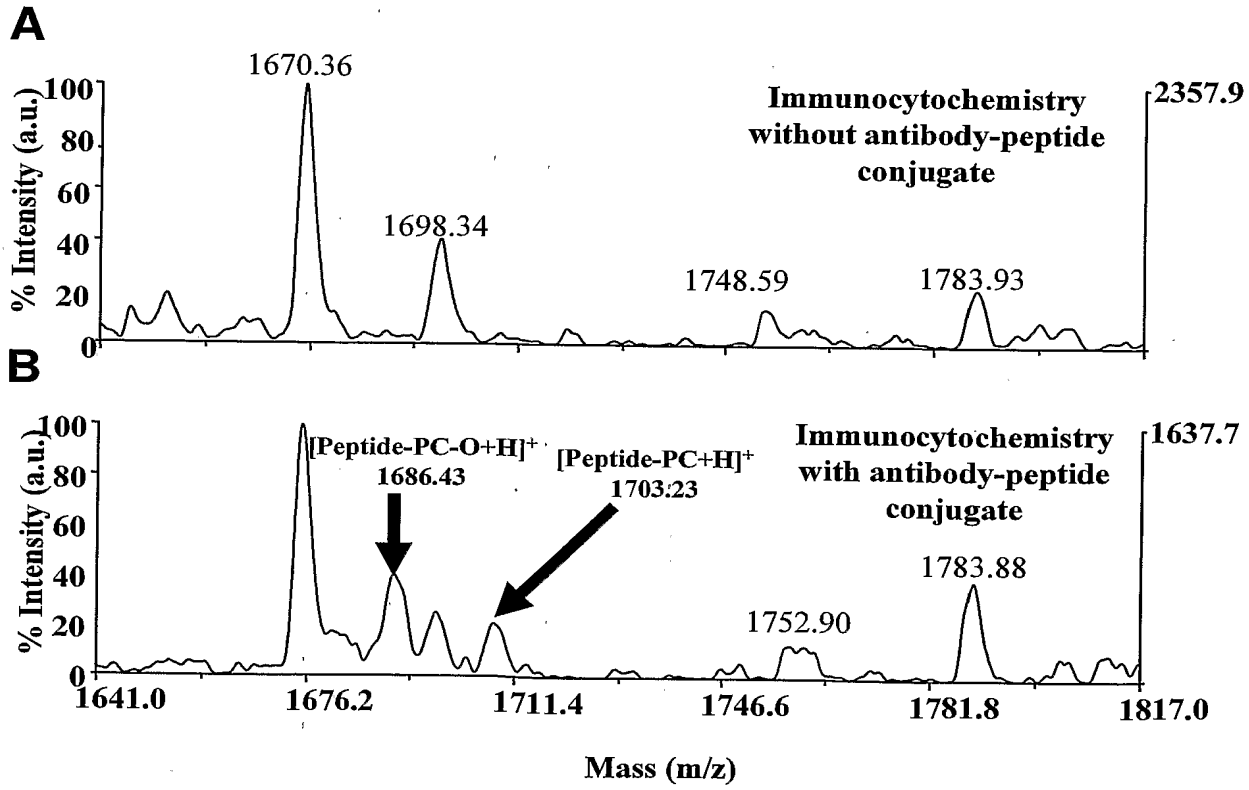


Figure 12

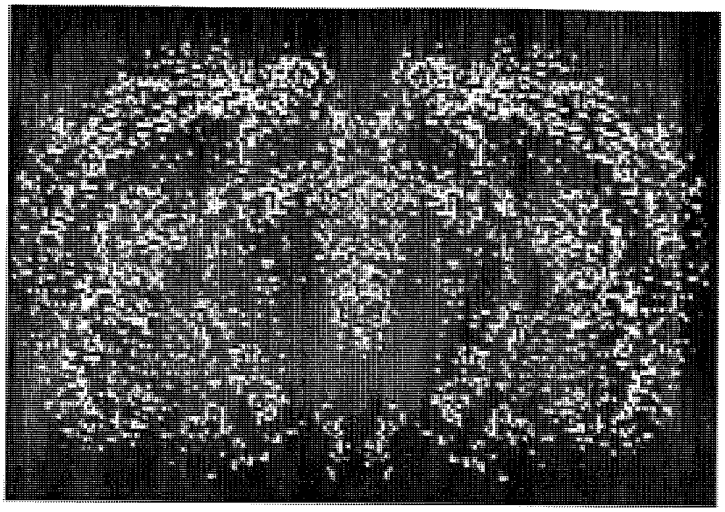


Figure 13

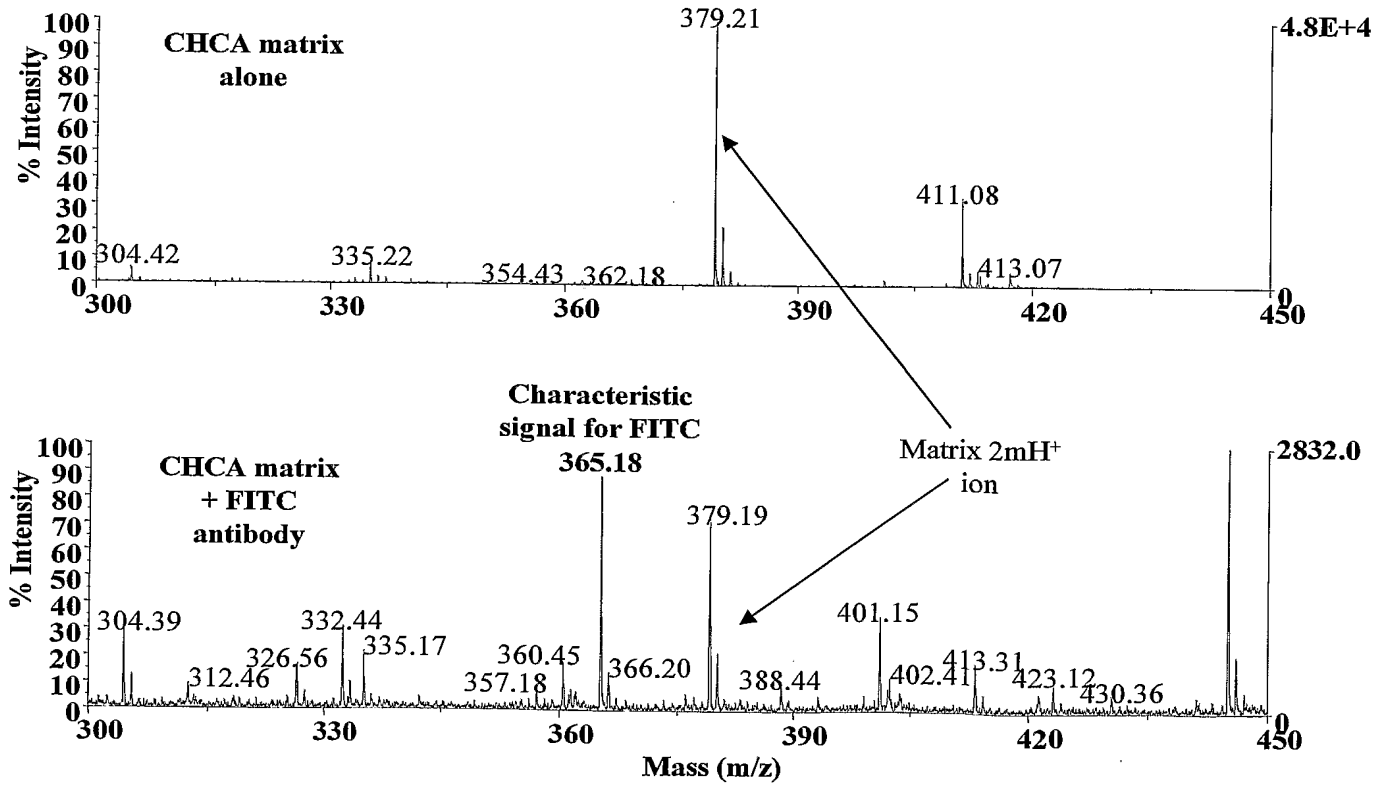


Figure 14

SEQUENCE LISTING

<110> CNRS
UNIVERSITE DES SCIENCES ET TECHNOLOGIES DE LILLE
EUROGENTEC SA
FOURNIER, Isabelle
LEMAIRE, Rémi
DESCHAMPS, Michel
SALZET, Michel
TABET, Jean-Claude
PROESS, Gottfried
RUDLOFF, Ivo
LEMAITRE, Marc

<120> Use of conjugates with linkers cleavable by photodissociation or fragmentation for mass spectrometry analysis of tissue sections

<130> 349763-D23437

<150> US 60/687,848
<151> June 7, 2005

<160> 11

<170> PatentIn version 3.1

<210> 1
<211> 9
<212> PRT
<213> artificial

<220>
<223> Tag peptide of oligo-peptide 1

<400> 1

Gly Arg Ala Leu Gly Val Phe Val Cys
1 5

<210> 2
<211> 12
<212> PRT
<213> artificial

<220>
<223> Tag peptide of oligo-peptide 2

<400> 2

Arg Pro Lys Pro Gln Gln Phe Phe Gly Leu Met Cys
1 5 10

<210> 3
<211> 10
<212> PRT
<213> artificial

<220>

<223> Tag peptide of oligo-peptide 3

<400> 3

Arg Pro Pro Gly Phe Ser Pro Phe Arg Cys
1 5 10

<210> 4
<211> 10
<212> PRT
<213> artificial

<220>
<223> Tag peptide of oligo-peptide Forward

<400> 4

Arg Pro Pro Gly Phe Ser Pro Phe Arg Cys
1 5 10

<210> 5
<211> 10
<212> PRT
<213> artificial

<220>
<223> Tag peptide of oligo-peptide Reverse

<400> 5

Arg Pro Pro Gly Phe Ser Pro Phe Arg Cys
1 5 10

<210> 6
<211> 18
<212> DNA
<213> artificial

<220>
<223> oligonucleotide of oligo-peptide 1

<400> 6
cacgtacagg atgtacag 18

<210> 7
<211> 17
<212> DNA
<213> artificial

<220>
<223> oligonucleotide of oligo-peptide 2

<400> 7
tcgagaggta catcgtg 17

<210> 8

<211> 17
<212> DNA
<213> artificial

<220>
<223> oligonucleotide of oligo-peptide 3

<400> 8
aagcggtacg agtagca 17

<210> 9
<211> 25
<212> DNA
<213> artificial

<220>

<223> oligonucleotide of oligo-peptide Forward

<400> 9
gacgtaccag gcggtagctg cattt 25

<210> 10
<211> 25
<212> DNA
<213> artificial

<220>
<223> oligonucleotide of oligo-peptide Reverse

<400> 10
caggactccc caaaggagaa cagga 25

<210> 11
<211> 13
<212> PRT
<213> artificial

<220>
<223> Tag peptide of antibody-peptide conjugate

<400> 11

Asp Ser Pro Glu Gly Leu Asn Arg Lys Gln Lys Pro Ala 13
1 5 10

Electronic Acknowledgement Receipt

EFS ID:	40411875
Application Number:	16272487
International Application Number:	
Confirmation Number:	6864
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Customer Number:	141429
Filer:	Robert E. Powers/Sylvana Householder
Filer Authorized By:	Robert E. Powers
Attorney Docket Number:	NATE-037001US 321329-2523
Receipt Date:	28-AUG-2020
Filing Date:	11-FEB-2019
Time Stamp:	13:35:12
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	NATE-037-001US_IDSTransmittal.pdf	111361 7c1996af91cd8d1ce7e45fa329e28fa316a13e4e	no	3

Warnings:

Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	NATE-037-001US_sb08.pdf	221600 d9c32d857ed8b99d4303f5bc8aadfdf8d6747680	no	1
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
3	Foreign Reference	JP2008542783A.pdf	8070734 016e60d67f129d43285989619405920ab92a2336	no	46
Warnings:					
Information:					
4	Foreign Reference	WO2007000669A2.pdf	2925132 3844f448200669 81ff161b868f358e5d1d07f659	no	65
Warnings:					
Information:					
		Total Files Size (in bytes):	11328827		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	F. W. M. Lu

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

VIA EFS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT
UNDER 37 C.F.R. §§ 1.56, 1.97, AND 1.98

In accordance with the duty of disclosure set forth in 37 C.F.R. §1.56, Applicant hereby submits the following information in conformance with 37 C.F.R. §§1.97 and 1.98. It is respectfully requested that the information be expressly considered during the prosecution of this application, and the references be made of record therein and appear among the "References Cited" on any patent to issue therefrom.

[X] Pursuant to 37 C.F.R. §1.98, a copy of each non-US patent document cited in the attached Form used in lieu of PTO/SB/08 is enclosed.

[] No copies of the publications listed on the attached Form used in lieu of PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98(d) because the publications were previously cited by or submitted to the Office in prior Application Serial No(s). ____ to which the above-identified application claims priority under 35 U.S.C. §120.

[X] No copies of any U.S. patents or U.S. patent application publications listed on the attached Form used in lieu of PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98.

[] Publication(s) listed on the attached Form used in lieu of PTO/SB/08 were cited in a foreign search or examination report corresponding to application serial no. and mailed on .

[] Enclosed is a copy of a non-English publication(s) _____ Pursuant to §609 of the M.P.E.P., Applicant submits the attached foreign search or examination report, which cites such non-English language publication(s).

[] Enclosed is a copy of a non-English publication(s) _____ English language publication _____ (copy enclosed) claims priority from this non-English publication.

[] Enclosed is an explanation of non-English publication(s) _____ for which an English translation is not available.

[] Enclosed is an English translation of non-English publication(s) _____ cited in the attached Form used in lieu of PTO/SB/08.

[] Enclosed is a copy of pending patent Application Serial No. _____.

[X] In accordance with **37 C.F.R. §1.97(b)**, no additional fee for submission of this Information Disclosure Statement is required, as it is filed within any one of the following time periods:

[] within three months from the filing date of this national application other than a CPA under 37 C.F.R. § 1.53(d);

[] within three months from the date of entry of the national stage as set forth in 37 C.F.R. §1.491 in this international application;

[X] before the mailing date of a first office action on the merits; or

[] before the mailing of a first office action after the filing of a request for continued examination under 37 C.F.R. § 1.114.

In accordance with 37 C.F.R. § 1.97(g), this Information Disclosure Statement shall not be construed as to mean that a search has been made.

In accordance with 37 C.F.R. § 1.97(h), the filing of this Information Disclosure Statement shall not be construed to be an admission that the information cited in the statement is, or is considered to be material to patentability as defined by 37 C.F.R § 1.56(b).

Remarks

It is respectfully requested that the Examiner consider the above-noted information and return an initialed copy of the attached Form used in lieu of PTO/SB/08 to the undersigned.

The Examiner is invited to contact the undersigned by telephone if it is felt that a telephone interview would advance the prosecution of the present application.

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 50-1283, under Order No. NATE-037/001US 321329-2523 from which the undersigned is authorized to draw.

Dated: August 28, 2020

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (202) 842-7899

By: /Robert E. Powers/
Robert E. Powers, Ph.D.
Reg. No. 77,802
Matthew Pavao, Ph.D.
Reg. No. 50,572

232841577 v1

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/272,487	02/11/2019	Joseph M. Beechem	NATE-037001US 321329-2523	6864

141429 7590 09/09/2020

Cooley LLP/NanoString Technologies, Inc
1299 Pennsylvania Avenue NW
Suite 700
Washington, DC 20004

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
09/09/2020	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zIPPatentDocketingMailboxUS@cooley.com

Office Action Summary	Application No. 16/272,487	Applicant(s) Beechem et al.	
	Examiner FRANK W LU	Art Unit 1634	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 2 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1-28 is/are pending in the application.

5a) Of the above claim(s) _____ is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☐ Claim(s) _____ is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☑ Claim(s) 1-28 are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date _____.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

1. Restriction to one of the following inventions is required under 35 U.S.C. 121:

I. Claims 1-14, drawn to a method for spatially detecting at least one target analyte in at least one cell from a tissue sample, classified in CPC C12Q 1/6806.

II. Claims 15-28, drawn to a method for spatially detecting at least one target analyte in at least one cell from a tissue sample, classified in CPC C12Q 1/6806.

2. The inventions are independent or distinct, each from the other because:

Groups I and II are distinct and independent inventions in that they are directed to different methods that have different method steps. As a result, different and distinct searches will have to be performed. For example, the search required for Group I such as step (4) of claim 1 is not required for Group II while the search required for Group II such as step (4) of claim 15 is not required for Group I.

3. Group I or II contains claims directed to the following patentably distinct species:

(1) at least one of the flow cell binding sites comprises a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3 (claim 6 or 19)

(2) at least one of the flow cell binding sites comprises a P7 flow cell adapter sequence, wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4 (claim 7 or 20)

The species are independent or distinct because these species are directed to different flow cell binding sites which comprise different flow cell adapter sequences. In addition, these species are not obvious variants of each other based on the current record.

Applicant is required under 35 U.S.C. 121 to elect a single disclosed species, or a single grouping of patentably indistinct species, for prosecution on the merits to which the claims shall be restricted if no generic claim is finally held to be allowable. Currently, generic claims are claims 1-5 and 8-14 or 15-18 and 21-28.

There is a search and/or examination burden for the patentably distinct species as set forth above because the searches for species (1) and (2) are not overlap.

Applicant is advised that the reply to this requirement to be complete must include (i) an election of a species to be examined even though the requirement may be traversed (37 CFR 1.143) **and (ii) identification of the claims encompassing the elected species or grouping of patentably indistinct species**, including any claims subsequently added. An argument that a claim is allowable or that all claims are generic is considered nonresponsive unless accompanied by an election.

The election may be made with or without traverse. To preserve a right to petition, the election must be made with traverse. If the reply does not distinctly and specifically point out supposed errors in the election of species requirement, the election shall be treated as an election without traverse. Traversal must be presented at the time of election in order to be considered timely. Failure to timely traverse the requirement will result in the loss of right to petition under 37 CFR 1.144. If claims are added after the election, applicant must indicate which of these claims are readable on the elected species or grouping of patentably indistinct species.

Should applicant traverse on the ground that the species, or groupings of patentably indistinct species from which election is required, are not patentably distinct, applicant should submit evidence or identify such evidence now of record showing them to be obvious variants or clearly admit on the record that this is the case. In either instance, if the examiner finds one of

the species unpatentable over the prior art, the evidence or admission may be used in a rejection under 35 U.S.C. 103 or pre-AIA 35 U.S.C. 103(a) of the other species.

Upon the allowance of a generic claim, applicant will be entitled to consideration of claims to additional species which depend from or otherwise require all the limitations of an allowable generic claim as provided by 37 CFR 1.141.

4. Group I or II further contains claims directed to the following patentably distinct species:

(1) at least one of the amplification primer binding sites comprises an i7 sequence, wherein the i7 sequence comprises the sequence set forth in SEQ ID NO: 1 (claim 11 or 24)

(2) at least one of the amplification primer binding sites comprises an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2 (claim 12 or 25)

The species are independent or distinct because these species are directed to different amplification primer binding sites which comprise different nucleotide sequences. In addition, these species are not obvious variants of each other based on the current record.

Applicant is required under 35 U.S.C. 121 to elect a single disclosed species, or a single grouping of patentably indistinct species, for prosecution on the merits to which the claims shall be restricted if no generic claim is finally held to be allowable. Currently, generic claims are claims 1-10, 13, and 14 or 15-23 and 26-28.

There is a search and/or examination burden for the patentably distinct species as set forth above because the searches for species (3) and (4) are not overlap.

Applicant is advised that the reply to this requirement to be complete must include (i) an election of a species to be examined even though the requirement may be traversed (37 CFR 1.143) **and (ii) identification of the claims encompassing the elected species or grouping of patentably indistinct species**, including any claims subsequently added. An

argument that a claim is allowable or that all claims are generic is considered nonresponsive unless accompanied by an election.

The election may be made with or without traverse. To preserve a right to petition, the election must be made with traverse. If the reply does not distinctly and specifically point out supposed errors in the election of species requirement, the election shall be treated as an election without traverse. Traversal must be presented at the time of election in order to be considered timely. Failure to timely traverse the requirement will result in the loss of right to petition under 37 CFR 1.144. If claims are added after the election, applicant must indicate which of these claims are readable on the elected species or grouping of patentably indistinct species.

Should applicant traverse on the ground that the species, or groupings of patentably indistinct species from which election is required, are not patentably distinct, applicant should submit evidence or identify such evidence now of record showing them to be obvious variants or clearly admit on the record that this is the case. In either instance, if the examiner finds one of the species unpatentable over the prior art, the evidence or admission may be used in a rejection under 35 U.S.C. 103 or pre-AIA 35 U.S.C. 103(a) of the other species.

Upon the allowance of a generic claim, applicant will be entitled to consideration of claims to additional species which depend from or otherwise require all the limitations of an allowable generic claim as provided by 37 CFR 1.141.

Applicant is reminded that upon the cancellation of claims to a non-elected invention, the inventorship must be corrected in compliance with 37 CFR 1.48(a) if one or more of the currently named inventors is no longer an inventor of at least one claim remaining in the application. A request to correct inventorship under 37 CFR 1.48(a) must be accompanied by an

application data sheet in accordance with 37 CFR 1.76 that identifies each inventor by his or her legal name and by the processing fee required under 37 CFR 1.17(i).

5. Papers related to this application may be submitted to Group 1600 by facsimile transmission. Papers should be faxed to Group 1600 via the PTO Fax Center. The faxing of such papers must conform with the notices published in the Official Gazette, 1096 OG 30 (November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 (December 28, 1993)(See 37 CAR § 1.6(d)). The CM Fax Center number is (571)273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frank Lu, Ph.D., whose telephone number is (571)272-0746. The examiner can normally be reached on Monday-Friday from 9 A.M. to 5 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen, can be reached on (571)272-0731.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FRANK W LU/
Primary Examiner, Art Unit 1634
September 2, 2020

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

BIB DATA SHEET

CONFIRMATION NO. 6864

SERIAL NUMBER	FILING or 371(c) DATE RULE	CLASS	GROUP ART UNIT	ATTORNEY DOCKET NO.
16/272,487	02/11/2019	435	1634	NATE-037001US 321329-2523

APPLICANTS

NanoString Technologies, Inc., Seattle, WA

INVENTORS

Joseph M. Beechem, Eugene, OR;
Dae Kim, Bellevue, WA;
Margaret Hoang, Seattle, WA;
Mark Gregory, Seattle, WA;
Erin Piazza, Edmonds, WA;
Denise Zhou, Seattle, WA;

**** CONTINUING DATA ***************************

This appln claims benefit of 62/771,212 11/26/2018
and claims benefit of 62/629,180 02/12/2018

**** FOREIGN APPLICATIONS ***************************

**** IF REQUIRED, FOREIGN FILING LICENSE GRANTED ****

02/21/2019

Foreign Priority claimed ❑ Yes ☑ No 35 USC 119(a-d) conditions met ❑ Yes ❑ No Verified and Acknowledged /FRANK LU/ Examiner's Signature	❑ Met after Allowance Initials	STATE OR COUNTRY	SHEETS DRAWINGS	TOTAL CLAIMS	INDEPENDENT CLAIMS
		OR	34	28	2

ADDRESS

Cooley LLP/NanoString Technologies, Inc
1299 Pennsylvania Avenue NW
Suite 700
Washington, DC 20004
UNITED STATES

TITLE

BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

FILING FEE RECEIVED		
3080	FEES: Authority has been given in Paper No._____ to charge/credit DEPOSIT ACCOUNT No._____ for following:	❑ All Fees ❑ 1.16 Fees (Filing) ❑ 1.17 Fees (Processing Ext. of time) ❑ 1.18 Fees (Issue) ❑ Other _____ ❑ Credit

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	Not Yet Assigned

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

RESPONSE TO RESTRICTION REQUIREMENT

This paper is submitted in response to the Restriction Requirement mailed September 9, 2020 (hereafter "the Restriction Requirement") in the above-referenced application. Accordingly, this Response is timely filed on or before November 9, 2020.

In response to Restriction Requirement, Applicant hereby elects Group II, without traverse. Applicant submits that claims 15-28 read on the elected invention.

The Applicant reserves the right to pursue non-elected subject matter in one or more continuation or divisional applications.

The Restriction Requirement also required the restriction of species.

The Restriction Requirement required the election of either a flow cell binding site comprising a P5 flow cell adapter sequence comprising the sequence set forth in SEQ ID NO: 3 or a flow cell binding site comprising a P7 flow cell adapter sequence comprising the sequence set forth in SEQ ID NO: 4. Applicant hereby elects flow cell binding sites comprising a P5 flow

cell adapter sequence comprising the sequence set forth in SEQ ID NO: 3, with traverse. Applicant submits that claims 15-19 and 21-28 read on the elected species.

As would be appreciated by the skilled artisan, P5 flow cell adapters and P7 flow cell adapters are commonly used together to perform sequencing reactions. Accordingly, Applicant submits that it would not be a serious burden on the Office to search both P5 flow cell adapters and P7 flow cell adapters.

The Restriction Requirement also required the election of either an amplification primer binding site comprising an i7 sequence, wherein the i7 sequence comprises the sequence set forth in SEQ ID NO: 1 or an amplification primer binding site comprising an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2. Applicant hereby elects an amplification primer binding site comprising an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2, with traverse. Applicant submits that claims 15-23 and 25-28 read on the elected species.

As would be appreciated by the skilled artisan, i5 and i7 amplification primer binding sites are commonly used together to perform sequencing reactions. Accordingly, Applicant submits that it would not be a serious burden on the Office to search both i5 amplification primer binding sites and i7 amplification primer binding sites.

Applicant's election is made without prejudice. As noted by the Examiner, upon the allowance of a generic claim, Applicant will be entitled to consideration of claims to not more than a reasonable number of species in addition to the elected species, provided that all claims to each additional species are written in dependent form or otherwise include all the limitations of an allowed generic claim as provided by 37 C.F.R. § 1.146.

An action on the merits of all the claims and a Notice of Allowance thereof are respectfully requested. Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 50-1283, under Order No. NATE-037/001US 321329-2523 from which the undersigned is authorized to draw.

Dated: November 9, 2020

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (617) 937-2400

By: / Matthew Pavao /
Matthew Pavao, Ph.D.
Reg. No. 50,572
Robert E. Powers, Ph.D.
Reg. No. 77,802

Electronic Acknowledgement Receipt

EFS ID:	41070103
Application Number:	16272487
International Application Number:	
Confirmation Number:	6864
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Customer Number:	141429
Filer:	Matthew Pavao/Francesca Vandermaelen
Filer Authorized By:	Matthew Pavao
Attorney Docket Number:	NATE-037001US 321329-2523
Receipt Date:	09-NOV-2020
Filing Date:	11-FEB-2019
Time Stamp:	14:17:13
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Response to Election / Restriction Filed	NATE-037_001US_Response.pdf	110306 12b91cb2178a7cdd1474915c2ebc9a168f051e10	no	3

Warnings:

Information:	
Total Files Size (in bytes):	110306

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/272,487	02/11/2019	Joseph M. Beechem	NATE-037001US 321329-2523	6864

141429 7590 01/25/2021

Cooley LLP/NanoString Technologies, Inc
1299 Pennsylvania Avenue NW
Suite 700
Washington, DC 20004

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
01/25/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zIPPatentDocketingMailboxUS@cooley.com

Office Action Summary	Application No. 16/272,487	Applicant(s) Beechem et al.	
	Examiner FRANK W LU	Art Unit 1634	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 11/9/2020.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☐ This action is **FINAL.** 2b) ☑ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 1-28 is/are pending in the application.

5a) Of the above claim(s) 1-14,20 and 24 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 15-19,21-23 and 25-28 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☑ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 2/11/2019 is/are: a)☐ accepted or b)☑ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date 5/8/2019,10/24/2019 and 8/28/2020.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Election/Restrictions

1. Applicant's election without traverse of Group II, claims 15-28 and with traverse of species (1) (at least one of the flow cell binding sites comprises a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3, see claim 19) and species (4) (at least one of the amplification primer binding sites comprises an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2 , see claim 25) in the reply filed on November 9, 2020 is acknowledged. The traversal is on the ground(s) that "[A]s would be appreciated by the skilled artisan, P5 flow cell adapters and P7 flow cell adapters are commonly used together to perform sequencing reactions. Accordingly, Applicant submits that it would not be a serious burden on the Office to search both P5 flow cell adapters and P7 flow cell adapters" and "[A]s would be appreciated by the skilled artisan, i5 and i7 amplification primer binding sites are commonly used together to perform sequencing reactions. Accordingly, Applicant submits that it would not be a serious burden on the Office to search both i5 amplification primer binding sites and i7 amplification primer binding sites".

The above arguments have been fully considered and have not been found persuasive toward the withdrawal of the election of species nor persuasive toward the relaxation of same such that species (1) to (4) will be examined together. First, although claim 17 requires performing PCR using a first amplification primer comprising a first flow cell binding site and a second amplification primer comprising a second flow cell binding site, since claim 19 only requires at least one of the flow cell binding sites comprises a P5 flow cell adapter sequence wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3 while claim 20 only requires at least one of the flow cell binding sites comprises a P7 flow cell

adapter sequence wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4, claim 17 does not require to use the P5 flow cell adapter sequence and the P7 flow cell adapter sequence together and applicant argument "[A]s would be appreciated by the skilled artisan, P5 flow cell adapters and P7 flow cell adapters are commonly used together to perform sequencing reactions. Accordingly, Applicant submits that it would not be a serious burden on the Office to search both P5 flow cell adapters and P7 flow cell adapters" is incorrect. Second, although claim 15 requires that the identifier oligonucleotide comprises a first amplification primer binding site and a second amplification primer binding site and claim 17 requires that a first amplification primer comprises a nucleic acid sequence complementary to the second amplification primer binding site and a second amplification primer comprises a nucleic acid sequence complementary to the first amplification primer binding site, since claim 24 only requires that at least one of the amplification primer binding sites comprises an i7 sequence while claim 25 only requires that at least one of the amplification primer binding sites comprises an i5 sequence, claim 15 does not require to use the i5 sequence and the i7 sequence together and applicant's argument "[A]s would be appreciated by the skilled artisan, i5 and i7 amplification primer binding sites are commonly used together to perform sequencing reactions. Accordingly, Applicant submits that it would not be a serious burden on the Office to search both i5 amplification primer binding sites and i7 amplification primer binding sites" is incorrect. Third, since SEQ ID No. 3 in claim 19 is different from SEQ ID No. 4 in claim 20 while SEQ ID No. 1 in claim 24 is different from SEQ ID No. 2 in claim 25, the searches for SEQ ID No. 3 in claim 19 and SEQ ID No. 4 in claim 20 are not overlap and the searches for SEQ ID No. 1 in claim 24 and SEQ ID No. 2 in claim 24 are not overlap, there is a burden on the Office to search claims 19 and 20 together and search claims 24 and 25 together. Therefore, the requirement on election of

species is still deemed proper and is made FINAL. However, the examiner agrees to combine species (1) to (4) together when this instant application is allowable. Claims 15-19, 21-23, and 25-28 will be examined.

Drawings

2. Some words in Figures 21, 22, 25, 30, and 31 cannot be recognized. Applicant is required to submit new Figures 21, 22, 25, 30, and 31. No new matter may be introduced in the required drawing. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d).

Specification

3. The disclosure is objected to because of the following informalities: (1) Figure 26 contains a nucleotide sequence having more than 10 nucleotides. However, either Figure 26 does not have a SEQ ID No. for this nucleotide sequence or Brief Description of The Drawings related to Figure 26 does not provide a SEQ ID No. for this nucleotide sequence; and (2) since case 15/819,151 has been patented, applicant is required to update this information in paragraph [00216] of the specification.

Appropriate correction is required.

Claim Objections

4. Claim 16 is objected to because of the following informalities: (1) "amplifying the collected identifier oligonucleotide" should be "said amplifying the collected identifier

oligonucleotide"; and (2) "performing PCR" should be "performing PCR using the collected identifier oligonucleotide as a template".

5. Claim 17 is objected to because of the following informality: "performing PCR" in line 1 should be "said performing PCR".

6. Claim 17 or 26 is objected to because of the following informality: "which the identifier oligonucleotide was released" should be "which the identifier oligonucleotide is released".

Appropriate correction is required.

Claim Rejections - 35 USC § 112

7. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

8. Enablement

Claims 15-19, 21-23, and 25-28 are rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, as failing to comply with the enablement requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention.

In *In re Wands*, 858 F.2d 731,737, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988) the court considered the issue of enablement in molecular biology. The Court summarized eight factors to be considered in a determination of "undue experimentation". These factors include: (a) the quantity of experimentation necessary; (b) the amount of direction or guidance presented; (c) the presence or absence of working examples; (d) the nature of the invention; (e) the state of the prior art; (f) the relative skill of those in the art; (g) the predictability of the art; and (h) the breadth of the claims. The Court also stated that although the level of skill in molecular biology is high, results of experiments in molecular biology are unpredictable.

To begin, there is no direction or guidance to show that at least one target analyte in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23, and 25-28. While the relative skill in the art is very high (the Ph.D. degree with laboratory experience), there is no predictability whether at least one target analyte in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23, and 25-28.

Since the specification does not show that at least one cell from a tissue sample is a permeabilized cell, claim 15 does not require that at least one cell from a tissue sample is a permeabilized cell and the least one target analyte is not on the surface of at least one cell, and it is known that very large molecules like proteins, polysaccharides or nucleic acids, do not diffuse across the cell membranes at all (see page 2 of "membrane permeability"), if at least one cell from a tissue sample is an eukaryotic cell with a non-permeable membrane and at least one probe comprises a protein and a nucleic acid, the at least one probe cannot enter the cell such that at least one target analyte in at least one cell in a tissue sample cannot contact with the at least one probe, and the at least one probe stays on the surface of the tissue sample after the contacting

step. Since steps (2) to (5) still can be performed based on amplification of the identification oligonucleotide of the at least one probe stayed stays on the surface of the tissue sample, the at least one target analyte in the at least one cell in the tissue sample cannot be spatially detected by the released identifier oligonucleotide by sequencing the amplified products produced in step (4) of claim 15.

In view of above discussions, the skilled artisan will have no way to predict the experimental results. Accordingly, it is concluded that undue experimentation is required to make the invention as it is claimed. These undue experimentation at least includes to test whether at least one target analyte in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23, and 25-28.

Claim Rejections - 35 USC § 103

9. The following is a quotation of 35 U.S.C. 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims the examiner presumes that the subject matter of the various claims was commonly owned as of the effective filing date of the claimed invention(s) absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and effective filing dates of each claim that was not commonly owned as of the effective filing date of the later invention in order for the examiner to consider the applicability of 35 U.S.C. 102(b)(2)(C) for any potential 35 U.S.C. 102(a)(2) prior art against the later invention.

10. Claims 15-18, 22, and 26-28 are rejected under 35 U.S.C. 103 as being unpatentable over Beechem *et al.,* (US 2020/0040385 A1, priority date: July 17, 2015) in view of McCasky Feazel *et al.,* (US Patent No. 6,100,030, published on August 8, 2000).

Regarding claims 15, since claim 15 does not require that a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site have specific structures or specific sequences, Beechem *et al.,* teach (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide (ie., the signal oligonucleotide), wherein the identifier oligonucleotide comprises: a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; and (3) collecting the released identifier oligonucleotide and identifying the released identifier oligonucleotide, thereby spatially detecting at least one target analyte in at least one cell from a tissue sample as recited in claim 15, wherein the unique nucleic acid sequence which identifies the target analyte bound to the target binding domain comprises between about 5 nucleotides and about 25 nucleotides as recited in claim 22, the nucleic acid sequence comprising a unique molecular identifier comprises about 5 to about 20 nucleotides as recited in claim 27 and the target binding domain comprises between about 10 nucleotides to about 70 nucleotides as recited in claim 28 (see paragraphs [0003] to [0008], [0023], and [0084], Figure 1 and claims 1-53).

Beechem *et al.,* do not disclose (4) amplifying the collected identifier oligonucleotide and (5) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (4) as recited in claim 15 wherein said amplifying the collected identifier oligonucleotide comprises performing PCR as recited in claim 16, said performing PCR comprises a first amplification primer and second amplification primer, wherein the first amplification primer comprises: a first flow cell binding site; a first nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide is released; and a nucleic acid sequence complementary to the second amplification primer binding site; and wherein the second amplification primer comprises: a second flow cell binding site; a second nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide is released; and a nucleic acid sequence complementary to the first amplification primer binding site as recited in claim 17, at least one of the flow cell binding sites comprises a flow cell adapter sequence suitable for sequencing as recited in claim 18, and the at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotide is released comprises about 1 to about 20 nucleotides as recited in claim 26. However, Beechem *et al.,* teach that "[D]etecting comprises a polymerase reaction, a reverse transcriptase reaction, hybridization to an oligonucleotide microarray, mass spectrometry, hybridization to a fluorescent molecular beacon, a sequencing reaction, or nCounter® Molecular Barcodes" (see paragraph [0012]).

Regarding claims 15-18 and 26, since claim 17 does not require that a first flow cell binding site, a first nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide is released, a second flow cell binding site, and a second nucleic acid sequence which identifies the specific location of the tissue sample from

which the identifier oligonucleotide is released have specific structures or specific sequences and claim 18 does not require that a flow cell adapter has a specific structure or a specific sequence and McCasky Feazel *et al.,* teach a AFLP technique using one primer pair (see Figure 1A), McCasky Feazel *et al.,* disclose amplifying an identifier oligonucleotide and identifying the identifier oligonucleotide by sequencing an amplified products as recited in claim 15, said performing PCR comprises a first amplification primer and second amplification primer, wherein the first amplification primer comprises: a first flow cell binding site; a first nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide is released; and a nucleic acid sequence complementary to the second amplification primer binding site; and wherein the second amplification primer comprises: a second flow cell binding site; a second nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide is released; and a nucleic acid sequence complementary to the first amplification primer binding site as recited in claim 17, at least one of the flow cell binding sites comprises a flow cell adapter sequence suitable for sequencing as recited in claim 18, and the at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotide is released comprises about 1 to about 20 nucleotides as recited in claim 26 (see Figure 1, columns 5, 11, and 12).

Therefore, it would have been *prima facie* obvious to one having ordinary skill in the art at the time the invention was made to have performed the methods recited in claims 15-18 and 26 by amplifying the collected identifier oligonucleotide and identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (4) of claim 15 wherein said amplifying the collected identifier oligonucleotide comprises performing PCR, said

performing PCR comprises a first amplification primer and second amplification primer, wherein the first amplification primer comprises: a first flow cell binding site; a first nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide is released; and a nucleic acid sequence complementary to the second amplification primer binding site; and wherein the second amplification primer comprises: a second flow cell binding site; a second nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide is released; and a nucleic acid sequence complementary to the first amplification primer binding site, at least one of the flow cell binding sites comprises a flow cell adapter sequence suitable for sequencing, and the at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotide is released comprises about 1 to about 20 nucleotides in view of the prior arts of Beechem *et al.,* and McCasky Feazel *et al.*. One having ordinary skill in the art would have been motivated to do so because Beechem *et al.,* have shown that "[D]etecting comprises a polymerase reaction, a reverse transcriptase reaction, hybridization to an oligonucleotide microarray, mass spectrometry, hybridization to a fluorescent molecular beacon, a sequencing reaction, or nCounter® Molecular Barcodes" (see paragraph [0012]) while, since McCasky Feazel *et al.,* disclose amplifying an identifier oligonucleotide and identifying the identifier oligonucleotide by sequencing amplified products as recited in claim 15 and claims 16-18 and 26 (see above and Figure 1, columns 5, 11, and 12) and the simple substitution of one kind of detection method (ie., the detection method taught by Beechem *et al.,*) from another kind of detection method (ie., the detection method taught by McCasky Feazel *et al.,* including PCR and sequencing) during the process of detecting at least one target analyte in at least one cell in a tissue sample as recited in claim 15, in the absence of convincing evidence to the contrary, would

have been *prima facie* obvious to one having ordinary skill in the art at the time the invention was made since the detection method taught by Beechem *et al.,* and the detection method taught by McCasky Feazel *et al.,* including PCR and sequencing are used for the same purpose (ie., detecting at least one target analyte in at least one cell in a tissue sample). One having ordinary skill in the art at the time the invention was made would have a reasonable expectation of success to detect at least one target analyte in at least one cell in a tissue sample using the detection method taught by McCasky Feazel *et al.,* in view of the prior arts of Beechem *et al.,* and McCasky Feazel *et al.*.

Furthermore, the motivation to make the substitution cited above arises from the expectation that the prior art elements will perform their expected functions to achieve their expected results when combined for their common known purpose. Support for making the obviousness rejection comes from the M.P.E.P. at 2144.06, 2144.07 and 2144.09.

Also note that there is no invention involved in combining old elements is such a manner that these elements perform in combination the same function as set forth in the prior art without giving unobvious or unexpected results. *In re Rose* 220 F.2d. 459, 105 USPQ 237 (CCPA 1955).

11. Claims 21 and 23 are rejected under 35 U.S.C. 103 as being unpatentable over Beechem *et al.,* in view of McCasky Feazel *et al.,* as applied to claims 15-18, 22, and 26-28 above, and further in view of Van Eijk *et al.,* (US 2003/0175729 A1, published on September 18, 2003).

The teachings of Beechem *et al.,* and McCasky Feazel *et al.,* have been summarized previously, *supra*.

Beechem *et al.,* and McCasky Feazel *et al.,* do not disclose that at least one of the flow cell binding sites comprises about 15 to about 40 nucleotides as recited in claim 21 and at least

one of the amplification primer binding sites comprises about 10 to about 40 nucleotides as recited in claim 23.

Van Eijk *et al.,* teach that conventional AFLP-adapters and -primers are about 10-30 base pairs in length (see paragraph [0161]).

Therefore, it would have been *prima facie* obvious to one having ordinary skill in the art at the time the invention was made to have performed the methods recited in claims 21 and 23 using an EcoRI primer and a Mse I primer which have about 30 base pairs in length such that at least one of the flow cell binding sites comprises about 15 nucleotides and at least one of the amplification primer binding sites comprises about 10 nucleotides in view of the prior arts of Beechem *et al.,* McCasky Feazel *et al.,* and Van Eijk *et al.*. One having ordinary skill in the art would have been motivated to do so because Van Eijk *et al.,* have suggested that conventional AFLP-adapters and -primers are about 10-30 base pairs in length (see paragraph [0161]). One having ordinary skill in the art at the time the invention was made would have a reasonable expectation of success to perform the methods recited in claims 21 and 23 using an EcoRI primer and a Mse I primer which have about 30 base pairs in length such that at least one of the flow cell binding sites comprises about 15 nucleotides and at least one of the amplification primer binding sites comprises about 10 nucleotides in view of the prior arts of Beechem *et al.,* McCasky Feazel *et al.,* and Van Eijk *et al.*.

Conclusion

12. Note that Agasti *et al.,* (J. Am. Chem. Soc., 134, 18499-18502, 2012) or Weissleder *et al.,* (US 2017/0275669 A1, priority date: June 12, 2013) in IDS filed on May 8, 2019 can be used as a 103 type of prior art to reject claim 15.

13. No claim is allowed.

14. Papers related to this application may be submitted to Group 1600 by facsimile transmission. Papers should be faxed to Group 1600 via the PTO Fax Center. The faxing of such papers must conform with the notices published in the Official Gazette, 1096 OG 30 (November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 (December 28, 1993)(See 37 CAR § 1.6(d)). The CM Fax Center number is (571)273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frank Lu, Ph.D., whose telephone number is (571)272-0746. The examiner can normally be reached on Monday-Friday from 9 A.M. to 5 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen, can be reached on (571)272-0731.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FRANK W LU/
Primary Examiner, Art Unit 1634
January 15, 2021

Notice of References Cited	Application/Control No. 16/272,487	Applicant(s)/Patent Under Reexamination Beechem et al.	
	Examiner FRANK W LU	Art Unit 1634	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
*	A	US-6100030-A	08-2000	McCasky Feazel; Rhonda J.	C12Q1/6827	435/6.12
*	B	US-20030152952-A1	08-2003	Van Ness, Jeffrey	C12Q1/6827	435/6.14
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	"membrane permeability". Printed on 1/15/2021.
	V	
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/272,487	Beechem et al.
	Examiner	Art Unit
	FRANK W LU	1634

CPC - Searched*		
Symbol	Date	Examiner
See attached files	01/15/2021	/FL/

CPC Combination Sets - Searched*		
Symbol	Date	Examiner
See attached files	01/15/2021	/FL/

US Classification - Searched*			
Class	Subclass	Date	Examiner
435	6.1, 6.11, 91.1, 91.2, 183	01/15/2021	/FL/
436	94, 501	01/15/2021	/FL/
536	23.1, 24.3, 24.33, 25.3	01/15/2021	/FL/

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
STN and EAST (see attached files) and Inventor Searches	01/15/2021	/FL/
Search SEQ ID NO: 2 and 3	01/15/2021	/FL/

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner

/FRANK W LU/ Primary Examiner, Art Unit 1634	January 15, 2021

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L3	2	probe near5 target near binding near5 oligonucleotide same force same (release or releasing or cleave or cleaving) same (amplifying or amplify or amplification) same sequencing same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/01/15 23:29
L4	2	probe near5 target near binding near5 oligonucleotide same (release or releasing or cleave or cleaving) same (amplifying or amplify or amplification) same sequencing same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/01/15 23:37
L5	198	probe near10 oligonucleotide same (release or releasing or cleave or cleaving) same (amplifying or amplify or amplification) same sequencing same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/01/15 23:38
L6	0	probe near10 oligonucleotide same linker same uv same (release or releasing or cleave or cleaving) same (amplifying or amplify or amplification) same sequencing same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/01/15 23:39
L7	10	probe near10 oligonucleotide same linker same (release or releasing or cleave or cleaving) same (amplifying or amplify or amplification) same sequencing same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/01/15 23:39
L8	8	7 AND ((C12Q2531/125 OR C12Q2535/122 OR C12Q2521/501 OR C12Q2525/155 OR C12Q2525/307 OR C12Q2565/501 OR C12Q2521/319 OR C12Q2521/325 OR C12Q2521/331 OR C12Q2525/161 OR C12Q2525/179 OR C12Q2525/191 OR C12Q2537/143 OR C12Q2563/179 OR C12Q1/6869 OR C12Q1/6806 OR C12Q1/6809 OR C12Q1/6816 OR C12Q1/6874 OR C12Q1/6837 OR C12Q1/6855 OR C12Q1/6844 OR C12Q1/6848 OR C12Q1/6862 OR C12Q2525/117 OR C12Q2525/186 OR C12Q2525/301 OR	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/01/15 23:42

		C12Q2545/114 OR C12Q2525/204 OR C12Q2525/207 OR C12Q2531/137 OR C12Q2533/107 OR C12Q2563/131 OR C12Q1/6883 OR C12Q1/707 OR C12Q2535/00 OR C12Q2537/1373 OR C12Q2600/106 OR C12Q2600/154 OR C12Q2600/156 OR C12N15/1068).CPC.)				

1/15/2021 11:42:56 PM
C:\Users\flu\Documents\EAST\Workspaces\~auto.02042018.005450.wsp

```
(FILE 'HOME' ENTERED AT 23:44:25 ON 15 JAN 2021)

     FILE 'MEDLINE, CAPLUS, BIOSIS, SCISEARCH, EMBASE' ENTERED AT 23:44:56
ON
     15 JAN 2021
L1           412 S PROBE AND TARGET BINDING AND OLIGONUCLEOTIDE
L2       2982039 S FORCE
L3       4948779 S RELEASE OR RELEASING OR CLEAVE OR CLEAVING
L4       1263231 S AMPLIFYING OR AMPLIFY OR AMPLIFICATION
L5       1915108 S SEQUENCING
L6       8724855 S DETECT OR DETECTING OR DETECTION
L7             0 S L1 AND L2 AND L3 AND L4 AND L5 AND L6
L8           955 S TARGET BINDING AND OLIGONUCLEOTIDE
L9             0 S L2 AND L3 AND L4 AND L5 AND L6 AND L8
L10            0 S L3 AND L4 AND L5 AND L6 AND L8
L11        13010 S TARGET BINDING
L12       601836 S OLIGONUCLEOTIDE
L13            0 S L2 AND L3 AND L4 AND L5 AND L6 AND L11 AND L12
L14            0 S L3 AND L4 AND L5 AND L6 AND L11 AND L12
L15           58 S L3 AND L4 AND L5 AND L6 AND  L12
L16            1 S L2 AND L15
L17           46 DUP REM L15 (12 DUPLICATES REMOVED)
```

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	Not Yet Assigned
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1.	5,320,814	06-14-1994	Walt et al.	
	2.	5,543,838	08-06-1996	Hosier et al.	
	3.	5,635,352	06-03-1997	Urdea et al.	
	4.	6,207,392	03-27-2001	Weiss et al.	
	5.	6,429,027	08-06-2002	Chee et al.	
	6.	6,596,257	07-22-2003	Bryan	
	7.	6,602,661	08-05-2003	Knezevic et al.	
	8.	6,690,470	02-10-2004	Baer et al.	
	9.	6,942,968	09-13-2005	Dickinson et al.	
	10.	7,060,507	06-13-2006	Akeson et al.	
	11.	7,132,519	11-07-2006	Monforte et al.	
	12.	7,214,477	05-08-2007	Emmert-Buck	
	13.	7,402,399	07-22-2008	Mukherjeei et al.	
	14.	7,569,392	08-04-2009	Levy et al.	
	15.	7,648,828	01-19-2010	Chan-Hui et al.	
	16.	7,728,287	06-01-2010	Felton et al.	
	17.	8,003,312	08-23-2011	Krutzik et al.	
	18.	8,221,972	07-17-2012	Lemaire et al.	
	19.	8,309,306	11-13-2012	Nolan et al.	
	20.	8,349,574	01-08-2013	Bates et al.	
	21.	8,362,415	01-29-2013	Felton et al.	
	22.	8,486,623	07-16-2013	Monforte et al.	

Examiner Signature:	/FRANK W LU/	Date Considered	01/16/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	Not Yet Assigned
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	23.	8,637,650	01-28-2014	Cherkasov et al.	
	24.	8,753,824	06-17-2014	Papin et al.	
	25.	8,865,414	10-21-2014	Hennig et al.	
	26.	8,906,700	12-09-2014	Lim et al.	
	27.	9,046,477	06-02-2015	Emedcoles et al.	
	28.	9,228,948	01-05-2016	Emedcoles et al.	
	29.	9,297,762	03-29-2016	Emedcoles et al.	
	30.	9,304,084	04-05-2016	Emedcoles et al.	
	31.	9,376,678	06-28-2016	Gormley et al.	
	32.	2001/0053334	12-20-2001	Chen et al.	
	33.	2002/0028457	03-07-2002	Empedocles et al.	
	34.	2002/0034737	03-21-2002	Drmanac	
	35.	2002/0034827	03-21-2002	Singh et al.	
	36.	2002/0039728	04-04-2002	Kain et al.	
	37.	2002/0051971	05-02-2002	Stuelpnagel et al.	
	38.	2002/0068018	06-06-2002	Pepper et al.	
	39.	2002/0177141	11-28-2002	Chee et al.	
	40.	2002/0187515	12-12-2002	Chee et al.	
	41.	2003/0008323	01-09-2003	Ravkin et al.	
	42.	2003/0013091	01-16-2003	Dimitrov	
	43.	2003/0028981	02-13-2003	Chandler et al.	
	44.	2003/0186426	10-02-2003	Brewer et al.	

Examiner Signature:	/FRANK W LU/	Date Considered	01/16/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	Not Yet Assigned
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	45.	2004/0000519	01-01-2004	Yong et al.	
	46.	2005/0048498	03-03-2005	Woudenberg et al.	
	47.	2005/0233318	10-20-2005	Chee et al.	
	48.	2007/0166708	07-19-2007	Dimitrov et al.	
	49.	2008/0038725	02-14-2008	Luo et al.	
	50.	2008/0038734	02-14-2008	Sorge et al.	
	51.	2010/0015607	01-21-2010	Geiss et al.	
	52.	2010/0047924	02-25-2010	Webster et al.	
	53.	2010/0112710	05-06-2010	Geiss et al.	
	54.	2010/0151472	06-17-2010	Nolan et al.	
	55.	2010/0209913	08-19-2010	Endress et al.	
	56.	2010/0261026	10-14-2010	Ferree et al.	
	57.	2010/0262374	10-14-2010	Hwang et al.	
	58.	2011/0086774	04-14-2011	Dunaway et al.	
	59.	2011/0151451	06-23-2011	Lemaire et al.	
	60.	2011/0223613	09-15-2011	Gut	
	61.	2011/0245111	10-06-2011	Chee	
	62.	2013/0023433	01-24-2013	Luo et al.	
	63.	2013/0178372	07-11-2013	Geiss et al.	
	64.	2014/0120532	05-01-2014	Lee et al.	
	65.	2014/0120550	05-01-2014	Baranov	
	66.	2014/0121117	05-01-2014	Tanner	

Examiner Signature:	/FRANK W LU/	Date Considered	01/16/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	Not Yet Assigned
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	67.	2014/0371088	12-18-2014	Webster	
	68.	2015/0080233	03-19-2015	Bendall et al.	
	69.	2015/0099650	04-09-2015	Sood et al.	
	70.	2015/0132763	05-14-2015	Amorese et al.	
	71.	2015/0141297	05-21-2015	Lim et al.	
	72.	2015/0287578	10-08-2015	Bendall et al.	
	73.	2015/0329852	11-19-2015	Nolan	
	74.	2015/0368697	12-24-2015	Samusik et al.	
	75.	2016/0194701	07-07-2016	Beechem et al.	
	76.	2016/0362730	12-15-2016	Alexander et al.	
	77.	2017/0016053	01-19-2017	Beechem et al.	
	78.	2017/0016909	01-19-2017	Beechem et al.	
	79.	2017/0275669	09-28-2017	Weissleder et al.	
	80.	2018/0142286	05-24-2018	Dunaway et al.	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (*if known*)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	81.	WO 97/14028 A2	04-17-1997	Luminex Corporation		
	82.	WO 97/07245 A1	02-27-1997	Rabani, Ely		
	83.	WO 99/18434 A1	04-15-1999	Trustees of Tufts College		

Examiner Signature:	/FRANK W LU/	Date Considered	01/16/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	Not Yet Assigned
Attorney Docket Number	NATE-037/001US 321329-2523

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (*if known*)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	84.	WO 00/73777 A1	12-07-2000	Bioarray Solutions, LLC		
	85.	WO 01/00875 A2	01-04-2001	Motorola, Inc		
	86.	WO 2012/106385 A2	08-09-2012	Apprise Bio, Inc.		
	87.	WO 2013/122996 A1	08-22-2013	The Johns Hopkins University		
	88.	WO 2014/200767 A1	12-18-2014	The General Hospital Corporation		
	89.	WO 2015/128272 A2	09-03-2015	Ventana Medical Systems, Inc.		
	90.	WO 2017/015099 A1	01-26-2017	Nanostring Technologies, Inc.		

NON PATENT LITERATURE DOCUMENTS

Examiner's Initials	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title ofthe item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[2]
	91.	AGASTI, S. S. et al., "Photocleavable DNA Barcode - Antibody Conjugates Allow Sensitive andn Multiplexed Protein Analysis in Single Cells," J. Am. Chem. Soc., 134:18499-18502 (2012).	
	92.	ALFANO, R. R. et al., "Optical Sensing, Imaging, and Manipulation for Biological and Biomedical Applications," SPIE-The International Society for Optical Engineering, July 2000, Vol. 4082, Taiwan, 342 pages.	
	93.	ANGELO, M. et al., "Multiplexed ion beam imaging of human breast tumors," Nature Medicine, 20(4):436-442 (2014).	
	94.	CESANO, A. et al., "Abstract 1371: Spatially-resolved, multiplexed digital characterizatio nof protein distribution and abundance in FFPE tissue sections," AACR 107th Annual Meeting, April 16-20, 2016, 5 pages.	

The words in the abstract cannot be recognized.

Examiner Signature:	/FRANK W LU/	Date Considered	01/16/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	Not Yet Assigned
Attorney Docket Number	NATE-037/001US 321329-2523

NON PATENT LITERATURE DOCUMENTS

Examiner's Initials	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title ofthe item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[2]
	95.	CHOW, S. et al., "Measurement of MAP Kinase Activation by Flow Cytometry Using Phospho-Specific Antibodies to MEK and ERK: Potential for Pharmacodynamic Monitoring of Signal Transduction Inhibitors," Cytometry (Communications in Clinical Cytometry), 46:72-78 (2001).	
	96.	CROSSETTO, N. et al., "Spatially resolved transcriptomics and beyond," Nature Reviews Genetics, 16:57-66 (2015).	
	97.	DIERCK, K. et al., "Quantitative multiplexed profiling of cellular signaling networks using phosphotyrosine-specific DNA-tagged SH2 domains," Nature Methods, 3:737-744 (2006).	
	98.	DRUMMOND, E. S. et al., "Proteomic analysis of neurons microdissected from formalinfixed, paraffin-embedded Alzheimer's disease brain tissue," Scientific Reports, 5:15456, 8 pages. doi: 10.1038/srep15456.	
	99.	FERGUSON, J. A. et al., "High-Density Fiber-Optic DNA Random Microsphere Array," Analytical Chemistry, 72(22):5618-5624 (2000).	
	100.	FREI, A. P. et al., "Highly multiplexed simultaneous detection of RNAs and proteins in single cells," Nature Methods, 13:269-275 (2016).	
	101.	GEISS, G. K. et al., "Direct multiplexed measurement of gene expression with color-coded probe pairs," Nature Biotechnology, 26(3):317-326 (2008).	
	102.	GULLBERG, M. et al., "Cytokine detection by antibody-based proximity ligation," PNAS, 101(22):8420-8424 (2004).	
	103.	LIND, K. & KUBISTA, M., "Development and evaluation of three real-time immuno-PCR assemblages for quantification of PSA," Journal of Immunological Methods, 304:107-116 (2005).	
	104.	Poster #3434, Power Point Presentation, "Digital spatial profiling platform allows for spatially-resolved, high-plex quantification of mRNA distribution and abundance on FFPE and fresh frozen tissue sections," AACR Annual Meeting, April 14-18, 2018, 1 page.	

Examiner Signature:	/FRANK W LU/	Date Considered	01/16/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	Not Yet Assigned
Attorney Docket Number	NATE-037/001US 321329-2523

NON PATENT LITERATURE DOCUMENTS

Examiner's Initials	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title ofthe item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[2]
	105.	STEEMERS, F. J. et al., "Screening unlabeled DNA targets with randomly ordered fiber-optic gene arrays," Nature Biotechnology, 18:91-94 (2000).	
	106.	THIERY-LAVENANT, G. et al., "Targeted Multiplex Imaging Mass Spectrometry in Transmission Geometry for Subcellular Spatial Resolution," Journal of the American Society for Mass Spectrometry, 24:609-614 (2013).	
	107.	ULLAL, A. V. et al., "Cancer cell profiling by barcoding allows multiplexed protein analysis in fine needle aspirates," Science Translational Medicine, 6(219):219ra9, 22 pages. doi: 10.1126/scitranslmed.3007361.	
	108.	WERNER, J. H. et al., "Current status of DNA sequencing by single molecule detection, Proc. SPIE 3602, Advances in Fluorescence Sensing Technology IV, pp. 355-366 (1999).	
	109.	ZINCHUK, V. et al., "Quantitative Colocalization Analysis of Multicolor Confocal Immunofluorescence Microscopy Images: Pushing Pixels to Explore Biological Phenomena," The Japan Society of Histochemistry and Cytochemistry, 40(4):101-111 (2007).	
	110.	ZOLLINGER, D. et al., "Abstract 3434: Digital spatial profiling platform allows for spatially resolved, high-plex quantification of mRNA distribution and abundance on FFPE and fresh frozen tissue sections," AACR Annual Meeting, April 14-18, 2018, 2 pages.	

203107908 v1

Examiner Signature:	/FRANK W LU/	Date Considered	01/16/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	F. W. M. Lu
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1.	10,501,777	12-10-2019	Beechem et al.	
	2.	10,640,816	05-05-2020	Beechem et al.	
	3.	2020/0040385	02-06-2020	Beechem et al.	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (*if known*)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	4.	JP 2008-542783 A	11-27-2008	Universite des Sciences et Technologie		Corresponds to WO 2007/000669 A2
	5.	WO 2007/000669 A2	01-04-2007	Centre National de la Recherche Scientifique (CNRS) et al.		

232841174 v1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

Examiner Signature:	/FRANK W LU/	Date Considered	01/14/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	F. W. M. Lu
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1.	2001/0002315	05-31-2001	Schultz et al.	
	2.	2001/0007775	07-12-2001	Seul et al.	
	3.	2001/0023078	09-20-2001	Bawendi et al.	
	4.	2001/0029049	10-11-2001	Walt et al.	
	5.	2001/0034034	10-25-2001	Bruchez et al.	

213808694 v1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

Examiner Signature:	/FRANK W LU/	Date Considered	01/14/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

DocCode – SCORE

SCORE Placeholder Sheet for IFW Content

Application Number: 16272487 Document Date: 07/26/2021

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Drawing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via DAV or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: March 1, 2019

PTO/SB/06 (09-11)
Approved for use through 1/31/2014. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD Substitute for Form PTO-875	Application or Docket Number 16/272,487	Filing Date 02/11/2019	☑ To be Mailed

ENTITY: ☑ LARGE ☐ SMALL ☐ MICRO

APPLICATION AS FILED - PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA		RATE ($)	FEE ($)
☐ BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A		N/A	
☐ SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A		N/A	
☐ EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A		N/A	
TOTAL CLAIMS (37 CFR 1.16(i))	minus 20 =	*		x $100 =	
INDEPENDENT CLAIMS (37 CFR 1.16(h))	minus 3 =	*		x $460 =	
☐ APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).				
☐ MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))					
* If the difference in column 1 is less than zero, enter "0" in column 2.				TOTAL	

APPLICATION AS AMENDED - PART II

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT	07/26/2021	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	* 20	Minus	** 28	= 0		x $100 =	0
	Independent (37 CFR 1.16(h))	* 1	Minus	*** 3	= 0		x $480 =	0
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	0

		(Column 1)		(Column 2)	(Column 3)			
AMENDMENT		CLAIMS REMAINING AFTER AMENDMENT		HIGHEST NUMBER PREVIOUSLY PAID FOR	PRESENT EXTRA		RATE ($)	ADDITIONAL FEE ($)
	Total (37 CFR 1.16(i))	*	Minus	**	=		x $0 =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=		x $0 =	
	☐ Application Size Fee (37 CFR 1.16(s))							
	☐ FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							
							TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.	LIE
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".	/ANGELA D JOHNSON/
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".	
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.	

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

DocCode – SEQ.TXT

SCORE Placeholder Sheet for IFW Content

Application Number: 16272487 Document Date: 07/26/2021

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Sequence Listing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via DAV or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: March 1, 2019

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	F. W. Lu

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

AMENDMENT/RESPONSE TO OFFICE ACTION

This paper is filed in response to the Non-final Office Action mailed January 25, 2021 in the above-identified application. Applicant hereby petitions for a three-month extension of time and submits the appropriate fee herewith. With the three-month extension of time, this paper is timely filed on or before July 26, 2021. Applicant submits the following Amendments and Remarks, and respectfully requests reconsideration of the application in view thereof.

Any extensions of time necessary to prevent abandonment of this application are hereby petitioned for under 37 C.F.R. §1.136(a), and any additional fees required (including fees for net addition of claims) are hereby authorized to be charged to our Deposit Account No. 50-1283.

Amendments to the Specification begin on page 2 of this paper.

Amendments to the Drawings begin on page 3 of this paper.

Amendments to the Claims appear on page 4 of this paper.

Remarks/Arguments begin on page 8 of this paper.

AMENDMENTS TO THE SPECIFICATION:

Please replace the copy of the Specification as filed on February 11, 2019 with the Substitute Specification submitted herewith. Applicant is submitting a marked-up copy and a clean copy of the Substitute Specification. No new matter is added.

APPENDIX A: Substitute Specification – Marked-up copy

APPENDIX B: Substitute Specification – Clean copy

AMENDMENTS TO THE DRAWINGS:

The attached eleven (11) sheets of drawings are replacement drawing sheets in compliance with 37 C.F.R. 1.84 and 1.121(d) for originally filed Figures 21, 22, 25, 30 and 31. Please replace Figures 21, 22, 25 30 and 31 filed on February 11, 2019 with the replacement drawing sheets attached hereto. No new matter is added.

APPENDIX C: Replacement Drawing Sheets

AMENDMENTS TO THE CLAIMS:

Set forth below in ascending order, with status identifiers, is a complete listing of all claims currently under examination. Changes to any amended claims are indicated by [[double brackets]], ~~strikethrough~~ and/or underlining. This listing also reflects any cancellation and/or addition of claims.

1. – 14. (Canceled)

15. (Original) A method for spatially detecting the abundance of at least one target analyte in at least one cell from a tissue sample comprising:

(1) contacting at least one target analyte in at least one cell in a tissue sample with at least ~~one probe~~three probes, thereby tiling the probes on the at least one target analyte,

wherein each of the probes ~~comprising~~comprise a target binding domain that binds to the at least one target analyte and an identifier oligonucleotide,

wherein the identifier oligonucleotide comprises:

a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain,

a nucleic acid sequence comprising a unique molecular identifier,

a first amplification primer binding site, and

a second amplification primer binding site,

wherein the tissue sample has been treated to facilitate binding of the probes to the at least one target analyte;

(2) providing a force to a location of the tissue sample sufficient to release the identifier ~~oligonucleotide~~oligonucleotides;

(3) collecting the released identifier ~~oligonucleotide~~oligonucleotides;

(4) amplifying the collected identifier ~~oligonucleotide~~oligonucleotides;

(5) identifying the released identifier ~~oligonucleotide~~oligonucleotides by sequencing the amplified products produced in step (4), thereby spatially detecting the abundance of the at least one target analyte in the at least one cell in a tissue sample.

16. (Currently Amended) The method of claim 15, wherein said amplifying the collected identifier ~~oligonucleotide~~oligonucleotides comprises performing PCR using the collected identifier oligonucleotides as templates.

17. (Currently Amended) The method of claim 16, wherein said performing PCR comprises a first amplification primer and second amplification primer,

wherein the first amplification primer comprises:

a first flow cell binding site;

a first nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier ~~oligonucleotide~~oligonucleotides ~~was~~are released; and

a nucleic acid sequence complementary to the second amplification primer binding site; and

wherein the second amplification primer comprises:

a second flow cell binding site;

a second nucleic acid sequence which identifies the specific location of the sample from which the identifier ~~oligonucleotide~~oligonucleotides ~~was~~are released; and

a nucleic acid sequence complementary to the first amplification primer binding site.

18. (Original) The method of claim 17, wherein at least one of the flow cell binding sites comprises a flow cell adapter sequence suitable for sequencing.

19. (Original) The method of claim 17, wherein at least one of the flow cell binding sites comprises a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3.

20. (Withdrawn) The method of claim 17, wherein at least one of the flow cell binding sites comprises a P7 flow cell adapter sequence, wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4.

21. (Original) The method of claim 17, wherein at least one of the flow cell binding sites comprises about 15 to about 40 nucleotides

22. (Original) The method of claim 15, wherein the unique nucleic acid sequence which identifies the target analyte bound to the target binding domain comprises between about 5 nucleotides and about 25 nucleotides

23. (Original) The method of claim 15, wherein at least one of the amplification primer binding sites comprises about 10 to about 40 nucleotides.

24. (Withdrawn) The method of claim 17, wherein at least one of the amplification primer binding sites comprises an i7 sequence, wherein the i7 sequence comprises the sequence set forth in SEQ ID NO: 1.

25. (Original) The method of claim 17, wherein at least one of the amplification primer binding sites comprises an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2.

26. (Currently Amended) The method of claim 17, wherein at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier ~~oligonucleotide~~oligonucleotides ~~was~~are released comprises about 1 to about 20 nucleotides.

27. (Original) The method of claim 15, wherein the nucleic acid sequence comprising a unique molecular identifier comprises about 5 to about 20 nucleotides.

28.(Original) The method of claim 15, wherein the target binding domain comprises between about 10 nucleotides to about 70 nucleotides.

29. (New) The method of claim 15, wherein each of the probes comprise a photo-cleavable linker between the identifier oligonucleotide and the target binding domain.

30. (New) The method of claim 29, wherein providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotides comprises exciting a location of the tissue sample with light of sufficient wavelength capable of cleaving the photo-cleavable linker, thereby releasing the identifier oligonucleotides.

31. (New) The method of claim 22, wherein the unique nucleic acid sequence which identifies the target analyte bound to the target binding domain comprises about 12 nucleotides.

32. (New) The method of claim 26, wherein at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotides are released comprises about 8 nucleotides.

33. (New) The method of claim 27, wherein the nucleic acid sequence comprising a unique molecular identifier comprises about 14 nucleotides.

34. (New) The method of claim 28, wherein the target binding domain comprises between about 35 to about 50 nucleotides.

REMARKS

Upon entry of these amendments, 15-34 are pending. Claims 1-14 are canceled herein without prejudice or disclaimer as being drawing to non-elected subject matter. Applicant reserves the right to pursue the subject matter of the canceled claims in one or more continuation or divisional applications. Claims 20 and 24 are withdrawn by the Office. Claims 15-17 and 26 are amended herein and claims 29-34 are added herein. Support for these amendments and new claims can be found throughout the specification and claims, as originally filed. No new matter is added.

Objections to the Drawings

The Office asserts that some words in Figures 21, 22, 25, 30 and 31 cannot be recognized and that Applicant is required to submit new Figures 21, 22, 25, 30 and 31. Applicant herewith submits replacement sheets for Figures 21, 22, 25, 30 and 31 as required by the Office. For clarity purposes, Applicant submits that Figure 21 has been divided into Figures 21A, 21B, 21C and 21D. Additionally, for clarity purposes, Applicant submits that Figure 22 has been divided into Figures 22A, 22B, 22C and 22D. No new matter has been added. Applicant submits that with submission of the replacement sheets, this objection is moot.

Objections to the Specification

The Office objects to the Specification because Figure 26 contains a nucleotide sequence having more than 10 nucleotides, but Figure 26 does not have a SEQ ID No. for this nucleotide sequence or the Brief Description of the Drawings related to Figure 26 does not provide a SEQ ID No. for this nucleotide sequence. Applicant submits herewith a substitute specification with an amended brief description of Figure 26 that explains that the nucleic acid sequence shown in Figure 26 corresponds to SEQ ID NO: 175. Applicant also submits herewith an updated Sequence Listing. No new matter has been added. Applicant submits that with the submission of the substitute specification and the updated sequence listing, this objection is moot.

The Office objects to the Specification because case 15/819,151, which is referenced in paragraph [00216] of the specification, has been patented. The Office asserts that the Applicant is required to update this information in paragraph [00216]. Applicant submits herewith a substitute specification with an amended [00216] that indicates that case 15/819,151 was issued as US Patent No. 10,415,080. No new matter has been added. Applicant submits that with the submission of the substitute specification, this objection is moot.

Objections to the Claims

The Office objects to claim 16 for alleged informalities. Without acquiescing as to the validity of the objection, and solely in an effort to expedite prosecution, claim 16 is amended herein according to the suggestion of the Office to recite "said amplifying the collected identifier oligonucleotides" and "performing PCR using the collected identifier oligonucleotides as templates". Applicant submits that these amendments obviate this objection.

The Office objects to claim 17 for alleged informalities. Without acquiescing as to the validity of the objection, and solely in an effort to expedite prosecution, claim 17 is amended herein according to the suggestion of the Office to recite "said performing PCR". Applicant submits that this amendment obviates this objection.

The Office objects to claims 17 and 26 for alleged informalities. Without acquiescing as to the validity of the objection, and solely in an effort to expedite prosecution, claims 17 and 26 are amended herein according to the suggestion of the Office to recite "which the identifier oligonucleotides are released". Applicant submits that these amendments obviate this objection.

Claim Rejections – 35 U.S.C. §112

Claims 15-19, 21-23 and 25-28 are rejected under 35 U.S.C. §112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, as allegedly failing to comply with the enablement requirement. The Office asserts that there is no direction or guidance to show that at least one target analyte in at least one cell form a tissue sample can be spatially detected using the methods recited in claims

15-19, 21-23 and 25-28. Specifically, the Office asserts that there would be undue experimentation to test whether at least one target analyte in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23 and 25-28. Applicant respectfully disagrees and traverses the rejection with respect to the pending claims, as amended herein.

The test of enablement is whether one reasonably skilled in the art could make or use the invention from the disclosures in the patent coupled with information known in the art without undue experimentation.[1] Importantly, a patent need not teach, and preferably omits, what is well known in the art.[2] Furthermore, the fact that experimentation may be complex does not necessarily make it undue, if the art typically engages in such experimentation.[3]

In order to make an enablement rejection, the Office has the initial burden to establish a reasonable basis to question the enablement provided for the claimed invention.[4] A specification disclosure which contains a teaching of the manner and process of making and using an invention in terms which correspond in scope to those used in describing and defining the subject matter sought to be patented must be taken as being in compliance with the enablement requirement, unless there is a reason to doubt the objective truth of the statements contained therein which must be relied on for enabling support.[5]

Applicant submits that the Office has failed to meet their initial burden of establishing a reasonable basis to question the enablement for the claimed invention and that the specification clearly and unambiguously teaches the skilled artisan how to make and use the claimed invention. Moreover, Applicant submits that, guided by the teachings of the specification, the skilled artisan would not be forced to undertake any undue experimentation.

---

[1] *see* MPEP §2164.01 citing *United States v. Telectronics, Inc.,* 857 F.2d 778, 785, 8 USPQ2d 1217, 1223 (Fed. Cir. 1988)
[2] *see* MPEP §2164.01 citing *In re Buchner,* 929 F.2d 660, 661, 18 USPQ2d 1331, 1332 (Fed. Cir. 1991); *Hybritech, Inc. v. Monoclonal Antibodies, Inc.,* 802 F.2d 1367, 1384, 231 USPQ 81, 94 (Fed. Cir. 1986), *cert. denied,* 480 U.S. 947 (1987); and *Lindemann Maschinenfabrik GMBH v. American Hoist & Derrick Co.,* 730 F.2d 1452, 1463, 221 USPQ 481, 489 (Fed. Cir. 1984)
[3] *see* MPEP §2164.01 citing *In re Certain Limited-Charge Cell Culture Microcarriers,* 221 USPQ 1165, 1174 (Int'l Trade Comm'n 1983), *aff'd. sub nom.,Massachusetts Institute of Technologyv.A.B. Fortia,* 774 F.2d 1104, 227 USPQ 428 (Fed. Cir. 1985)
[4] *see* MPEP §2164.04
[5] *Id.*

As an initial matter, Applicant submits that the skill level of a person of ordinary skill in the art in the context of the instant application is high, at least that of a PhD in biology, if not higher.

Without acquiescing as to the validity of the rejection, and solely in an effort to expedite prosecution, claim 1 is amended herein to recite that the tissue sample has been treated to facilitate binding of the probes to the at least one target analyte. Such a step is described in the instant specification at least at paragraph [00266], which explains that there are reagents, for example buffers and other reagents, that are known in the art to be useful in facilitating binding of a protein and/or a nucleic acid in a sample. That is, Applicant submits that it would be well within the purview of a person of ordinary skill in the art (*i.e.* a person with at least a PhD in biology) to treat a tissue sample such that at least one probe can bind to at least one target analyte in at least one cell in a tissue sample, as such a step is a routine laboratory procedure that is commonly performed in both research and industrial labs within the field of the present disclosure.

Moreover, contrary to the assertions of the Office that there is no direction or guidance to show that at least one target analyte in at least one cell form a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23 and 25-28, the specification includes numerous examples that describe the detection of more than one target analyte in more than one cell in a tissue sample using the instantly claimed methods. More specifically, Example 11 of the instant specification (*see* paragraphs [00418]-[00425] and Figures 28-34) describes that the claimed method was used to detect the spatial abundance of 96 different target analytes in FFPE samples in 22 common cell lines. Applicant submits that these experimental results unambiguously demonstrate that the claimed methods can be used by a person of ordinary skill in the art to detect multiple target analytes in multiple cells in a tissue sample, and therefore there is no need for undue experimentation.

Thus, Applicant submits that the skilled artisan, guided both by the teachings of the specification, specifically those described *supra*, as well as by common knowledge known in the art, would readily be able to practice the instantly claimed method without any undue experimentation. For at least the reasons described *supra*, Applicant submits that the instant

claims comply with the enablement requirement of 35 USC §112(a) or 35 U.S.C. §112 (pre-AIA), first paragraph.

Reconsideration of the claims and withdrawal of the rejection are respectfully requested.

Claim Rejections – 35 USC §103

Claims 15-18, 22, and 26-28 are rejected under 35 U.S.C. §103 as allegedly being unpatentable over Beechem *et al.* (US 2020/0040385 A1; hereafter "Beechem"), in view of McCasky Feazel *et al.* (US Patent No. 6,100,030; hereafter :McCasky). Applicant respectfully disagrees and traverses the rejection with respect to the pending claims, as amended herein.

The consistent criterion for determination of obviousness is whether the prior art would suggest to one of ordinary skill in the art that the claimed process should be carried out and would have a reasonable likelihood of success.[6] The rationale to support a conclusion that a claim would have been *prima facie* obvious is that all the claimed elements were known in the prior art and one skilled in the art could have combined the elements as claimed by known methods with no change in their respective functions, and the combination yielded nothing more than predictable results to one of ordinary skill in the art.[7] The burden is on the Examiner to show that the references expressly or implicitly suggest all of the claim limitations[8]. It is important to note the claimed invention as a whole, including all claim limitations, must be considered.[9] Finally, evidence showing there was no reasonable expectation of success or the results are greater than expected support a conclusion of nonobviousness[10].

Without acquiescing as to the validity of the rejection, and solely in an effort to expedite prosecution, claim 15 is amended herein to recite "contacting at least one target analyte in at least one cell in a tissue sample with at least three probes, thereby tiling the probes on the at least one target analyte". Moreover, Applicant submits that claim 15 recites that the identifier oligonucleotides comprise both a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain and a unique molecular identifier.

---

6 In re Dow Chemical Co., 837 F.2d 469 (Fed. Cir.1988).
7 KSR, 550 U.S. at 416, 82 USPQ2d at 1395.
8 MPEP § 2142
9 MPEP § 2142.02 and § 2143.03
10 *In re Rinehart,* 531 F.2d 1048, 189 USPQ 143 (CCPA 1976).

Applicant submits that this specific combination of features is not taught or suggested by Beechem and McCasky, either alone or in combination.

In addition to not being taught or suggested by the cited art, Applicant submits that the tiling of multiple probes onto a target analyte and the use of identifier oligonucleotides comprising a unique molecular identifier in the instantly claimed method unexpectedly results in superior spatial detection of the abundance of target analytes in a tissue sample.

First, as explained in the instant specification at paragraph [00208], the term tiling is used to describe when more than one probe of the present disclosure is bound to a target analyte (*e.g.* as shown in Figure 27 of the instant specification). The specification further explains that by tiling multiple probes onto a target analyte, the target analyte will be individually detected multiple times, increasing the overall accuracy of the measurement. Finally, in the case where multiple probes are tiled onto a single target RNA, the specification explains that there may be instances where one of the probes is incorrectly detected too many times and/or another probe is detected too few times. However, because multiple probes have been tiled onto the target RNA, these outlier probes can be discarded during analysis if the signals from the other probes are detected at a similar level, thereby further increasing the accuracy of the measurement of RNA abundance.

Second, as explained in the instant specification at paragraph [00319], unique molecular identifiers are short nucleic acid sequences that are used to quantify and reduce quantitative bias caused by nucleic acid amplification prior to sequencing reactions.

The instantly claimed method uses the combination of tiling and unique molecular identifies to more accurately detect the spatial abundance of specific target analytes in a tissue sample, as is shown in Example 11 of the instant specification. In this example, the instantly claimed method was used to detect the spatial abundance of 96 different target analytes in FFPE samples in 22 common human cell lines. As described in Example 11, for most of the 96 different target analytes, there was more than one probe that hybridized to the target analyte, thereby creating a tiling effect on many of the target analytes. As shown in Figures 31, 33 and 34, the use of tiling allowed for robust detection of different target analytes. Moreover, as shown in Figures 31, 33 and 34, the tiling approach allowed for the easy and rapid identification of

outlier probes that can be discarded during analysis to increase the accuracy of the abundance measurement. Finally, because of the robust and accurate detection, Figures 31, 33 and 34 demonstrate that the spatial abundance of certain targets can be compared across different cell types. The ability to more accurately detect spatial abundance in tissue samples to the extent that different samples and/or different areas on a single sample can be compared is important in clinical applications directed towards ascertaining the boundaries of cancerous and non-cancerous portions of patient biopsies.

Applicant submits that claim 15, as amended herein, recites a unique combination of features not disclosed by the cited art, and that these features produce a method that yields unexpectedly superior results. Accordingly, claim 15, from which all other claims subject to the rejection properly depend, is nonobvious in view of Beechem and McCasky.

Reconsideration of the claims and withdrawal of the rejection are respectfully requested.

Claims 21 and 23 are rejected under 35 U.S.C. §103 as allegedly being unpatentable over Beechem, *et al.*, in view of McCasky Feazel *et al.*, as applied to claims 15-18, 22, and 26-28 above, and further in view of Van Eijk *et al.* (US 2003/0175729 A1; hereafter "Van Eijk").

Claim 15, from which claims 21 and 23 properly depend, is nonobvious in view of Beechem and McCasky for at least the reasons described *supra*. Van Eijk fails to cure the deficiencies of Beechem and McCasky. Accordingly, Applicant submits that the instant claims are nonobvious in view of Beechem, McCasky and Van Eijk, either alone or in combination.

Reconsideration of the claims and withdrawal of the rejection are respectfully requested.

CONCLUSION

In view of the foregoing, Applicant respectfully submits that no further impediments exist to the allowance of this application and, therefore, requests an indication of allowability. However, the Examiner is requested to call the undersigned if any questions or comments arise.

The Director is hereby authorized to charge any appropriate fees under 37 C.F.R. §§1.16, 1.17, and 1.21 that may be required by this paper, and to credit any overpayment, to Deposit Account No. 50-1283 **referencing Docket No.** NATE-037/001US 321329-2523**.**

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 50-1283, under Order No. NATE-037/001US 321329-2523 from which the undersigned is authorized to draw.

Dated: July 26, 2021

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (617) 937-2400

By: / Matthew Pavao /
Matthew Pavao, Ph.D.
Reg. No. 50,572
Robert E. Powers, Ph.D.
Reg. No. 77,802

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	F. W. Lu

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT IN SUPPORT OF SUBSTITUTE SPECIFICATION

Pursuant to 37 C.F.R. 1.121(b)(3) and 1.125, I hereby state that the content of the Substitute Specification submitted herewith and that of the specification filed on February 11, 2019, are the same. Accordingly, no new matter is added.

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 50-1283, under Order No. NATE-037/001US 321329-2523 from which the undersigned is authorized to draw.

Dated: July 26, 2021

Respectfully submitted,
COOLEY LLP

By: /Robert E. Powers/
Robert E. Powers, Ph.D.
Reg. No. 77,802
Matthew Pavao, Ph.D.
Reg. No. 50,572

USPTO CUSTOMER NO. 141429

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (202) 842-7899

253972217
253972217 v1

BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

BACKGROUND OF THE INVENTION

[0001] This application claims priority to, and the benefit of, U.S. Provisional Application No. 62/629,180, filed February 12, 2018 and U.S. Provisional Application No. 62/771,212, filed November 26, 2018. The contents of each of the aforementioned patent applications are incorporated herein by reference in their entireties.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on May 19, 2021, is named "NATE-037_001US_Sequence_Listing.txt" and is about 50,049 bytes in size.

BACKGROUND OF THE INVENTION

[0003] Standard immunohistochemical and *in situ* hybridization methods allow for simultaneous detection of, at most, six to ten protein or nucleic acid targets, with three to four targets being typical. There exists a need for probes, compositions, methods, and kits for simultaneous, multiplexed detection and quantification of protein and/or nucleic acid expression in a user-defined region of a tissue, user-defined cell, and/or user-defined subcellular structure within a cell. Furthermore, there is a need for such systems to be adaptable for use with existing sequencing technologies that are already available to a large number of end users.

SUMMARY OF THE INVENTION

[0004] The present disclosure relates to probes, compositions, methods, and kits for simultaneous, multiplexed, spatial detection and quantification of protein and/or nucleic acid expression in a user-defined region of a tissue, user-defined cell, and/or user-defined subcellular structure within a cell.

[0005] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, a second amplification primer binding site and optionally, a constant nucleic acid sequence to minimize ligation bias; (5) amplifying the ligation product produced in step (4); and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0006] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises: a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, a second amplification primer binding site and optionally, a constant nucleic acid sequence to minimize ligation bias, and wherein at least one of the first or second amplification primer binding sites identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; (5) amplifying the ligation product produced in step (4); and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0007] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises: a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, a second amplification primer binding site and optionally, a constant nucleic acid sequence to minimize ligation bias; (5) amplifying the extension product produced in step (4) using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprise a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0008] The nucleic acid adapter of step (4) can be a partially double-stranded nucleic acid molecule. A partially double-stranded nucleic acid adapter can comprise a double-stranded annealed region, a first single-stranded mismatched region and a second single-stranded mismatched region. The first single-stranded mismatched region and the second single stranded mismatched region can be present on opposing sides of the double-stranded annealed region.

[0009] The nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be present in the double-stranded annealed region of a partially double-stranded nucleic acid adapter.

[0010] The constant nucleic acid sequence to minimize ligation bias can be present in the double-stranded annealed region of a partially double-stranded nucleic acid adapter.

[0011] A unique molecular identifier can be present in at least one of the first or second single-stranded mismatched regions of a partially double-stranded nucleic acid adapter.

[0012] The first amplification primer binding site can be present in the first single-stranded mismatched region of a partially double-stranded nucleic acid adapter and the second amplification primer binding site can be present in the second single-stranded mismatched region of the same partially double-stranded nucleic acid adapter.

[0013] The methods of the present disclosure described in the preceding can further comprise prior to step (4), performing an end repair reaction. The methods can also further comprise prior to step (4), performing a tailing reaction to attach a single nucleotide overhang to the 3' ends of the identifier oligonucleotide. The methods can further comprise, prior to step (4), performing an end repair reaction and a tailing reaction to attach a single nucleotide overhang to the 3' ends of the identifier oligonucleotide. The tailing reaction and the end repair reaction can be performed sequentially or concurrently.

[0014] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a first amplification primer binding site, and a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier, a second amplification primer binding site and optionally, a constant nucleic acid sequence to minimize ligation bias; (5) amplifying the ligation product produced in step (4); and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0015] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises: a first amplification primer binding site and a unique nucleic acid sequence which identifies the target

analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises a nucleic acid sequence comprising a unique molecular identifier, a second amplification primer binding site, and optionally, a constant nucleic acid sequence to minimize ligation bias; (5) amplifying the extension product produced in step (4) using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprise a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0016] The nucleic acid adapter of step (4) can be a partially double-stranded nucleic acid molecule. A partially double-stranded nucleic acid adapter can comprise a double-stranded annealed region and a single-stranded mismatched region.

[0017] The nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be present in the double-stranded annealed region of a partially double-stranded nucleic acid adapter.

[0018] The constant nucleic acid sequence to minimize ligation bias can be present in the double-stranded annealed region of a partially double-stranded nucleic acid adapter. The constant nucleic acid sequence can also comprise a cleavable moiety. The cleavable moiety can be an enzymatically cleavable moiety. The enzymatically cleavable moiety can be a USER sequence.

[0019] A unique molecular identifier can be present in the single-stranded mismatched region of a partially double-stranded nucleic acid adapter.

[0020] The second amplification primer binding site can be present in the single-stranded mismatched region of a partially double-stranded nucleic acid adapter.

[0021] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a first

amplification primer binding site, and a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a single stranded nucleic acid template, wherein the nucleic acid template comprises a region complementary to the unique nucleic acid sequence of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier, a nucleic acid sequence complementary to a second amplification primer binding site and optionally, an affinity molecule; (5) extending the identifier oligonucleotide of step (4) to form an extension product complementary to the single stranded nucleic acid template, wherein the extension product comprises the identifier oligonucleotide, the nucleic acid sequence complementary to the unique molecular identifier, and the second amplification primer binding site; (6) amplifying the extension product produced in step (5) using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprise a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0022] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a first amplification primer binding site and a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a single stranded nucleic acid template, wherein the nucleic acid template comprises a region complementary to the unique nucleic acid sequence of the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier, a nucleic acid

sequence complementary to a second amplification primer binding site and optionally, an affinity molecule; (5) extending the identifier oligonucleotide of step (4) to form an extension product complementary to the single stranded nucleic acid template, wherein the extension product comprises the identifier oligonucleotide, the nucleic acid sequence complementary to the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, the nucleic acid sequence complementary to the unique molecular identifier and the second amplification primer binding site; (6) amplifying the extension product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0023] The single stranded nucleic acid template can further comprise an affinity molecule. In aspects in which the single stranded nucleic acid template comprises an affinity molecule, the methods of the present disclosure described in the preceding can further comprise an affinity purification step between steps (4) and (5).

[0024] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a first amplification primer binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier and a second amplification primer binding site, and wherein the first the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) performing nick repair such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying

the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0025] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide and a first amplification primer binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide and a second amplification primer binding site, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence comprising a unique molecular identifier, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) performing nick repair such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0026] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier

oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a first amplification primer binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier and a second amplification primer binding site, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping; (5) performing a gap extension and nick repair reaction such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0027] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide and a first amplification primer binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide and a second amplification primer binding site, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence comprising a unique molecular identifier, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein the first and the second nucleic acid probes

hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping; (5) performing a gap extension and nick repair reaction such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0028] In aspects in which the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released is located on the first nucleic acid probe, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be located 5' to the first amplification primer binding site.

[0029] In aspects in which a unique molecular identifier is located on the second nucleic acid probe, the unique molecular identifier can be located 3' to the second amplification primer binding site.

[0030] In aspects in which the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a unique molecular identifier are present in the first nucleic acid probe, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a unique molecular identifier can be located 5' to the first amplification primer binding site.

[0031] In aspects in which the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a unique molecular identifier are present in the second nucleic acid probe, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the unique molecular identifier can be located 3' to the second amplification primer binding site.

[0032] In aspects in which a unique molecular identifier is present in the first nucleic acid probe and the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released is present in the second nucleic acid probe, the unique molecular identifier can be located 5' to the first amplification primer binding site and the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be located 3' to the second amplification binding site.

[0033] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a first amplification primer binding site, a nucleic acid sequence comprising a first unique molecular identifier and a first flow cell binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a second unique molecular identifier, a nucleic acid sequence comprising a third unique molecular identifier and a second flow cell binding site, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) performing nick repair such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0034] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the

identifier oligonucleotide, a first amplification primer binding site, a nucleic acid sequence comprising a first unique molecular identifier and a first flow cell binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a second unique molecular identifier and a second flow cell binding site, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence comprising a third unique molecular identifier, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) performing nick repair such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0035] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a first amplification primer binding site, a nucleic acid sequence comprising a first unique molecular identifier and a first flow cell binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to the identifier oligonucleotide, a nucleic acid sequence comprising a second unique molecular identifier, a nucleic acid sequence comprising a third unique molecular identifier and a second flow cell binding site, and wherein the first and the second nucleic acid

probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping; (5) performing a gap extension and nick repair reaction such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0036] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to the identifier oligonucleotide, a first amplification primer binding site, a nucleic acid sequence comprising a first unique molecular identifier and a first flow cell binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to the identifier oligonucleotide, a nucleic acid sequence comprising a second unique molecular identifier and a second flow cell binding site, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence comprising a third unique molecular identifier, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping; (5) performing a gap extension and nick repair reaction such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0037] In aspects in which the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the first unique molecular identifier are present in the first nucleic acid probe, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the first unique molecular identifier can be located 5' to the first flow cell binding site.

[0038] In aspects in which the second and the third unique molecular identifiers are present in the second nucleic acid probe, the second and the third unique molecular identifiers can be located 3' to the second flow cell binding site.

[0039] In some aspects, the first unique molecular identifier can be present in the first nucleic acid probe and can be located 5' to the first flow cell binding site. In other aspects, the second unique molecular identifier can be present in the second nucleic acid probe and can be located 3' to the second flow cell binding site.

[0040] In some aspects, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the third unique molecular identifier can be present in the first nucleic acid probe and can be located 5' to the first flow cell binding site.

[0041] In some aspects, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the third unique molecular identifier can be present in the second nucleic acid probe and can be located 3' to the second flow cell binding site.

[0042] In some aspect, the third unique molecular identifier can be present in the first nucleic acid probe and can be located 5' to the first flow cell binding site. In this same aspect, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be present in the second nucleic acid probe and can be located 3' to the second flow cell binding site.

[0043] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding

domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) amplifying the released identifier oligonucleotide using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; (5) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0044] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a first amplification primer binding site and a second amplification primer binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) amplifying the released identifier oligonucleotide using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein at least one of the amplification primers comprises a nucleic acid sequence comprising a unique molecular identifier; and (5) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0045] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding

domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain and a capture probe binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a capture probe, wherein the capture probe comprises an affinity molecule and a region complementary to the capture probe binding site; and (5) Identifying the released identifier oligonucleotide by sequencing the amplified hybridized product produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0046] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a capture probe binding site and a multiplexing probe binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a capture probe and a multiplexing probe, wherein the capture probe comprises an affinity molecule and a region complementary to the capture probe binding site, and wherein the multiplexing probe comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a region complementary to the multiplexing probe binding site; and (5) Identifying the released identifier oligonucleotide by sequencing the hybridized product produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0047] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the

released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises: a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and wherein the second nucleic acid probe comprises: a nucleic acid complementary to a portion of the identifier oligonucleotide, and a second amplification primer binding site, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) ligating the hybridized first and second nucleic acid probes together; (6) amplifying the ligation product produced in step (5); and (7) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0048] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises: a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) amplifying the collected identifier oligonucleotide; (5) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0049] A In all methods of the present disclosure, the ligation process can be a nick ligation process. The nick ligation process can be a nick repair process.

[0050] In all methods of the present disclosure, the sequencing can be an enzyme free sequencing method.

[0051] In all methods of the present disclosure, the identifier oligonucleotide can be double-stranded. In aspects in which the identifier oligonucleotide is double-stranded, at least one of the two strands of the identifier oligonucleotide can comprise at least two separate nucleic acid

molecules. Alternatively, at least one 3' end of an identifier oligonucleotide can comprise a single nucleotide overhang.

[0052] In all methods of the present disclosure, the identifier oligonucleotide can be single-stranded.

[0053] In all methods of the present disclosure, the unique nucleic acid sequence which identifies the target analyte bound to a target binding domain can comprise between about 5 nucleotides and about 40 nucleotides preferably about 35 nucleotides, preferably still about 10 nucleotides.

[0054] In all methods of the present disclosure, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can comprise between about 6 nucleotides and about 15 nucleotides, preferably about 12 nucleotides, preferably still about 10 nucleotides.

[0055] In all methods of the present disclosure, at least one of a first nucleic acid probe or a second nucleic acid probe can comprise an affinity molecule. For example, at least one of a first nucleic acid probe or a second nucleic acid probe can comprise a biotin.

[0056] In all methods of the present disclosure, an amplification primer binding site can comprise between about 18 nucleotides and about 40 nucleotides, preferably about 32 nucleotides, preferably still about 25 nucleotides. An amplification primer binding site can comprise an i7 sequence, wherein the i7 sequence comprises the sequence set forth in SEQ ID NO: 1. An amplification primer binding site can comprise an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2.

[0057] In all methods of the present disclosure, an amplification primer can comprise a flow cell adapter sequence, wherein the flow cell adapter sequence is suitable for sequencing. An amplification primer can comprise a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3. An amplification primer can comprise a P7 flow cell adapter sequence, wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4.

[0058] In all methods of the present disclosure, a flow cell binding site can comprise a flow cell adapter sequence, wherein the flow cell adapter sequence is suitable for sequencing. A flow cell binding site can comprise a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3. A flow cell binding site can

comprise a P7 flow cell adapter sequence, wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4.

[0059] In all methods of the present invention, at least one of the amplification primers can comprise an affinity molecule. For example, at least one of the amplification primers cam comprise a biotin.

[0060] In all methods of the present disclosure, amplification can comprise performing PCR. Performing PCR can comprise an amplification primer.

[0061] An amplification primer can comprise a flow cell binding site. An amplification primer can comprise a nucleic sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released. An amplification primer can comprise a nucleic acid sequence complementary to an amplification primer binding site.

[0062] Any of the above aspects can be combined with any other aspect.

[0063] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. In the Specification, the singular forms also include the plural unless the context clearly dictates otherwise; as examples, the terms “a,” “an,” and “the” are understood to be singular or plural and the term “or” is understood to be inclusive. By way of example, “an element” means one or more element. Throughout the specification the word “comprising,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from the context, all numerical values provided herein are modified by the term “about.”

[0064] Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. The references cited herein are not admitted to be prior art to the claimed invention. In the case of conflict, the present Specification, including definitions, will control. In addition, the materials, methods, and examples are

illustrative only and are not intended to be limiting. Other features and advantages of the disclosure will be apparent from the following detailed description and claim.

BRIEF DESCRIPTION OF THE DRAWINGS

[0065] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

[0066] The above and further features will be more clearly appreciated from the following detailed description when taken in conjunction with the accompanying drawings.

[0067] Figure 1 is a schematic of a two-ended adapter ligation method of the present disclosure.

[0068] Figure 2 is a schematic of a one-ended adapter ligation method of the present disclosure.

[0069] Figure 3 is a schematic of a templated-primer extension method of the present disclosure.

[0070] Figure 4 is a schematic of a template-extended identifier oligonucleotide of the present disclosure.

[0071] Figure 5 is a schematic of a short probe hybridization method of the present disclosure.

[0072] Figure 6 is a schematic of a short probe hybridization method of the present disclosure.

[0073] Figure 7 is a schematic of a short probe hybridization method of the present disclosure.

[0074] Figure 8 is a schematic of a long probe hybridization method of the present disclosure.

[0075] Figure 9 is a schematic of a long probe hybridization method of the present disclosure.

[0076] Figure 10 is a schematic of a long probe hybridization method of the present disclosure.

[0077] Figure 11 is a schematic of a direct-PCR method of the present disclosure.

[0078] Figure 12 is a schematic of an enzyme free method of the present disclosure.

[0079] Figure 13 is a schematic of a multiplexed enzyme free method of the present disclosure.

[0080] Figure 14 is a schematic of a probe of the present disclosure indirectly binding to a target nucleic acid.

[0081] Figure 15 is a schematic of an identifier oligonucleotide-short nucleic acid probe complex of the present disclosure.

[0082] Figure 16 is a schematic of a short probe hybridization method of the present disclosure.

[0083] Figure 17 is a schematic of an identifier oligonucleotide-short nucleic acid probe complex of the present disclosure.

[0084] Figure 18 is a schematic of a short probe hybridization method of the present disclosure.

[0085] Figure 19 is a schematic of a direct-PCR method of the present disclosure.

[0086] Figure 20 is a schematic overview of the methods of the present disclosure.

[0087] Figure 21A, Figure 21B, Figure 21C and Figure 21D show the spatial detection of protein target analytes using the methods of the present disclosure.

[0088] Figure 22A, Figure 22B, Figure 22C and Figure 22D show the spatial detection of RNA target analytes using the methods of the present disclosure.

[0089] Figure 23 shows the spatial detection of protein target analytes using the methods of the present disclosure.

[0090] Figure 24 shows the spatial detection of RNA target analytes using the methods of the present disclosure.

[0091] Figure 25 shows the spatial detection of protein target analytes using the methods of the present disclosure.

[0092] Figure 26 is a schematic of a probe of the present disclosure. The nucleic acid sequence shown in Figure 26 corresponds to SEQ ID NO: 175.

[0093] Figure 27 shows the use of probe tiling in the methods of the present disclosure.

[0094] Figure 28 shows the regions of interest selected on a tissue microarray.

[0095] Figure 29 is a series of graphs showing the read depth achieved using the methods of the present disclosure.

[0096] Figure 30 is a series of graphs showing the spatial detection of RNA target analytes in negative control samples using the methods of the present disclosure.

[0097] Figure 31 is a series of graphs showing the spatial detection of RNA target analytes in a HEK293 sample (top panel) and a Jurkat cell sample (bottom panel) using the methods of the present disclosure.

[0098] Figure 32 is a series of graphs showing the spatial detection of RNA target analytes in sixteen FFPE samples using the methods of the present disclosure.

[0099] Figure 33 is a graph showing the spatial detection of RNA target analytes in a HEK293 sample using the methods of the present disclosure.

[00100] Figure 34 is a graph showing the spatial detection of RNA target analytes in a Jurkat cell sample using the methods of the present disclosure.

DETAILED DESCRIPTION OF THE INVENTION

[00101] The present disclosure is based in part on probes, compositions, methods, and kits for simultaneous, multiplexed spatial detection and quantification of protein and/or nucleic acid expression in a user-defined region of a tissue, user-defined cell, and/or user-defined subcellular structure within a cell using existing sequencing methods.

[00102] The present disclosure provides a comparison of the identity and abundance of target proteins and/or target nucleic acids present in a first region of interest (*e.g.*, tissue type, a cell (including normal and abnormal cells), and a subcellular structure within a cell) and the identity and abundance of target proteins and/or target nucleic acids present in a second region of interest. There is no pre-defined upper limit to the number of regions of interest and comparisons that can be made; the upper limit relates to the size of the region of interest relative the size of the sample. As examples, when a single cell represents a region of interest, then a section may have hundreds to thousands of regions of interest; however, if a tissue section includes only two cell types, then the section may have only two regions of interest (each including only one cell type).

[00103] The present disclosure provides a higher degree of multiplexing than is possible with standard immunohistochemical or *in situ* hybridization methods. Standard immunohistochemical methods allow for maximal simultaneous detection of six to ten protein targets, with three to four protein targets being more typical. Similarly, *in situ* hybridization methods are limited to simultaneous detection of fewer than ten nucleic acid targets. The present disclosure provides detection of large combinations of nucleic acid targets and/or protein targets from a defined region of a sample. The present disclosure provides an increase in objective measurements by digital quantification and increased reliability and consistency, thereby enabling comparison of results among multiple centers.

[00104] Various compositions and methods of the present disclosure are described in full detail herein.

[00105] In one aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "two-ended adapter ligation method".

[00106] A two ended-adapter ligation method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The probes and samples of the present disclosure are described in further detail herein. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain.

[00107] Following contacting the at least on target analyte with the at least one probe, a two-ended adapter ligation method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00108] Following release of the identifier oligonucleotide, a two-ended adapter ligation method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location, thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00109] Following collection of the released identifier oligonucleotide, a two-ended adapter ligation method can further comprise: (4) ligating to the released identifier oligonucleotide collected in step (3) at least one nucleic acid adapter.

[00110] The nucleic acid adapter can comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated

"ROI #1", the nucleic acid adapter would comprise a nucleic acid sequence that corresponds to "ROI #1".

[00111] The nucleic acid adapter can also comprise a unique molecular identifier.

[00112] The nucleic acid adapter can also comprise a first amplification primer binding site. In other aspects, the nucleic acid adapter can also comprise a second amplification primer binding site.

[00113] In some aspects, the nucleic acid adapter can also comprise a constant nucleic acid sequence to minimize ligation bias caused by differences in sequences of particular identifier oligonucleotides.

[00114] The nucleic acid adapter can be a partially double-stranded nucleic acid molecule. In aspects in which the nucleic acid adapter is partially double-stranded, the nucleic acid adapter comprises a double-stranded annealed region, a first single-stranded mismatched region, and a second single-stranded mismatched region. The first single-stranded mismatched region and the second single stranded mismatched region can be present on opposing sides of the double-stranded annealed region.

[00115] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released can be present in the double-stranded annealed region of the nucleic acid adapter.

[00116] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a constant nucleic acid sequence to minimize ligation bias, the constant nucleic acid sequence to minimize ligation bias can be present in the double-stranded annealed region of the nucleic acid adapter.

[00117] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a unique molecular identifier, the unique molecular identifier can be present in at least one of the first or second single-stranded mismatched regions of the nucleic acid adapter.

[00118] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a first and a second amplification primer binding site, the first amplification primer binding site can be present in the first single-stranded mismatched region of the nucleic acid adapter and the

second amplification primer binding site can be present in the second single-stranded mismatched region of the nucleic acid adapter.

[00119] After ligation of the at least one nucleic acid adapter, a two-ended adapter ligation method can further comprise: (5) amplifying the ligation product produced in step (4) using amplification primers that bind to the first and second amplification primer binding sites; and (6) identifying the released oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the sample.

[00120] A two-ended adapter ligation method of the present disclosure can further comprise, prior to step (4), performing an end repair reaction using methods known in the art. The method can also further comprise, prior to step (4), performing a tailing reaction to attach a single nucleotide overhang to the 3' ends of the identifier oligonucleotide using methods known in the art. In aspects, the end repair reaction and the tailing reaction can be performed sequentially or concurrently.

[00121] In preferred aspects of a two-ended adapter ligation method, a nucleic acid adapter is ligated to both ends of the released and collected identifier oligonucleotide.

[00122] In other aspects of a two-ended adapter ligation method, at least one of the amplification primers used in step (5) to amplify the ligation product produced in step (4) comprises a nucleic acid sequence which identifies the specific location of the tissue sample form which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated "ROI #1", at least one of the amplification primers would comprise a nucleic acid sequence that corresponds to "ROI #1".

[00123] Figure 1 shows a schematic of a preferred aspect of a two-ended adapter ligation method of the present disclosure. In this aspect, the probe comprises a target binding domain comprising an antibody that binds to a target protein. In the upper left panel, the probe binds to the target protein. In the upper right panel, a UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target protein bound to the target binding domain. In the bottom panel, a nucleic acid adapter is ligated to both ends of the identifier oligonucleotide. In this non-limiting example, the nucleic acid adapter is partially double-stranded and comprises a double-stranded annealed region, a first single-stranded mismatched region, and a second single-stranded mismatched region. Present in

the double-stranded annealed region is a constant nucleic acid sequence to minimize ligation bias and a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. Present in the first single-stranded mismatched region is a first amplification primer binding site. Present in the second single-stranded mismatched region is a unique molecular identifier and a second amplification primer binding site. Following ligation of the nucleic acid adapters to the identifier oligonucleotide, the product is amplified using amplification primers that bind the first and the second amplification primer binding sites and sequenced to identify the target protein bound by the probe.

[00124] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide dually ligated to two nucleic acid adapters for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide dually ligated to two nucleic acid adapters comprises an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which is capable of identifying a target analyte in a sample. Each end of the identifier oligonucleotide is attached to a nucleic acid adapter molecule, wherein the nucleic acid adapter molecule is partially double-stranded and comprises a double-stranded annealed region, a first single-stranded mismatched region and a second single-stranded mismatched region. The first single-stranded mismatched region and the second single stranded mismatched region are present on opposing sides of the double-stranded annealed region. The double-stranded mismatch region comprises a constant nucleic acid sequence to minimize ligation bias and a nucleic acid sequence nucleic acid sequence which is capable of identifying a specific location of a sample. The first single-stranded mismatched region comprises a first amplification primer binding site. The second single-stranded mismatched region comprises a second amplification primer binding site and a nucleic acid sequence comprising a unique molecular identifier. A schematic of an identifier oligonucleotide dually ligated to two nucleic acid adapters is shown in the bottom panel of Figure 1.

[00125] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "one-ended adapter ligation method".

[00126] A one-ended adapter ligation method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that

the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. The identifier oligonucleotide can also comprise a first amplification primer binding site. In some aspects, the identifier oligonucleotide also comprises at least one 3' end with a single nucleotide overhang.

[00127] Following contacting the at least one target analyte with the at least one probe, a one-ended adapter ligation method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00128] Following release of the identifier oligonucleotide, a one-ended adapter ligation method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00129] Following collection of the released identifier oligonucleotide, a one-ended adapter ligation method can further comprise: (4) ligating to the released oligonucleotide collected in step (3) at least one nucleic acid adapter;

[00130] The nucleic acid adapter can comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated "ROI #1", the nucleic acid adapter would comprise a nucleic acid sequence that corresponds to "ROI #1". The nucleic acid adapter can also comprise a unique molecular identifier. The nucleic acid adapter can also comprise a second amplification primer binding site.

[00131] In some aspects, the nucleic acid adapter can also comprise a constant nucleic acid sequence to minimize ligation bias caused by differences in sequences of particular identifier oligonucleotides. The constant nucleic acid sequence can comprise a cleavable moiety. The cleavable moiety can be enzymatically cleavable. In a non-limiting example, the enzymatically cleavable moiety can be a USER sequence, wherein the USER sequence comprises the sequence GUGUATUG.

[00132] The nucleic acid adapter can comprise any combination of the features described above.

[00133] The nucleic acid adapter can be a partially double-stranded nucleic acid molecule. In aspects in which the nucleic acid adapter is partially double-stranded, the nucleic acid adapter comprises a double-stranded annealed region and a single-stranded mismatched region.

[00134] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released can be present in the double-stranded annealed region of the nucleic acid adapter.

[00135] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a constant nucleic acid sequence to minimize ligation bias, the constant nucleic acid sequence to minimize ligation bias can be present in the double-stranded annealed region of the nucleic acid adapter.

[00136] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a unique molecular identifier, the unique molecular identifier can be present in the single-stranded mismatched region.

[00137] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a second amplification primer binding site, the second amplification primer binding site can be present in the single-stranded mismatched region of the nucleic acid adapter.

[00138] After ligation of the at least one nucleic acid adapter, a two-ended adapter ligation method can further comprise: (5) amplifying the ligation product produced in step (4); and (6) identifying the released oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the sample.

[00139] In other aspects of a one-ended adapter ligation method of the present disclosure, at least one of the amplification primers used in step (5) to amplify the ligation product produced in step

(4) comprises a nucleic acid sequence which identifies the specific location of the tissue sample form which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated "ROI #1", at least one of the amplification primers would comprise a nucleic acid sequence that corresponds to "ROI #1".

[00140] Figure 2 shows a schematic of a preferred aspect of a one-ended adapter ligation method of the present disclosure. In this aspect, the probe comprises a target binding domain that is an antibody that binds to a target protein. In the upper left panel, the probe binds to the target protein. In upper right panel, the UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target protein and a first amplification primer binding site. In this non-limiting example, the identifier oligonucleotide is double-stranded with one strand that comprises three separate nucleic acid molecules. The identifier oligonucleotide also comprises one 3' end with a single nucleotide overhang.

[00141] In the bottom panel of Figure 2, a nucleic acid adapter is ligated to the end of the identifier oligonucleotide that comprises the 3' single nucleotide overhang. In this non-limiting example, the nucleic acid adapter is partially double-stranded and comprises a double-stranded annealed region and a single-stranded mismatched region. Present in the double-stranded annealed region is a constant nucleic acid sequence to minimize ligation bias and a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. Present in the single-stranded mismatched region is a unique molecular identifier and a second amplification primer binding site. Following ligation of the nucleic acid adapter to the identifier oligonucleotide, the product is amplified using amplification primers that bind to the first and the second amplification primer binding sites and sequenced to identify the target protein bound by the probe.

[00142] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide ligated to one nucleic acid adapter for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide ligated to one nucleic acid adapter comprises an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which is capable of identifying a target analyte in a sample and a first amplification primer binding site. One end of the identifier oligonucleotide is attached to a nucleic acid adapter molecule,

wherein the nucleic acid adapter molecule is partially double-stranded and comprises a double-stranded annealed region and a single-stranded mismatched region and a second single-stranded mismatched region. The double-stranded mismatch region comprises a constant nucleic acid sequence to minimize ligation bias and a nucleic acid sequence nucleic acid sequence which is capable of identifying a specific location of a sample. The single-stranded mismatched region comprises a second amplification primer binding site. A schematic of an identifier oligonucleotide ligated to one nucleic acid adapter is shown in the bottom panel of Figure 2.

[00143] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "templated-primer extension method".

[00144] A templated-primer extension method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. The identifier oligonucleotide can also comprise a first amplification primer binding site.

[00145] Following contacting the at least one target analyte with the at least one probe, a templated-primer extension method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00146] Following release of the identifier oligonucleotide, a templated-primer extension method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier

oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00147] Following collection of the released identifier oligonucleotide, a templated-primer extension method can further comprise: (4) hybridizing to the released identifier oligonucleotide collected in step (3) a single stranded nucleic acid template.

[00148] The single stranded nucleic acid template can comprise a region complementary to the unique nucleic acid sequence of the identifier oligonucleotide, thereby allowing for the hybridization of the single stranded nucleic acid template and the collected identifier oligonucleotide.

[00149] The single stranded nucleic acid template can also comprise a nucleic acid sequence comprising a unique molecular identifier.

[00150] The single stranded nucleic acid template can also comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released.

[00151] The single stranded nucleic acid template can also comprise a nucleic acid sequence that is complementary to a second amplification primer binding site.

[00152] The single stranded nucleic acid template can comprise any combination of the features described above.

[00153] Following hybridization of the identifier oligonucleotide to the single stranded nucleic acid template, a templated-primer extension method can further comprise: (5) extending the identifier oligonucleotide of step (4) to form an extension produce complementary to the single stranded nucleic acid template, wherein the extension product comprises the identifier oligonucleotide and the sequence complementary to the single stranded nucleic acid template; (6) amplifying the extension product of step (6) using amplification primers that hybridize to the first and second amplification primer binding sites; and (7) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the sample.

[00154] In some aspects, the single stranded nucleic acid template can comprise an affinity molecule. In aspects in which the single stranded nucleic acid template comprises an affinity

molecule, a templated-primer extension method can further comprise an affinity purification step between steps (4) and (5).

[00155] Figure 3 shows a schematic of a preferred aspect of a templated-primer extension method of the present disclosure. In this aspect, the probe comprises a target binding domain that is an antibody that binds to a target protein. In the upper left panel, the probe binds to the target protein. In upper right panel, the UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target protein and a first amplification primer binding site. In the lower right panel, the identifier oligonucleotide is hybridized to a single stranded nucleic acid template. In this non-limiting example, the single-stranded nucleic acid template comprises an affinity molecule, a nucleic acid sequence complementary to the unique nucleic acid sequence of the identifier oligonucleotide, a first unique molecular identifier, and a sequence complementary to a second amplification primer biding site. The identifier oligonucleotide is extended to form an extension product complementary to the single stranded nucleic acid template. As shown in the lower left panel, the extension product comprises the identifier oligonucleotide, the nucleic acid sequence complementary to the first unique molecular identifier, and the second amplification primer binding site. Following the extension reaction, the primer extension product is amplified using amplification primers that bind to the first and the second amplification primer binding sites. In this non-limiting example, one of the amplification primers comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. The amplified product is then sequenced to identify the target protein bound by the probe.

[00156] In one aspect, the present disclosure provides a composition of a template-extended identifier oligonucleotide for spatially detecting at least one target analyte in a sample. The template-extended identifier oligonucleotide comprises a first flow cell adapter sequence suitable for sequencing, followed by a first unique molecular identifier, followed by an identifier oligonucleotide, followed by a second unique molecular identifier, followed by a second amplification primer binding site, followed by a third unique molecular identifier, followed by a second flow cell adapter sequence suitable for sequencing. The identifier oligonucleotide comprises a first amplification primer binding site and a unique nucleic acid sequence which is

capable of identifying a target analyte in a sample. A schematic of a template-extended identifier oligonucleotide is shown in the bottom panel of Figure 4.

[00157] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "short probe hybridization method".

[00158] A short probe hybridization method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain.

[00159] Following contacting the at least one target analyte with the at least one probe, a short probe hybridization method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00160] Following release of the identifier oligonucleotide, a short probe hybridization method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00161] Following collection of the released identifier oligonucleotide, a short probe hybridization method can further comprise: (4) hybridizing to the release identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe.

[00162] The first or the second nucleic acid probe can comprise a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. The first or the second nucleic acid probe can also comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. The first or the second nucleic acid probe can also comprise a nucleic acid sequence comprising unique molecular identifier. The first nucleic acid probe can comprise a first amplification primer binding site. The second nucleic acid probe can comprise a second amplification primer binding site.

[00163] The first or the second nucleic acid probe can comprise any combination of the features described above. In a preferred aspect depicted in Figure 5, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released is located 5' to the first amplification primer binding site and the unique molecular identifier is located 3' to the second amplification primer binding site.

[00164] In another preferred aspect depicted in Figure 6, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this same preferred aspect, the second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and the unique molecular identifier are located 5' to the first amplification primer binding site.

[00165] In another preferred aspect depicted in Figure 7, the first nucleic acid probe comprises a first amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this same preferred aspect, the second nucleic acid probe

comprises a second amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and the unique molecular identifier are located 3' to the second amplification primer binding site.

[00166] In another preferred aspect depicted in Figure 15, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence comprising a unique molecular identifier is located 3' to the first amplification binding site.

[00167] In another preferred aspect depicted in Figure 17, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence comprising a unique molecular identifier is located 5' to the first amplification binding site.

[00168] The first nucleic acid probe and the second nucleic acid probe can hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping. Alternatively, the first nucleic acid probe and the second nucleic acid probe can hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping.

[00169] Following hybridization of the first and the second nucleic acid probe to the identifier oligonucleotide, a short probe hybridization method can further comprise: (5) in the aspect in which the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping, ligating the first and the second nucleic acid probes together, for example, by performing a nick repair

reaction. Alternatively, in the aspect in which the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping, the method comprises ligating the first and the second nucleic acid probes together, for example, by performing a gap extension reaction and a nick repair reaction, such that the first and the second nucleic acid probes are ligated together.

[00170] Following ligation of the first and the second nucleic acid probe, a short probe hybridization method can further comprise: (6) amplifying the ligation product produced in step (5) using amplification primers that hybridize to the first and second amplification primer binding sites; and (7) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the sample.

[00171] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide-short nucleic acid probe complex for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide-short nucleic acid probe complex comprises an identifier oligonucleotide hybridized to a first nucleic acid probe and a second nucleic acid probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a target analyte in a sample. The first nucleic acid probe comprises a first amplification primer binding site, followed by a unique nucleic acid sequence capable of identifying a specific location in a sample, followed by a region complementary to the identifier oligonucleotide. The second nucleic acid probe comprises a second amplification primer binding site, followed by a nucleic acid sequence comprising a unique molecular identifier, followed by a region complementary to the identifier oligonucleotide. A schematic of an identifier oligonucleotide-short nucleic acid probe complex is depicted in Figure 5.

[00172] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide-short nucleic acid probe complex for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide-short nucleic acid probe complex comprises an identifier oligonucleotide hybridized to a first nucleic acid probe and a second nucleic acid probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a target analyte in a sample. The first nucleic acid probe comprises a first amplification primer binding site, followed by a nucleic acid sequence comprising a unique molecular identifier, followed by a region complementary to the identifier oligonucleotide,

wherein the nucleic acid sequence comprising a unique molecular identifier is located 3' to the first amplification primer binding site. The second nucleic acid probe comprises a second amplification primer binding site followed by a region complementary to the identifier oligonucleotide. A schematic of an identifier oligonucleotide-short nucleic acid probe complex is depicted in Figure 15.

[00173] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide-short nucleic acid probe complex for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide-short nucleic acid probe complex comprises an identifier oligonucleotide hybridized to a first nucleic acid probe and a second nucleic acid probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a target analyte in a sample. The first nucleic acid probe comprises a first amplification primer binding site, followed by a nucleic acid sequence comprising a unique molecular identifier, followed by a region complementary to the identifier oligonucleotide, wherein the nucleic acid sequence comprising a unique molecular identifier is located 5' to the first amplification primer binding site. The second nucleic acid probe comprises a second amplification primer binding site followed by a region complementary to the identifier oligonucleotide. A schematic of an identifier oligonucleotide-short nucleic acid probe complex is depicted in Figure 17.

[00174] Figure 16 shows a schematic overview of an exemplary short probe hybridization method of the present disclosure. First at least one target analyte in a sample is contacted with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target domain.

[00175] Following contacting the at least one target analyte with the at least one probe, a force is then provided to a location of the sample sufficient to release the identifier oligonucleotide. The identifier oligonucleotide is collected following release, as shown in the top panel of Figure 16.

[00176] As shown in the second panel from the top of Figure 16, the released identifier oligonucleotide is then hybridized to a first nucleic acid probe and a second nucleic acid probe. In this non-limiting example, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. The nucleic acid sequence comprising the unique molecular identifier is located 3' to the first amplification primer binding site. The second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this non-limiting example, the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping. Following hybridization to the identifier oligonucleotide, the first and second probe are ligated together, for example, by performing a nick repair reaction.

[00177] Following ligation of the first and second nucleic acid probes, the ligation product is amplified via PCR using amplification primers that hybridize to the first and second amplification primer binding sites. As shown in the second panel from the bottom of Figure 16, the amplification primer that hybridizes to the second amplification primer binding site comprises a first flow cell binding site, a first nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the second amplification primer binding site. The amplification primer that hybridizes to the first amplification primer binding site comprises a second flow cell binding site, a second nucleic acid sequence which identifies the specific location from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the first amplification primer binding site. The PCR product shown in the bottom panel of Figure 16 is then sequenced to identify the released oligonucleotide, thereby spatially detecting the at least one target analyte in the sample.

[00178] Figure 18 shows a schematic overview of an exemplary short probe hybridization method of the present disclosure. First, at least one target analyte in a sample is contacted with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target

nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target domain.

[00179] Following contacting the at least one target analyte with the at least one probe, a force is then provided to a location of the sample sufficient to release the identifier oligonucleotide. The identifier oligonucleotide is collected following release, as shown in the top panel of Figure 18.

[00180] As shown in the second panel from the top of Figure 18, the released identifier oligonucleotide is then hybridized to a first nucleic acid probe and a second nucleic acid probe. In this non-limiting example, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. The nucleic acid sequence comprising the unique molecular identifier is located 5' to the first amplification primer binding site. The second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this non-limiting example, the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping. Following hybridization to the identifier oligonucleotide, the first and second probe are ligated together, for example, by performing a nick repair reaction.

[00181] Following ligation of the first and second nucleic acid probes, the ligation product is amplified via PCR using amplification primers that hybridize to the first and second amplification primer binding sites. As shown in the second panel from the bottom of Figure 18, the amplification primer that hybridizes to the second amplification primer binding site comprises a first flow cell binding site, a first nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the second amplification primer binding site. The amplification primer that hybridizes to the first amplification primer binding site comprises a second flow cell binding site, a second nucleic acid sequence which identifies the specific location from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the first amplification primer binding site. The PCR product shown in the bottom panel Figure 18 is then

sequenced to identify the released oligonucleotide, thereby spatially detecting the at least one target analyte in the sample.

[00182] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "long probe hybridization method".

[00183] A long probe hybridization method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain.

[00184] Following contacting the at least one target analyte with the at least one probe, a long probe hybridization method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00185] Following release of the identifier oligonucleotide, a long probe hybridization method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00186] Following collection of the released identifier oligonucleotide(s), a long probe hybridization method can further comprise: (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe.

[00187] The first or the second nucleic acid probe can comprise a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. The first or the second nucleic acid probe can also comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released.

[00188] The first or the second nucleic acid probe can also comprise a first unique molecular identifier. The first or the second nucleic acid probe can also comprise a second unique molecular identifier. The first or the second nucleic acid probe can also comprise a third unique molecular identifier.

[00189] The first nucleic acid probe can comprise a first amplification primer binding site.

[00190] The first nucleic acid probe can also comprise a first flow cell binding site. The second nucleic acid probe can comprise a second flow cell binding site.

[00191] The first and the second nucleic acid probes can comprise any combination of the features described above. In a preferred aspect depicted in Figure 8, the first nucleic acid probe comprises a first flow cell binding site, a first unique molecular identifier, a first amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second flow cell binding site, a second unique molecular identifier, a third unique molecular identifier and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and the first unique molecular identifier are located 5' to the first flow cell binding site and the second and the third unique molecular identifiers are located 3' to the second flow cell binding site.

[00192] In another preferred aspect depicted in Figure 9, the first nucleic acid probe comprises a first flow cell binding site, a first unique molecular identifier, a second unique molecular identifier, a first amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second flow cell binding site, a third unique molecular identifier and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the first unique molecular identifier, the second unique molecular

identifier and the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released are located 5' to the first flow cell binding site and the third unique molecular identifier is located 3' to the second flow cell binding site.

[00193] In another preferred aspect depicted in Figure 10, the first nucleic acid probe comprises a first flow cell binding site, a first unique molecular identifier, a first amplification primer binding site and a nucleic acid sequence complementary to the identifier oligonucleotide. In this same preferred aspect, the second nucleic acid probe comprises a second flow cell binding site, a second unique molecular identifier, a third unique molecular identifier, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the first unique molecular identifier is located 5' to the first flow cell binding site and the second unique molecular identifier, the third unique molecular identifier, and the nucleic acid sequence which identifies the specific location of the sample form which the identifier oligonucleotide was released are located 3' to the second amplification primer binding site.

[00194] The first nucleic acid probe and the second nucleic acid probe can hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping. Alternatively, the first nucleic acid probe and the second nucleic acid probe can hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping.

[00195] Following hybridization of the first and the second nucleic acid probe to the identifier oligonucleotide, a long probe hybridization method can further comprise: (5) in the aspect in which the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping, performing a nick repair reaction such that the first and the second nucleic acid probes are ligated together. Alternatively, in the aspect in which the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping, the method comprises performing a gap extension and a nick repair reaction such that the first and the second nucleic acid probes are ligated together.

[00196] The name method can further comprise: (6) amplifying the ligation product produced in step (5) using amplification primers that hybridize to the first and second amplification primer

binding sites; and (7) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the sample.

[00197] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide-long nucleic acid probe complex for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide-long nucleic acid probe complex comprises an identifier oligonucleotide hybridized to a first nucleic acid probe and a second nucleic acid probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a target analyte in a sample. The first nucleic acid probe comprises a first flow cell binding site suitable for sequencing, followed by a first unique molecular identifier, followed by a first amplification primer binding site, followed by a unique nucleic acid sequence capable of identifying a specific location in a sample, followed by a region complementary to the identifier oligonucleotide. The second nucleic acid probe comprises a second flow cell binding site, followed by a second unique molecular identifier, followed by a third unique molecular identifier, followed by a region complementary to the identifier oligonucleotide. A schematic of an identifier oligonucleotide-short nucleic acid probe complex is depicted in Figure 8.

[00198] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "direct PCR method".

[00199] A direct PCR method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. The identifier oligonucleotide can also comprise a first amplification primer binding site, a second amplification primer binding site, or a unique molecular identifier. The identifier oligonucleotide can comprise any combination of these features. Any of these features

can also be flanked by regions comprising constant nucleic acid sequences of about 1 nucleotide to about 10 nucleotides.

[00200] Following contacting the at least one analyte with the at least one probe, a direct PCR method further comprises: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00201] A direct PCR method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00202] Following release of the identifier oligonucleotide, a direct PCR method can further comprise: (4) amplifying the released identifier oligonucleotide using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site. In some aspects, at least one of the amplification primers comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated "ROI #1", at least one of the amplification primers would comprise a nucleic acid sequence that corresponds to "ROI #1". In still other aspects, at least one of the amplification primers comprises a unique molecular identifier.

[00203] Following amplification, a direct PCR method of the present disclosure can further comprise: (5) identifying the released oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the sample.

[00204] Figure 11 shows a schematic of a preferred aspect of a direct PCR method of the present disclosure. In this aspect, the probe comprises a target binding domain comprising a nucleic acid sequence that is complementary to a target nucleic acid. In the upper panel, the probe hybridizes to the target nucleic acid. In the lower panel, a UV photo-cleavable linker located between the

target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a first amplification primer binding site, a second amplification primer binding site, a unique molecular identifier, and a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. Located between these four features are constant spacer regions that are 3 nucleotides in length. The identifier oligonucleotide is double-stranded and comprises a strand that comprises 3 separate nucleic acid molecules. After release, the identifier oligonucleotide is amplified using a first amplification primer that hybridizes to the first amplification primer binding site and comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a second amplification primer that hybridizes to the second amplification primer binding site. The amplified product is then sequenced to identify the target nucleic acid bound by the probe.

[00205] Figure 19 shows a schematic of a preferred aspect of a direct PCR method of the present disclosure. In this aspect, the identifier oligonucleotide comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier, a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain and a second amplification primer binding site, as shown in the top panel of Figure 19. The identifier oligonucleotide is amplified using amplification primers that hybridize to the first and second amplification primer binding sites. As shown in the middle panel of Figure 19, the amplification primer that hybridizes to the second amplification primer binding site comprises a first flow cell binding site, a first nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the second amplification primer binding site. The amplification primer that hybridizes to the first amplification primer binding site comprises a second flow cell binding site, a second nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the first amplification primer binding site. The PCR product shown in the bottom panel of Figure 19 is sequenced to identify the released oligonucleotide, thereby spatially detecting the at least one target analyte in the sample.

[00206] In one aspect, the present disclosure provides a composition of a direct-PCR compatible identifier oligonucleotide for spatially detecting at least one target analyte in a sample. A direct-

PCR compatible identifier oligonucleotide comprises a first amplification primer binding site, followed a unique nucleic acid sequence which is capable of identifying a target analyte in a sample, followed by a unique molecular identifier, followed by a second amplification primer binding site. A schematic of a direct-PCR compatible identifier oligonucleotide is depicted in the lower panel of Figure 11.

[00207] In one aspect, the present disclosure provides a composition of a direct-PCR compatible identifier oligonucleotide for spatially detecting at least one target analyte in a sample. A direct-PCR compatible identifier oligonucleotide comprises a first amplification primer binding site, followed a nucleic acid sequence comprising a unique molecular identifier, followed by a unique nucleic acid sequence which is capable of identifying a target analyte in a sample, followed by a second amplification primer binding site. A schematic of a direct-PCR compatible identifier oligonucleotide is shown in the top panel of Figure 19.

[00208] In some aspects of the methods of the present disclosure at least one, or at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least ten, or at least eleven, or at least twelve, or at least thirteen, or at least fourteen, or at least fifteen, or at least sixteen, or at least seventeen, or at least eighteen, or at least nineteen, or at least twenty, or at least thirty, or at least forty, or at least fifty, or at least sixty, or at least seventy, or at least eighty, or at least ninety, or at least one hundred probes can been to a single target analyte. As used herein, the term "tiling" is used to describe when more than one probe of the present disclosure is bound to a target analyte. The top panel of Figure 27 shows the tiling of probes onto a target RNA. Tiling multiple probes onto a target analyte means that each target analyte will be individually detected multiple times, increasing the overall accuracy of the measurement. In a non-limiting example, as shown in the bottom panel of Figure 27, in the case where 10 probes are tiled onto a single target RNA, one of the probes may be incorrectly detected too many times (outlier high count probe), while another probe may be incorrectly detected too few times (outlier low count probe). However, the other 8 probes may be detected at a similar level, indicating that the two outliers should be discarded during analysis and the signals from the 8 probes used to generate a more accurate measurement of the abundance of the target RNA.

[00209] The present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as an "enzyme free method".

[00210] An enzyme free method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. An identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. An identifier oligonucleotide can also comprise a capture probe binding site.

[00211] Following contacting at least one target analyte with at least one probe, an enzyme free method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00212] Following release of an identifier oligonucleotide, an enzyme free method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00213] Following collection of a released identifier oligonucleotide, an enzyme free method can further comprise: (4) hybridizing to a released identifier oligonucleotide a capture probe.

[00214] A capture probe can comprise a region complementary to the capture probe binding site. A capture probe can also comprise an affinity molecule.

[00215] Following hybridization of a capture probe, an enzyme free method can further comprise: (5) Identifying a released identifier oligonucleotide by sequencing the hybridized product produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[00216] A hybridized product produced in step (4) can be sequenced using an enzyme free method of sequencing. Enzyme-free methods of sequencing have been described in, e.g., US2014946386 and US15/819,151 (US Patent No. 10,415,080), each of which is incorporated herein by reference in its entirety.

[00217] Figure 12 shows a schematic of a preferred aspect of an enzyme free method of the present disclosure. In this aspect, the probe comprises a target binding domain comprising a nucleic acid sequence that is complementary to a target nucleic acid. In the top panel, the probe hybridizes to the target nucleic acid. In the middle panel, a UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target nucleic acid bound to the target binding domain and a capture probe binding site. After release, the identifier oligonucleotide is hybridized to a capture probe, as depicted in the bottom panel. The capture probe comprises a nucleic acid sequence complementary to the capture probe binding site and an affinity molecule. The hybridized product is then sequenced using enzyme free sequencing methods to identify the target nucleic acid bound by the probe.

[00218] In one aspect, the present disclosure provides a composition of a hybridized identifier oligonucleotide-capture probe complex for spatially detecting at least one target analyte in a sample. A hybridized identifier oligonucleotide-capture probe complex comprises an identifier oligonucleotide hybridized to a capture probe. An identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a specific a target analyte in a sample and a capture probe binding site. A capture probe comprises an affinity molecule and a region complementary to the capture probe binding site. A schematic of a hybridized identifier oligonucleotide-capture probe complex is depicted in the bottom of panel Figure 12.

[00219] The present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "multiplexed enzyme free method".

[00220] A multiplexed enzyme free method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. The identifier oligonucleotide can also comprise a capture probe binding site. The identifier oligonucleotide can also comprise a multiplexing probe binding site.

[00221] Following contacting the at least one target analyte with the at least one probe, a multiplexed enzyme free method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00222] Following release of the identifier oligonucleotide, a multiplexed enzyme free method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00223] Following collection of the released identifier oligonucleotide, a multiplexed enzyme free method can further comprise: (4) hybridizing to the released identifier oligonucleotide a capture probe and a multiplexing probe.

[00224] A capture probe can comprise a region complementary to the capture probe binding site. A capture probe can also comprise an affinity molecule.

[00225] A multiplexing probe can comprise a region complementary to the multiplexing probe binding site. A multiplexing probe can also comprise a nucleic acid sequence which identifies

the specific location of the sample from which the identifier oligonucleotide was released and a region complementary to the multiplexing probe binding site.

[00226] Following hybridization of a capture probe and a multiplexing probe, a multiplexed enzyme free method can further comprise: (5) Identifying the released identifier oligonucleotide by sequencing the hybridized product produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[00227] A hybridized product produced in step (4) can be sequenced using an enzyme free method of sequencing. Enzyme-free methods of sequencing have been described in, e.g., US2014946386 and US15/819,151, each of which is incorporated herein by reference in its entirety.

[00228] Figure 13 shows a schematic of a preferred aspect of a multiplexed enzyme free method of the present disclosure. In this aspect, the probe comprises a target binding domain comprising a nucleic acid sequence that is complementary to a target nucleic acid. In the top panel, the probe hybridizes to the target nucleic acid. In the middle panel, a UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a capture probe binding site, and a multiplexing probe binding site, as shown in the middle panel. After release, the identifier oligonucleotide is hybridized to a capture probe and a multiplexing probe as shown in the lower panel. The capture probe comprises a nucleic acid sequence complementary to the capture probe binding site and an affinity molecule. The multiplexing probe comprises a nucleic acid sequence complementary to the multiplexing probe binding site and a nucleic acid sequence which identifies the specific location of a sample from which the identifier oligonucleotide was released. The hybridized product is then sequenced using enzyme free sequencing methods to identify the target nucleic acid bound by the probe.

[00229] In one aspect, the present disclosure provides a composition of a hybridized identifier oligonucleotide-capture probe-multiplex probe complex for spatially detecting at least one target analyte in a sample. A hybridized identifier oligonucleotide-capture probe-multiplex probe complex comprises an identifier oligonucleotide hybridized to a capture probe and a multiplexing probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a specific a target analyte in a sample, a capture probe binding site and a

multiplexing probe binding site. The capture probe comprises an affinity molecule and a region complementary to the capture probe binding site. The multiplexing probe comprises a nucleic acid sequence which identifies the specific location of a sample from which the identifier oligonucleotide was released and a region complementary to the multiplexing probe binding site. A schematic of a hybridized identifier oligonucleotide-capture probe-multiplex probe complex is depicted in the bottom panel of Figure 13.

[00230] Figure 20 is an exemplary schematic of overview of the methods of the present disclosure. First, a sample on a microscope slide is contacted with a plurality of probes of the present disclosure (step 1 in Figure 20). The slide is then imaged and particular regions of interest (ROIs) are selected (step 2 in Figure 20). A specific ROI is then illuminated by UV light to release identifier oligonucleotides from probes bound within the ROI. The released identifier oligonucleotides are then collected via aspiration with a microcapillary. Following aspiration, the identifier oligonucleotides are transferred to a particular well within a 96 well plate. Steps 4 and 5 are then repeated for each ROI identified in step 2. After all ROIs have been illuminated and all released identifier oligonucleotides collected, the identifier oligonucleotides are sequenced using next generation sequencing methods to spatially detect at least one target analyte in the sample.

[00231] As described in the preceding, the present disclosure provides probes for the compositions and methods of spatially detecting at least one target analyte in a sample. The present disclosure provides probes comprising a target binding domain and an identifier oligonucleotide. The target binding domain is a region of the probe that specifically binds to at least one target analyte in a sample.

[00232] Probes of the present disclosure can be used for spatially detecting a target nucleic acid. In this aspect, the target binding domain can be a target nucleic acid-binding region. The target nucleic acid-binding region is preferably at least 15 nucleotides in length, and more preferably is at least 20 nucleotides in length. In specific aspects, the target nucleic acid-binding region is approximately 10 to 500, 20 to 400, 25, 30 to 300, 35, 40 to 200, or 50 to 100 nucleotides in length. Probes and methods for binding and identifying a target nucleic acid have been described in, *e.g.*, US2003/0013091, US2007/0166708, US2010/0015607, US2010/0261026, US2010/0262374, US2010/0112710, US2010/0047924, and US2014/0371088, each of which is incorporated herein by reference in its entirety.

[00233] The target nucleic acid-binding region can directly hybridize to a target nucleic acid present in a sample. Alternatively, the probes of the present disclosure can indirectly hybridize to a target nucleic acid present in a sample (via an intermediary oligonucleotide). Figure 14 illustrates a probe (or composition) of this aspect. The probe includes a target nucleic-acid binding domain which binds to a synthetic oligonucleotide (the intermediary oligonucleotide) that in turn binds to a target nucleic acid in a biological sample. It could be said that the intermediary oligonucleotide is a probe, as defined herein, since it comprises a nucleic acid backbone and is capable of binding a target nucleic acid. In these aspects, a probe's target nucleic acid-binding region hybridizes to a region of an intermediary oligonucleotide (*i.e.*, a synthetic oligonucleotide) which is different from the target nucleic acid present in a sample. Thus, the probe's target binding region is independent of the ultimate target nucleic acid in the sample. This allows economical and rapid flexibility in an assay design, as the target (present in a sample)-specific components of the assay are included in inexpensive and widely-available synthetic DNA oligonucleotides rather than the more expensive probes. Such synthetic oligonucleotides are simply designed by including a region that hybridizes to the target nucleic acid present in a sample and a region that hybridizes to a probe. Therefore, a single set of indirectly-binding probes can be used to detect an infinite variety of target nucleic acids (present in a sample) in different experiments simply by replacing the target-specific (synthetic) oligonucleotide portion of the assay.

[00234] A target nucleic acid may be DNA or RNA and preferably messenger RNA (mRNA) or miRNA.

[00235] Probes of the present disclosure can be used for detecting a target protein. In this aspect, the target binding domain can be a target protein-binding region. A target protein-binding region includes molecules or assembles that are designed to bind to at least one target protein, at least one target protein surrogate, or both and can, under appropriate conditions, form a molecular complex comprising the probe and the target protein. The target-protein binding region can include an antibody, a peptide, an aptamer, or a peptoid. The antibody can be obtained from a variety of sources, including but not limited to polyclonal antibody, monoclonal antibody, monospecific antibody, recombinantly expressed antibody, humanized antibody, plantibodies, and the like. The terms protein, polypeptide, peptide, and amino acid sequence are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be

linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids or synthetic amino acids. The terms also encompass an amino acid polymer that has been modified, for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term amino acid refers to either natural and/or unnatural or synthetic amino acids, including but not limited to glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. Probes and methods for binding and identifying a target protein have been described, *e.g.*, in US2011/0086774, the contents of which is incorporated herein by reference in its entirety.

[00236] An identifier oligonucleotide is a nucleic acid molecule that identifies the target analyte bound to the target binding domain. The identifier oligonucleotide comprises a unique nucleic acid sequence that identifies the target analyte bound to the target binding domain of the probe. In a non-limiting example, a probe with a target binding domain that binds to the protein P53 comprises an identifier oligonucleotide with a unique nucleic acid sequence that corresponds to P53, while a probe with a target binding domain that binds to the protein P97 comprises an identifier oligonucleotide with a unique nucleic acid sequence that corresponds to P97.

[00237] An identifier oligonucleotide can be DNA, RNA, or a combination of DNA and RNA.

[00238] In some aspects, an identifier oligonucleotide can comprise at least one amplification primer binding site. An amplification primer binding site is a nucleic acid sequence capable of binding to an amplification primer. An amplification primer can be used to amplify the nucleic molecule to which it is bound using methods known in the art, including, but not limited to, polymerase chain reaction (PCR).

[00239] In some aspects, an identifier oligonucleotide can comprise at least one unique molecular identifier.

[00240] An identifier oligonucleotide can be a single-stranded, a double-stranded, or a partially double-stranded nucleic acid molecule. In the aspects in which an identifier oligonucleotide is double-stranded or partially double-stranded, at least one of the two strands can comprise at least two separate nucleic acid molecules which, without being bound by theory, allows for denaturing of the identifier oligonucleotide at lower temperatures.

[00241] An identifier oligonucleotide can also comprise at least one 3' end that comprises a single nucleotide overhang.

[00242] An identifier oligonucleotide can also comprise a capture probe binding site. A capture probe binding site is a nucleic acid sequence to which a capture probe can bind.

[00243] A capture probe of the present disclosure can comprise a nucleic acid sequence complementary to a capture probe binding site. A capture probe can also comprise an affinity molecule.

[00244] An identifier oligonucleotide can also comprise a multiplexing probe binding site. A multiplexing probe binding site is a nucleic acid sequence to which a multiplexing probe can bind.

[00245] A multiplexing probe of the present disclosure can comprise a nucleic acid sequence complementary to a multiplexing probe binding site. A multiplexing probe can also comprise a nucleic acid sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released.

[00246] A probe of the present disclosure can include a region which permits the release of an identifier oligonucleotide following the application of a suitable force. In one non-limited example, the region is a cleavable motif (*e.g.*, a restriction enzyme site or cleavable linker). The cleavable motif allows release of an identifier oligonucleotide from a bound target nucleic acid or protein and the identifier oligonucleotide can then be collected and detected. The region which permits the release of an identifier oligonucleotide can be positioned between the target-binding domain and the identifier oligonucleotide, allowing for the release of the identifier oligonucleotide from the target binding domain. An identifier oligonucleotide is said to be releasable when it can be separated (*i.e.*, cleaved and released) from the remainder of the probe. Examples of cleavable motives include but are not limited to photo-cleavable linkers. Photo-cleavable linkers can be cleaved by light provided by a suitable coherent light source (*e.g.*, a laser and a UV light source) or a suitable incoherent light source (*e.g.*, an arc-lamp and a light-emitting diode (LED)).

[00247] In some aspects, the identifier oligonucleotide is collected from a solution proximal to, e.g., at least immediately above or surrounding, the point at which the identifier oligonucleotide is released or the at least one cell. The proximal solution may be collected by aspirating, e.g., via a pipette, a capillary tube, a microarray pin, a flow cell comprising holes, or another suitable aspirating system known in the art or any combination thereof. The capillary tube may comprise an optical device capable of transmitting a light force, e.g., UV light, to the at least one cell. The

pipette or a microarray pin may be attached to an array comprising a plurality of pipettes or microarray pins. The proximal solution may comprise an anionic polymer, e.g., dextran sulfate, and/or salmon sperm DNA and/or the collected signal oligonucleotide may be added to a solution comprising an anionic polymer, e.g., dextran sulfate, and/or salmon sperm DNA. Other non-specific blocking agents known in the art in addition to or instead of salmon sperm DNA may be used.

[00248] In some aspects, the identifier oligonucleotide is collected from a tissue, at least one cell or proximal to the point at which the identifier oligonucleotide is released via liquid laminar, turbulent, or transitional flow. The flow may be via a channel, e.g., having 25 to 500 m depth between the tissue and a fluidic device or impermeable barrier placed over the tissue.

[00249] In aspects where the target-binding domain of a probe is an antibody, the probe can be prepared using a cysteine bioconjugation method that is stable, site-specific to, preferably, the antibody's hinge-region heavy-chain. This preparation method provides relatively controllable identifier oligonucleotides to antibody stoichiometric ratios. A probe can comprise a plurality (i.e., more than one, e.g., 2, 3, 4, 5, or more) identifier oligonucleotides per antibody. Generally, "heavier" probes, which comprise 3 or 4 identifier oligonucleotides per antibody, are significantly less sensitive than antibodies lacking an identifier oligonucleotide or "lighter" probes, which comprise 1 or 2 identifier oligonucleotides per antibody.

[00250] In aspects, probes are provided to a sample at concentrations typically less than that used for immunohistochemistry (IHC) or for in situ hybridization (ISH). Alternately, the concentration may be significantly less than that used for IHC or ISH. For example, the probe concentration may be 2 fold less, 5 fold less, 10 fold less, 20 fold less, 25 fold less, 30 fold less, 50 fold less, 60 fold less, 70 fold less, 80 fold less, 90 fold less, 100 fold less, 200 fold less, 300 fold less, 400 fold less, 500 fold less, 600 fold less, 700 fold less, 800 fold less, 900 fold less, 1000 fold less, 2000 fold less, or less and any number in between. In aspects, probes are provided at a concentration of 100 nM, 70 nM, 60 nM, 50 nM, 40 nM, 30 nM, 20 nM, 10 nM, 9 nM, 8 nM, 7 nM, 6 nM, 5 nM, 4 nM, 3 nM, 2 nM, 1 nM, 0.9 nM, 0.8 nM, 0.7 nM, 0.6 nM, 0.5 nM, 0.4 nM, 0.3 nM, 0.2 nM, 0.1 nM, 0.09 nM, 0.08 nM, 0.07 nM, 0.06 nM, 0.05 nM, 0.04 nM, 0.03 nM, 0.02 nM, 0.01 nM, and less and any concentration in between.

[00251] Background noise, during protein detection, can be reduced by performing a negative purification of the intact probe molecule. This can be done by conducting an affinity purification

of the antibody or photo-cleavable linker after collection of eluate from a region of interest. Normally, released signal oligonucleotides will not be pulled out of solution. A protein-G or -O mechanism in a pipet tip, tube, or plate can be employed for this step. Such devices and reagents commercially available.

[00252] Background noise, during nucleic acid detection, can be reduced by performing a negative purification of the intact probe molecule. This can be done by conducting an affinity purification of the target binding domain or photo-cleavable linker after collection of eluate from a region of interest. Normally, released signal oligonucleotides will not be pulled out of solution. To assist in the negative purification, a universal purification sequence may be included in a probe, e.g., in the target binding domain.

[00253] Protein-targeting probes and nucleic acid-targeting probes may be applied simultaneously as long as conditions allow for binding of both a protein target and a nucleic acid target. Alternately, protein-targeting probes and nucleic acid-targeting probes may be applied sequentially when conditions allowing for binding of both a protein target and a nucleic acid target are not possible.

[00254] A set of probes is synonymous with a composition of probes. A set of probes includes at least one species of probes, *i.e.*, directed to one target. A set of probes preferably includes at least two, e.g., 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more species of probes. A probe set may include one or multiple copies of each species of probe.

[00255] A first set of probes only may be applied to a sample. Alternately, a second set (or higher number) of probes may be later applied to the sample. The first set and second (or higher number) may target only nucleic acids, only proteins, or a combination thereof.

[00256] In the present disclosure, two or more targets (*i.e.*, proteins, nucleic acids, or a combination thereof) are detected; 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more targets, and any number there between, are detected.

[00257] A set of probes may be pre-defined based upon the cell type or tissue type to be targeted. For example, if the tissue is a breast cancer, then the set of probes will include probes directed to proteins relevant to breast cancer cells (e.g., Her2, EGFR, and PR) and/or probes directed to proteins relevant to normal breast tissues. Additionally, the set of probes may be pre-defined

based upon developmental status of a cell or tissue to be targeted. Alternately, the set of probes may be pre-defined based upon subcellular localizations of interest, e.g., nucleus, cytoplasm, and membrane. For example, antibodies directed to Foxp3, Histone H3, or P-S6 label the nucleus, antibodies directed to CD3, CD4, PD-1, or CD45RO label the cytoplasm, and antibodies directed to PD-L1 label membranes.

[00258] A probe may be chemically synthesized or may be produced biologically using a vector into which a nucleic acid encoding the probe has been cloned.

[00259] Any probe or set of probes described herein may be used in methods and kits of the present disclosure.

[00260] For the herein-described probes, association of a unique nucleic acid sequence to a specific target nucleic acid or target protein is not fixed.

[00261] As described in the preceding, probes of the present disclosure can be used to detect a target nucleic acid or target protein present in any sample, *e.g.*, a biological sample. As will be appreciated by those in the art, the sample may comprise any number of things, including, but not limited to: cells (including both primary cells and cultured cell lines) and tissues (including cultured or explanted). In aspects, a tissue sample (fixed or unfixed) is embedded, serially sectioned, and immobilized onto a microscope slide. As is well known, a pair of serial sections will include at least one cell that is present in both serial sections. Structures and cell types, located on a first serial section will have a similar location on an adjacent serial section. The sample can be cultured cells or dissociated cells (fixed or unfixed) that have been immobilized onto a slide. A sample can be a formalin-fixed paraffin-embedded (FFPE) tissue sample.

[00262] In aspects, a tissue sample is a biopsied tumor or a portion thereof, i.e., a clinically-relevant tissue sample. For example, the tumor may be from a breast cancer. The sample may be an excised lymph node.

[00263] The sample can be obtained from virtually any organism including multicellular organisms, *e.g.*, of the plant, fungus, and animal kingdoms; preferably, the sample is obtained from an animal, *e.g.*, a mammal. Human samples are particularly preferred.

[00264] In some aspects, the probes, compositions, methods, and kits described herein are used in the diagnosis of a condition. As used herein the term diagnose or diagnosis of a condition includes predicting or diagnosing the condition, determining predisposition to the condition, monitoring treatment of the condition, diagnosing a therapeutic response of the disease, and

prognosis of the condition, condition progression, and response to particular treatment of the condition. For example, a tissue sample can be assayed according to any of the probes, methods, or kits described herein to determine the presence and/or quantity of markers of a disease or malignant cell type in the sample (relative to the non-diseased condition), thereby diagnosing or staging a disease or a cancer.

[00265] In general, samples attached to a slide can be first imaged using fluorescence (e.g., fluorescent antibodies or fluorescent stains (e.g., DAPI)) to identify morphology, regions of interest, cell types of interest, and single cells and then expression of proteins and/or nucleic acids can be digitally counted from the sample on the same slide.

[00266] Compositions and kits of the present disclosure can include probes and other reagents, for example, buffers and other reagents known in the art to facilitate binding of a protein and/or a nucleic acid in a sample, *i.e.*, for performing hybridization reactions.

[00267] A kit also will include instructions for using the components of the kit, including, but not limited to, information necessary to hybridize labeled oligonucleotides to a probe, to hybridize a probe to a target-specific oligonucleotide, to hybridize a target-specific oligonucleotide to a target nucleic acid and/or to hybridize a probe to target protein.

[00268] A region of interest may be a tissue type present in a sample, a cell type, a cell, or a subcellular structure within a cell.

[00269] Together, a comparison of the identity and abundance of the target proteins and/or target nucleic acids present in a first region of interest (*e.g.*, tissue type, a cell type (including normal and abnormal cells), and a subcellular structure within a cell) and the identity and abundance of the target proteins and/or target nucleic acids present in second region of interest or more regions of interest can be made using the methods of the present disclosure.

[00270] As described in the preceding, the products produced by the methods of the present disclosure can be used for nucleic acid amplification. In a preferred aspect, the nucleic acid amplification can be solid-phase nucleic acid amplification. Thus, in further aspects the invention provides a method of solid-phase nucleic acid amplification of template polynucleotide molecules which comprises: preparing a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends using the methods of the present disclosure and carrying out a solid-phase nucleic acid amplification reaction wherein said template polynucleotide molecules are amplified. Compositions and methods for nucleic acid

amplification and sequencing have been described in, e.g., US9376678, which is incorporated herein by reference in its entirety.

[00271] The term "solid-phase amplification" as used herein refers to any nucleic acid amplification reaction carried out on or in association with a solid support such that all or a portion of the amplified products are immobilized on the solid support as they are formed. In particular, the term encompasses solid-phase polymerase chain reaction (solid-phase PCR), which is a reaction analogous to standard solution phase PCR, except that one or both of the forward and reverse amplification primers is/are immobilized on the solid support.

[00272] Although the invention encompasses "solid-phase" amplification methods in which only one amplification primer is immobilized (the other primer usually being present in free solution), it is preferred for the solid support to be provided with both the forward and the reverse primers immobilized. In practice, there will be a "plurality" of identical forward primers and/or a "plurality" of identical reverse primers immobilized on the solid support, since the PCR process requires an excess of primers to sustain amplification. References herein to forward and reverse primers are to be interpreted accordingly as encompassing a "plurality" of such primers unless the context indicates otherwise.

[00273] As will be appreciated by the skilled reader, any given PCR reaction requires at least one type of forward primer and at least one type of reverse primer specific for the template to be amplified. However, in certain aspects the forward and reverse primers may comprise template-specific portions of identical sequence, and may have entirely identical nucleotide sequence and structure (including any non-nucleotide modifications). In other words, it is possible to carry out solid-phase amplification using only one type of primer, and such single-primer methods are encompassed within the scope of the invention. Other aspects may use forward and reverse primers which contain identical template-specific sequences but which differ in some other structural features. For example one type of primer may contain a non-nucleotide modification which is not present in the other.

[00274] In other aspects of the invention the forward and reverse primers may contain template-specific portions of different sequence.

[00275] Amplification primers for solid-phase PCR are preferably immobilized by covalent attachment to the solid support at or near the 5' end of the primer, leaving the template-specific portion of the primer free for annealing to its cognate template and the 3' hydroxyl group free for

primer extension. Any suitable covalent attachment means known in the art may be used for this purpose. The chosen attachment chemistry will depend on the nature of the solid support, and any derivatization or functionalization applied to it. The primer itself may include a moiety, which may be a non-nucleotide chemical modification, to facilitate attachment. In one particularly preferred aspect the primer may include a sulphur-containing nucleophile, such as phosphorothioate or thiophosphate, at the 5' end. In the case of solid-supported polyacrylamide hydrogels (as described below), this nucleophile will bind to a "C" group present in the hydrogel. The most preferred means of attaching primers and templates to a solid support is via 5' phosphorothioate attachment to a hydrogel comprised of polymerised acrylamide and N-(5-bromoacetamidylpentyl)acrylamide (BRAPA).

[00276] The terms "cluster" and "colony" are used interchangeably herein to refer to a discrete site on a solid support comprised of a plurality of identical immobilized nucleic acid strands and a plurality of identical immobilized complementary nucleic acid strands. The term "clustered array" refers to an array formed from such clusters or colonies. In this context the term "array" is not to be understood as requiring an ordered arrangement of clusters.

[00277] The invention also encompasses methods of sequencing the amplified nucleic acids generated by solid-phase amplification. Thus, the invention provides a method of nucleic acid sequencing comprising amplifying a library of nucleic acid templates by the methods of the present disclosure described above, using solid-phase amplification as described above to amplify this library on a solid support, and carrying out a nucleic acid sequencing reaction to determine the sequence of the whole or a part of at least one amplified nucleic acid strand produced in the solid-phase amplification reaction.

[00278] Sequencing, as referred to herein, can be carried out using any suitable "sequencing-by-synthesis" technique, wherein nucleotides are added successively to a free 3' hydroxyl group, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction. The nature of the nucleotide added is preferably determined after each nucleotide addition.

[00279] The initiation point for the sequencing reaction may be provided by annealing of a sequencing primer to a product of the whole genome or solid-phase amplification reaction. In this connection, one or both of the adapters added during formation of the template library may include a nucleotide sequence which permits annealing of a sequencing primer to amplified products derived by whole genome or solid-phase amplification of the template library.

[00280] The products of solid-phase amplification reactions wherein both forward and reverse amplification primers are covalently immobilized on the solid surface are so-called "bridged" structures formed by annealing of pairs of immobilized polynucleotide strands and immobilized complementary strands, both strands being attached to the solid support at the 5' end. Arrays comprised of such bridged structures provide inefficient templates for nucleic acid sequencing, since hybridization of a conventional sequencing primer to one of the immobilized strands is not favored compared to annealing of this strand to its immobilized complementary strand under standard conditions for hybridization.

[00281] In order to provide more suitable templates for nucleic acid sequencing it is preferred to remove substantially all or at least a portion of one of the immobilized strands in the "bridged" structure in order to generate a template which is at least partially single-stranded. The portion of the template which is single-stranded will thus be available for hybridization to a sequencing primer. The process of removing all or a portion of one immobilized strand in a "bridged" double-stranded nucleic acid structure may be referred to herein as "linearization".

[00282] Bridged template structures may be linearized by cleavage of one or both strands with a restriction endonuclease or by cleavage of one strand with a nicking endonuclease. Other methods of cleavage can be used as an alternative to restriction enzymes or nicking enzymes, including inter alfa chemical cleavage (e.g. cleavage of a diol linkage with periodate), cleavage of abasic sites by cleavage with endonuclease, or by exposure to heat or alkali, cleavage of ribonucleotides incorporated into amplification products otherwise comprised of deoxyribonucleotides, photochemical cleavage or cleavage of a peptide linker.

[00283] It will be appreciated that a linearization step may not be essential if the solid-phase amplification reaction is performed with only one primer covalently immobilized and the other in free solution.

[00284] In order to generate a linearized template suitable for sequencing it is necessary to remove "unequal" amounts of the complementary strands in the bridged structure formed by amplification so as to leave behind a linearized template for sequencing which is fully or partially single stranded. Most preferably one strand of the bridged structure is substantially or completely removed.

[00285] Following the cleavage step, regardless of the method used for cleavage, the product of the cleavage reaction may be subjected to denaturing conditions in order to remove the portion(s)

of the cleaved strand(s) that are not attached to the solid support. Suitable denaturing conditions will be apparent to the skilled reader with reference to standard molecular biology protocols.

[00286] Denaturation (and subsequent re-annealing of the cleaved strands) results in the production of a sequencing template which is partially or substantially single-stranded. A sequencing reaction may then be initiated by hybridization of a sequencing primer to the single-stranded portion of the template.

[00287] Thus, the nucleic acid sequencing reaction may comprise hybridizing a sequencing primer to a single-stranded region of a linearized amplification product, sequentially incorporating one or more nucleotides into a polynucleotide strand complementary to the region of amplified template strand to be sequenced, identifying the base present in one or more of the incorporated nucleotide(s) and thereby determining the sequence of a region of the template strand.

[00288] One preferred sequencing method which can be used in accordance with the invention relies on the use of modified nucleotides that can act as chain terminators. Once the modified nucleotide has been incorporated into the growing polynucleotide chain complementary to the region of the template being sequenced there is no free 3'-OH group available to direct further sequence extension and therefore the polymerase cannot add further nucleotides. Once the nature of the base incorporated into the growing chain has been determined, the 3' block may be removed to allow addition of the next successive nucleotide. By ordering the products derived using these modified nucleotides it is possible to deduce the DNA sequence of the DNA template. Such reactions can be done in a single experiment if each of the modified nucleotides has attached a different label, known to correspond to the particular base, to facilitate discrimination between the bases added at each incorporation step. Alternatively, a separate reaction may be carried out containing each of the modified nucleotides separately.

[00289] The modified nucleotides may carry a label to facilitate their detection. Preferably this is a fluorescent label. Each nucleotide type may carry a different fluorescent label. However the detectable label need not be a fluorescent label. Any label can be used which allows the detection of an incorporated nucleotide.

[00290] One method for detecting fluorescently labelled nucleotides comprises using laser light of a wavelength specific for the labelled nucleotides, or the use of other suitable sources of

illumination. The fluorescence from the label on the nucleotide may be detected by a CCD camera or other suitable detection means.

[00291] The invention is not intended to be limited to use of the sequencing method outlined above, as essentially any sequencing methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used. Suitable alternative techniques include, for example, Pyrosequencing, FISSEQ (fluorescent in situ sequencing), MPSS (massively parallel signature sequencing) and sequencing by ligation-based methods.

[00292] In methods of the present disclosure, the unique nucleic acid sequence present in the identifier oligonucleotide of a probe which identifies the target analyte bound to the target binding domain of the probe can comprise between about 5 nucleotides and about 50 nucleotides. Preferably, the sequence comprises between about 20 nucleotides and about 40 nucleotides. Even more preferably, the sequence comprises about 35 nucleotides. In some preferred aspects, the sequence comprises 10 nucleotides.

[00293] In methods of the present disclosure, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released comprises between about 6 nucleotides and about 15 nucleotides. Preferably, the sequence comprises about 12 nucleotides.

[00294] In methods of the present disclosure, an amplification primer binding sites comprises between about 18 nucleotides and about 40 nucleotides. Preferably, an amplification primer binding sites comprises about 32 nucleotides.

[00295] In some aspects of the methods of the present disclosure, an amplification primer binding site can comprise an i7 sequence, wherein the i7 sequence comprises the sequence set forth in SEQ ID NO: 1.

[00296] In some aspects of the methods of the present disclosure, an amplification primer binding site can comprise an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2.

[00297] In some aspects of the methods of the present disclosure, an amplification primer can comprise a flow cell adapter sequence, wherein the flow cell adapter sequence is suitable for sequencing. Preferably, at least one amplification primer used in the methods of the present disclosure comprises a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3. Preferably still, at least one amplification

primer used in the methods of the present disclosure comprises a P7 flow cell adapter sequence, wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4.

[00298] In methods of the present disclosure, a unique molecular identifier can comprise between about 6 nucleotides and about 30 nucleotides. Preferably, a unique molecular identifier can comprise about 15 nucleotides. The terms unique molecular identifier and random molecular tags are used interchangeably herein. Using methods known in that art, unique molecular identifiers are random sequences that can be used to correct for biases in amplification prior to sequencing.

[00299] In methods of the present disclosure, a constant nucleic acid sequence to minimize ligation bias comprises between about 1 nucleotide and about 15 nucleotides. Preferably, the constant sequence comprises about 8 nucleotides.

[00300] In some aspects, a flow cell binding site can comprise between about 15 to about 40 nucleotides. A flow cell binding site can comprise about 29 nucleotides. A flow cell binding site can comprise about 24 nucleotides.

[00301] In some aspects, a target binding domain can comprise between about 10 to about 70 nucleotides. A target binding domain can comprise between about 30 to about 55 nucleotides. A target binding domain can comprise between about 35 to about 50 nucleotides.

[00302] In some aspects, a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain can comprise between about 20 to about 40. A unique nucleic acid sequence which identifies the target analyte bound to the target binding domain can comprise about 25 nucleotides, or about 35 nucleotides, or about 12 nucleotides.

[00303] In some aspects, an amplification primer binding site can comprise between about 20 to about 50 nucleotides. An amplification primer binding site can comprise about 33 nucleotides, or about 34 nucleotides.

[00304] In some aspects, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can comprise between about 1 to about 20 nucleotides. A nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can comprise about 8 nucleotides.

[00305] In some aspects, a nucleic acid sequence comprising a unique molecular identifier can comprise between about 5 to about 20 nucleotides. A nucleic acid sequence comprising a unique molecular identifier can comprise about 14 nucleotides.

[00306] As used herein, the terms "region of interest" and "ROI" are used in their broadest sense to refer to a specific location within a sample that is to be analyzed using the methods of the present disclosure.

[00307] As used herein, the term "adjacent" can mean within about 1 nucleotide, or within about 2 nucleotides, or within about 3 nucleotides, or within about 4 nucleotides, or within about 5 nucleotides, or within about 6 nucleotides, or within about 7 nucleotides, or within about 8 nucleotides, or within about 9 nucleotides, or within about 10 nucleotides, or within about 11 nucleotides, or within about 12 nucleotides, or within about 13 nucleotides, or within about 14 nucleotides, or within about 15 nucleotides, or within about 16 nucleotides, or within about 17 nucleotides, or within about 18 nucleotides, or within about 19 nucleotides, or within about 20 nucleotides, or within about 21 nucleotides, or within about 22 nucleotides, or within about 23 nucleotides, or within about 24 nucleotides, or within about 25 nucleotides, or within about 26 nucleotides, or within about 27 nucleotides, or within about 28 nucleotides, or within about 29 nucleotides, or within about 30 nucleotides, or within about 40 nucleotides, or within about nucleotides, or within about 50 nucleotides, or within about 60 nucleotides, or within about 70 nucleotides, or within about 80 nucleotides, or within about 90 nucleotides, or within about 100 nucleotides.

[00308] As used herein, the term "spatially detecting" is used in its broadest sense to refer to the identification of the presence of a specific target analyte within a specific region of interest in a sample. Spatially detecting can comprise quantifying the amount of a specific target analyte present within a specific region of interest in a sample. Spatially detecting can further comprise quantifying the relative amount of a first target analyte within a specific region of interest in a sample as compared to the amount of at least a second target analyte within a specific region of interest in a sample. Spatially detecting can also comprise quantifying the relative amount of a specific target analyte within a first region of interest in a sample compared to the amount of the same target analyte in at least a second region of interest in the same sample or different sample.

[00309] In some aspects of the methods and compositions of the present disclosure, a target analyte can be any molecule within a sample that is to be spatially detected. Target analytes include, but are not limited to, nucleic acid molecules and protein molecules. When the target analyte is a protein, the protein can be referred to as a target protein. When the target analyte is a nucleic acid, the nucleic acid can be referred to as a target nucleic acid. Target nucleic acids can include, but

are not limited to, mRNA molecules, micro RNA (miRNA) molecules, tRNA molecules, rRNA molecules, gDNA or any other nucleic acid present within a sample.

[00310] In some aspects of the methods and compositions of the present disclosure, the term target binding domain is used in its broadest sense to refer to a portion of a probe of the present disclosure that binds to, either directly or indirectly, a target analyte located in a sample. A target binding domain can comprise nucleic acid, protein, at least one antibody, an aptamer, or any combination thereof. A target binding domain can comprise DNA, RNA or any combination thereof. A target binding domain can comprise any number of modified nucleotides and/or nucleic acid analogues.

[00311] In the aspect that the target analyte to be spatially detected is a target protein, a target binding domain can be a protein-target binding domain. A protein-target binding domain can comprise an antibody or antibody fragment that binds to the target protein.

[00312] In the aspect that the target analyte to be spatially detected is a target nucleic acid, a target binding domain can be a target nucleic acid-binding region. A target nucleic acid-binding region can comprise a nucleic acid that is complementary to the target nucleic acid to be spatially detected. A target nucleic acid-binding region can comprise a nucleic acid that hybridizes to the target nucleic acid to be detected.

[00313] As used herein, the term "hybridize" is used in its broadest sense to mean the formation of a stable nucleic acid duplex. In one aspect, "stable duplex" means that a duplex structure is not destroyed by a stringent wash under conditions such as, for example, a temperature of either about 5 °C below or about 5 °C above the Tm of a strand of the duplex and low monovalent salt concentration, *e.g.,* less than 0.2 M, or less than 0.1 M or salt concentrations known to those of skill in the art. A duplex can be "perfectly matched", such that the polynucleotide and/or oligonucleotide strands making up the duplex form a double stranded structure with one another such that every nucleotide in each strand undergoes Watson-Crick base pairing with a nucleotide in the other strand. The term "duplex" comprises, but is not limited to, the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2-aminopurine bases, PNAs, and the like, that can be employed. A duplex can comprise at least one mismatch, wherein the term "mismatch" means that a pair of nucleotides in the duplex fail to undergo Watson-Crick bonding.

[00314] As used herein, the term "hybridization conditions," will typically include salt concentrations of less than about 1 M, more usually less than about 500 mM and even more usually less than about 200 mM. Hybridization temperatures can be as low as 5 °C, but are typically greater

than 22 °C, more typically greater than about 30 °C, and often in excess of about 37 °C. Hybridizations are usually performed under stringent conditions, *e.g.,* conditions under which a probe will specifically hybridize to its target analyte. Stringent conditions are sequence-dependent and are different in different circumstances. Longer fragments can require higher hybridization temperatures for specific hybridization. As other factors can affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents and extent of base mismatching, the combination of parameters is more important than the absolute measure of any one alone. Certain hybridization conditions will promote the formation of a duplex between the entire length of a target binding domain and the target analyte. Other hybridization conditions will promote the formation of a duplex only along certain portions of the target binding domain.

[00315] In some aspects of the methods and compositions of the present disclosure, a probe can comprise a target binding domain directly, or indirectly linked to an identifier oligonucleotide. In the context of a probe, an identifier oligonucleotide is a polynucleotide that comprises a nucleic acid sequence that identifies the target analyte bound to the target binding domain of that probe. That is to say, the identifier oligonucleotide comprises a specific nucleic acid sequence that is *a priori* assigned to the specific target analyte bound to the target binding to which the identifier oligonucleotide is attached. In a non-limiting example, a probe designated as "probe *X*" designed to spatially detect "target analyte *X*" comprises a target binding domain designated "target binding domain *X*" linked to an identifier oligonucleotide designated "identifier oligonucleotide *X*". Target binding domain *X* binds to target analyte *X* and identifier oligonucleotide *X* comprises a nucleic acid sequence, designated as "nucleic acid sequence *X*", which corresponds to target analyte *X*. Thus, if a skilled artisan practicing the methods of the present disclosure were to collect identifier oligonucleotides released from a region of interest in sample and obtain nucleic acid sequence *X* after sequencing, the skilled artisan would understand that to indicate that target analyte *X* was present in that region of interest. The amount, or number of sequencing reads, of nucleic acid sequence *X* can be used to determine the quantify, in absolute or relative terms, the amount of target analyte *X* within the region of interest.

[00316] As used herein, the term "amplification primer binding site" is used in its broadest sense to refer to a nucleic acid sequence that is complementary to, or at least partially complementary to at least one amplification primer, wherein the amplification primer is a short single-stranded or

partially single-stranded oligonucleotide that is sufficient to prime DNA and/or RNA synthesis, for example, by PCR.

[00317] In some aspects of the methods and compositions of the present disclosure, a target binding domain can be linked to an identifier oligonucleotide by a cleavable linker. Suitable cleavable linkers include, but are not limited to, chemically cleavable linkers (*e.g.* a linker that is cleaved when exposed to a particular chemical, combination of chemicals or reaction conditions), a photo-cleavable linker (*e.g.* a linker that is cleaved when exposed to light of a sufficient wavelength or light comprising a sufficient range of wavelengths), or an enzymatically cleavable linker (*e.g.* a linker that is cleaved by a specific enzyme or class of enzymes). Thus, as used herein the phrase "providing a force to a location of the sample sufficient to release an identifier oligonucleotide" is used in its broadest sense to describe changing the conditions within a certain region of interest in a sample such that, for any probe bound to a target analyte within that region of interest, the linker between the target binding domain of the probe and the identifier oligonucleotide of the probe is cleaved, thereby separating the identifier oligonucleotide from the target binding domain so that the identifier oligonucleotide can be subsequently collected from solution. For example, in aspects wherein a probe comprises a chemically cleavable linker between the target binding domain and the identifier oligonucleotide, providing a force to a location of the sample sufficient to release an identifier oligonucleotide can comprise exposing that location of the sample to the specific chemical, combination of chemicals or reaction conditions that catalyze the cleavage of the linker. In another non-limiting example, in aspects wherein a probe comprises a photo-cleavable linker between the target binding domain and the identifier oligonucleotide, providing a force to a location of the sample sufficient to release an identifier oligonucleotide can comprise exposing/exciting that location of the sample with light of a sufficient wavelength capable of cleaving the photo-cleavable linker. In another non-limiting example, in aspects wherein a probe comprises an enzymatically cleavable linker between the target binding domain and the identifier oligonucleotide, providing a force to a location of the sample sufficient to release an identifier oligonucleotide can comprise exposing that location of the sample to an amount of enzyme sufficient to catalyze the cleavage of the linker.

[00318] Providing a force to a location of the sample sufficient to release an identifier oligonucleotide can result in at least about 10%, or at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least

about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 99% of probes bound to target analytes within that location of the sample to undergo cleavage of the linker connecting the target binding domain and the identifier oligonucleotide.

[00319] As would be appreciated by one skilled in the art, the term "unique molecular identifier" or "UMI" refer to short nucleic acid sequences that are used to quantify and reduce quantitative bias caused by nucleic acid amplification prior to sequencing reactions.

[00320] In some aspects of the methods and compositions of the present disclosure, an affinity moiety can comprise biotin, avidin, streptavidin, nucleic acid, or any combination thereof.

[00321] In some aspects of the methods and compositions of the present disclosure, a probe can comprise at least at least about 5, about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00322] In some aspects of the methods and compositions of the present disclosure, a target binding domain can comprise at least at least about 5, about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00323] In some aspects of the methods and compositions of the present disclosure, an identifier oligonucleotide can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00324] In some aspects of the methods and compositions of the present disclosure, an amplification primer can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00325] In some aspects of the methods and compositions of the present disclosure, a nucleic acid probe can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00326] In some aspects of the methods and compositions of the present disclosure, a nucleic acid complementary to a portion of a identifier oligonucleotide can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95 or at least about 100 nucleotides.

[00327] In some aspects of the methods and compositions of the present disclosure, a nucleic acid sequence comprising a molecular identifier can comprise at least about 5, or at least about 10 nucleotides, or at least about 15, or at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 nucleotides.

[00328] In some aspects of the methods and compositions of the present disclosure, an amplification primer binding site can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35, at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65 or at least about 70 nucleotides.

[00329] In some aspects of the methods and compositions of the present disclosure, a flow cell adapter sequence suitable for sequencing can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95 or at least about 100 nucleotides.

[00330] In some aspects of the methods and compositions of the present disclosure, a flow cell binding site can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35, at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 90, at least about 95 or at least about 100 nucleotides.

[00331] In some aspects of the methods and compositions of the present disclosure, a nucleic acid sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released can comprise at least about 5, or at least about 10 nucleotides, or at least about 15, or at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 nucleotides.

[00332] In some aspects of the methods and compositions of the present disclosure, a unique nucleic acid sequence which identifies the target analyte bound to a target binding domain can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35, at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 90, at least about 95 or at least about 100 nucleotides.

[00333] In some aspects of the methods and compositions of the present disclosure, a probe, a target binding domain, an identifier oligonucleotide, an amplification primer, a nucleic acid probe, a nucleic acid complementary to a portion of a identifier oligonucleotide, a nucleic acid sequence comprising a molecular identifier, an amplification primer binding site, a flow cell adapter sequence, a flow cell binding site, a nucleic acid sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released, a unique nucleic acid sequence which identifies the target analyte bound to a target binding domain or any combination thereof can comprise at least one natural base, can comprise no natural bases, can comprise at least one modified nucleotide or nucleic acid analog, can comprise no modified

nucleotides or nucleic acid analogs, can comprise at least one universal base, can comprise no universal bases, can comprise at least one degenerate base or can comprise no degenerate bases.

[00334] In some aspects of the methods and compositions of the present disclosure, a probe, a target binding domain, an identifier oligonucleotide, an amplification primer, a nucleic acid probe, a nucleic acid complementary to a portion of a identifier oligonucleotide, a nucleic acid sequence comprising a molecular identifier, an amplification primer binding site, a flow cell adapter sequence, a flow cell binding site, a nucleic acid sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released, a unique nucleic acid sequence which identifies the target analyte bound to a target binding domain or any combination thereof can comprise any combination natural bases (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more natural bases), modified nucleotides or nucleic acid analogs (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more modified or analog nucleotides), universal bases (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more universal bases), or degenerate bases (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more degenerative bases). When present in a combination, the natural bases, modified nucleotides or nucleic acid analogs, universal bases and degenerate bases can be arranged in any order.

[00335] The terms "modified nucleotides" or "nucleic acid analogues" include, but are not limited to, locked nucleic acids (LNA), bridged nucleic acids (BNA), propyne-modified nucleic acids, zip nucleic acids (ZNA®), isoguanine and isocytosine. Preferably, the modified nucleotides or nucleic acid analogues are locked nucleic acids (LNAs).

[00336] The term "locked nucleic acids (LNA)" as used herein includes, but is not limited to, a modified RNA nucleotide in which the ribose moiety comprises a methylene bridge connecting the 2' oxygen and the 4' carbon. This methylene bridge locks the ribose in the 3'-endo confirmation, also known as the north confirmation, that is found in A-form RNA duplexes. The term inaccessible RNA can be used interchangeably with LNA. The term "bridged nucleic acids (BNA)" as used herein includes, but is not limited to, modified RNA molecules that comprise a five-membered or six-membered bridged structure with a fixed 3'-endo confirmation, also known as the north confirmation. The bridged structure connects the 2' oxygen of the ribose to the 4' carbon of the ribose. Various different bridge structures are possible containing carbon, nitrogen, and hydrogen atoms. The term "propyne-modified nucleic acids" as used herein includes, but is not limited to, pyrimidines, namely cytosine and thymine/uracil, that comprise a propyne modification at the C5 position of the nucleic acid base. The term "zip nucleic acids

(ZNA®)" as used herein includes, but is not limited to, oligonucleotides that are conjugated with cationic spermine moieties.

[00337] The term "universal base" as used herein includes, but is not limited to, a nucleotide base does not follow Watson-Crick base pair rules but rather can bind to any of the four canonical bases (A, T/U, C, G) located on the target nucleic acid. The term "degenerate base" as used herein includes, but is not limited to, a nucleotide base that does not follow Watson-Crick base pair rules but rather can bind to at least two of the four canonical bases A, T/U, C, G), but not all four. A degenerate base can also be termed a Wobble base; these terms are used interchangeably herein.

[00338] As used in this Specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

[00339] Unless specifically stated or obvious from context, as used herein, the term "or" is understood to be inclusive and covers both "or" and "and".

[00340] Unless specifically stated or obvious from context, as used herein, the term "about" is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from the context, all numerical values provided herein are modified by the term "about."

[00341] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although other probes, compositions, methods, and kits similar, or equivalent, to those described herein can be used in the practice of the present disclosure, the preferred materials and methods are described herein. It is to be understood that the terminology used herein is for the purpose of describing particular aspects only, and is not intended to be limiting.

EXAMPLES

[00342] Example 1—two-ended adapter ligation method for 96 multiplexed samples

[00343] In this example, a two-ended adapter ligation method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. The nucleic acid adapters used in this experiment were partially double-stranded. The nucleic acid adapters comprised a first strand and a second strand. The first strand comprised a 5' phosphate moiety for ligation. The first strand also comprised a constant nucleic acid sequence to minimize ligation

bias (GCGTAGTG), a nucleic acid sequence comprising a unique molecular identifier, a unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a first amplification primer binding site (SEQ ID NO: 2). The second strand comprised a single overhanging thymine nucleotide at the 3' end, a sequence complementary to the constant nucleic acid sequence to minimize ligation bias present in the first strand, a sequence complementary to the unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released present in the first strand and a second amplification primer binding site (SEQ ID NO: 1).

[00344] To form the partially double stranded nucleic acid adapters, first strand oligonucleotides and second strand oligonucleotides were combined in equimolar proportion for a final total oligonucleotide concentration of 28 μM in buffer comprising 50 mM NaCl. The oligonucleotide mixture was heated at 95°C for 2 minutes and cooled at ambient temperature for 30 minutes, thereby annealing the first stand and second strand oligonucleotides together to form the partially double-stranded nucleic acid adapters. Annealed nucleic acid adaptors were diluted to final concentration ranging between 0.02 μM to 0.002 μM in a solution of 10 mM Tris pH 8 and 0.05% Tween20.

[00345] Collected identifier oligonucleotides were end repaired and A-tailed using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) with a modified protocol. End repair/A-tail master mix was prepared by combining the following: 627.8 μL of PCR-grade H2O, 143.9 μL of NEBNext Ultra II End Prep Reaction Buffer, and 61.7 μL of NEBNext Ultra II End Prep Enzyme Mix. 8.3 μL of end repair/A-tail master mix was added to 4 μL of each sample of identifier oligonucleotides. The reaction was incubated for 30 minutes at 20°C with a heated lid of >75°C, followed by a second incubation for 30 minutes at 65°C. The repaired/A-tailed identifier oligonucleotide mixtures were then stored at 4°C.

[00346] Following end repair and A-tailing, the nucleic acid adaptors were ligated to the repaired/A-tailed identifier oligonucleotides by adding 6.4 μL of NEBNext Ultra II Ligation Master Mix, 0.2 μL of NEBNext Ligation Enhancer, and 1 μL of the nucleic acid adapter dilution to each repaired/A-tailed identifier oligonucleotide mixture. These reactions were incubated for 15 minutes at 20°C with the heated lid off and subsequently quenched with 1 μL

0.5M EDTA. All of the reactions were then pooled into a single 15 mL conical tube to form a pooled adapter-ligated sample.

[00347] The pooled adaptor-ligated sample was purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 350 µL of AMPure XP beads and 3.15 mL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). AMPure XP bead cleanup was performed with 3.5 mL of diluted AMPure XP beads and eluted in 200 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. AMPure XP bead cleanup was then repeated with 400 µL of AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20 to obtain purified adapter-ligated samples.

[00348] Following AMPure XP cleanup, PCR reactions with purified adaptor-ligated sample were prepared to amplify the adapter-ligated identifier oligonucleotides. To 6 µL of the purified adapter-ligated sample, 10 µL of NEBNext Ultra II Q5 Master Mix, 0.2 µL of 100 µM forward and reverse primers, and 3.6 µL of PCR-grade H2O was added. The forward primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the first amplification primer binding site located on the first strand of the nucleic acid adapter. Table 1 provides the sequences of the forward primers used.

Table 1. Forward primers for two-ended adapter ligation.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACGCTCAGATATAGCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	5
AATGATACGGCGACCACCGAGATCTACACGCTCAGAATAGAGGCACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	6
AATGATACGGCGACCACCGAGATCTACACGCTCAGACCTATCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	7
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGGCTCTGAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	8
AATGATACGGCGACCACCGAGATCTACACGCTCAGAAGGCGAAGACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	9
AATGATACGGCGACCACCGAGATCTACACGCTCAGATAATCTTAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	10
AATGATACGGCGACCACCGAGATCTACACGCTCAGACAGGACGTACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	11
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGTACTGACACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	12

[00349] The reverse primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the second

amplification primer binding site located on the second strand of the nucleic acid adapter. Table 2 provides the sequences of the reverse primers used.

Table 2. Reverse primers for two-ended adapter ligation

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	13
CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	14
CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	15
CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	16
CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	17
CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	18
CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	19
CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	20
CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	21
CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	22
CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	23
CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	24

[00350] The optimal number of PCR cycles was determined empirically with triplicate PCR reactions. Alternately, the optimal number of PCR cycles could have been determined using real-time/qPCR. The PCR program used comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 1 minute at 65°C

(4) Repeating steps (2) and (3) nine times

(5) 5 minutes at 65°C

[00351] The amplified products were purified using 18 µL of AMPure XP beads and eluting with 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00352] The amplified products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The amplified products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with a custom spike-in primer comprising the nucleotide sequence ACACTCTTTAAGACGACGTCGCTATGGCCTCTCC (SEQ ID NO: 25).

[00353] Example 2—one-ended adapter ligation method for 96 multiplexed samples

[00354] In this example, a one-ended adapter ligation method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. The nucleic acid adapters used in this experiment were partially double-stranded. The nucleic acid adapters comprised a first strand and a second strand. The first strand comprised a 5' phosphate moiety for ligation. The first strand also comprised a constant nucleic acid sequence to minimize ligation bias (CACTACGC), a nucleic acid sequence comprising a unique molecular identifier, a unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a first amplification primer binding site (SEQ ID NO: 1). The second strand comprised a single overhanging thymine nucleotide at the 3' end, a sequence complementary to the constant nucleic acid sequence to minimize ligation bias present in the first strand and a sequence complementary to the unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released present in the first strand.

[00355] To form the partially double stranded nucleic acid adapters, first strand oligonucleotides and second strand oligonucleotides were combined in equimolar proportion for a final total oligonucleotide concentration of 28µM in 50 mM NaCl. The oligonucleotide mixture was heated to 95°C for 2 minutes and cooled at ambient temperature for 30 minutes, thereby annealing the first stand and second strand oligonucleotides together to form the partially double-stranded nucleic acid adapters. Annealed nucleic acid adaptors were diluted to a final concentration ranging between 0.02 µM to 0.002 µM in a solution of 10 mM Tris pH 8 and 0.05% Tween20.

[00356] Nucleic acid adapters were ligated to the collected identifier oligonucleotides by addition of 10 µL of 2X rapid ligation buffer (Enzymatics), 1 µL of T4 DNA Rapid Ligase (Enzymatics),

and 1 µL of annealed nucleic acid adapter dilutions to each sample of collected identifier oligonucleotides. Samples were incubated for 15 minutes at 20°C with the heated lid off and subsequently quenched with 1 µL 0.5M EDTA. All of the reactions were then pooled into a single 15 mL conical tube to form a pooled adapter-ligated sample.

[00357] The pooled adaptor-ligated sample was purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 350 µL of AMPure XP beads and 3.15 mL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). AMPure XP bead cleanup was performed with 3.5 mL of diluted AMPure XP beads and eluted in 200 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. AMPure XP bead cleanup was then repeated with 400 µL of AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20 to obtain purified adapter-ligated samples.

[00358] Following AMPure XP cleanup, PCR reactions with purified adaptor-ligated sample were prepared to amplify the adapter-ligated identifier oligonucleotides. To 6 µL of the purified adapter-ligated sample, 10 µL of NEBNext Ultra II Q5 Master Mix, 0.2 µL of 100 µM forward and reverse primers, and 3.6 µL of PCR-grade H_2O was added. The forward primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the first amplification primer binding site located on the first strand of the nucleic acid adapter. Table 3 provides the sequences of the forward primers used.

Table 3. Forward primers for two-ended adapter ligation.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACGCTCAGATATAGCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	5
AATGATACGGCGACCACCGAGATCTACACGCTCAGAATAGAGGCACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	6
AATGATACGGCGACCACCGAGATCTACACGCTCAGACCTATCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	7
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGGCTCTGAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	8
AATGATACGGCGACCACCGAGATCTACACGCTCAGAAGGCGAAGACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	9
AATGATACGGCGACCACCGAGATCTACACGCTCAGATAATCTTAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	10
AATGATACGGCGACCACCGAGATCTACACGCTCAGACAGGACGTACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	11
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGTACTGACACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	12

[00359] The reverse primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the second amplification primer binding site located on the second strand of the nucleic acid adapter. Table 4 provides the sequences of the reverse primers used.

Table 4. Reverse primers for two-ended adapter ligation.

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	13
CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	14
CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	15
CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	16
CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	17
CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	18
CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	19
CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	20
CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	21
CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	22
CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	23
CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	24

[00360] The optimal number of PCR cycles was determined empirically with triplicate PCR reactions. Alternately, the optimal number of PCR cycles could have been determined using real-time/qPCR. The PCR program used comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 1 minute at 65°C

(4) Repeating steps (2) and (3) nine times

(5) 5 minutes at 65°C

[00361] The amplified products were purified using 18 µL of AMPure XP beads and eluting with 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00362] The amplified products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The amplified products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with a custom spike-in primer comprising the nucleotide sequence ACACTCTTTAAGACGACGTCGCTATGGCCTCTCC (SEQ ID NO: 25).

[00363] Example 3—templated-primer extension method for 96 multiplexed samples

[00364] In this example, a templated-primer extension method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. The single stranded nucleic acid templates used in this example comprised a 3' biotin moiety, a region complementary to the unique nucleic acid sequences present in the collected identifier oligonucleotides, a nucleic acid sequence comprising a unique molecular identifier and a second amplification primer binding sequence (GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT, SEQ ID NO: 26). Table 5 provides the sequences of the single-stranded nucleic acid templates used in this example.

Table 5. Single stranded nucleic acid templates for templated-primer extension method

Single-stranded nucleic acid templates	SEQ ID NO
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNTTGAAGCACACCGTTTTTCTTTCTTCTTTCACGG	27
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNACCCACAGGTTATACGGGATTATCCGGTTATCCA	28
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNCGACACCGAGTTCGACCGTTATGTTGGTAGGATC	29
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNCGGTGTGTAAGCGTAACGATGTTGGTGTCGCTCT	30
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNCAGACACTGCGACAACTCACGATCATGACACAGA	31
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNATATTCTGTACTCAGTGCCTATCCACCTAATAGG	32
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNTTCAGTTATAATGTGTCCAGCAGAAGCAGGAATT	33
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNGTCCTTTGTTGGGCGGACCGTAATGAGGAATTTG	34
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNGATGAGACTTCTACATGTCCGATGTTTTTGTGCT	35

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNACTC ACACATAGTACTGACACGTAAGATAGGATG	36
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTTAC CCTATCTCGTCTATGTACGTCAGGCTGAAT	37
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNATCA ACGTAGGGTAAGGTCATATTTTTACCTTAC	38
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTTCC CTCTTTCTCCGCTTATGGATGAAAGGACAG	39
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCCTG CACAGTGAGTTTCTTTCACTCTAACTCTCT	40
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTGTC GCTCTAGTGTGACTTTTCCACCTCGCATCT	41
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNATAT CTTTCTCGGGTAAAGATTAGGCGTCCGATA	42
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCGAT TAGCCGTAGACGCAACTCATTGCCGAAGAT	43
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTGTG AGCATTTCAGTACGAGTGATGCAGATAAAC	44
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTATA GTTACCAAGTACTATGGGTTGGTGGAAGCC	45
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCCAA TTATACTGTCTGTTATGTTCTCGGATAAGC	46
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTCAG GTGCTTGTAGGCTCATGATAGGGGTAATGC	47
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCTCT GCTGTAATCTCAGCTCCACTTGTTTCTAAG	48
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNGTGC ATATTGCAGCTGAGCCAGCTCAATTTGAAG	49
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCCGT TGATTTACGCAACAGCGGCTTATATAGCTC	50
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCATC ATCGACAGTTCGCAGCCCTATAACATGATA	51
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNATCG CAGGATGGTACAGCATCATACATGATGAGC	52
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCTGA TAAGTCGTAGGAATGTCGCTTAATACGGAT	53
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNATGG CGGTTTCGGGTCCTGCACTATTCCTAATAA	54
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCCAG TACGGGTACTAATAAGTGTCATATCTATTG	55
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTGTT GGAGAGGTTAGAGGTGAGGAGGCGAAGATA	56

[00365] Single stranded nucleic acid templates were ordered from Integrated DNA Technologies, Inc. and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). Individual single stranded nucleic acid templates were normalized to a standard concentration

and then pooled to be equimolar. The pool of single stranded nucleic acid templates was diluted to 0.83 nM in a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00366] The collected identifier oligonucleotides were hybridized to the single stranded nucleic acid templates and extended by addition of 10 µL of NEBNext Ultra II Q5 Master Mix (New England Biolabs), 4 µL of the diluted single stranded nucleic acid template pool and 4 µL of H_2O to 2 µL of each sample of identifier oligonucleotides. The following PCR program was used to extend the identifier oligonucleotides:

(1) 30 seconds at 98°C, 10×

(2) 1 minutes at 98°C,

(3) 1 minutes at 68°C

(4) 1 minutes at 72°C

(5) Repeating steps (2)-(4) ten times

(6) 2 minutes at 72°C

[00367] The extension products were stored at 4°C. Magnetic streptavidin beads (MyOne Streptavidin C1 beads, Thermo Fisher Scientific) were washed in 1X Binding and Washing Buffer (5 mM Tris-HCl, 0.5 mM EDTA, 1M NaCl), and 5 µL of streptavidin beads were added to each extension product sample. The extension product samples were incubated with the beads on an orbital mixer for a minimum of 15 minutes. Following incubation, the samples were heated to 95°C for 3 minutes and transferred to a magnetic plate. Supernatant was extracted immediately after sufficient bead pelleting to yield the purified extension product samples.

[00368] The purified extension product samples were amplified by adding to 7.5 µL of each purified extension product sample, 12.5 µL of NEBNext Ultra II Q5 Master Mix, 0.25 µL of 100 µM forward primer, 1 µL of 25 µM reverse primer and 3.8 µL of PCR-grade H_2O. The forward primer comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the first amplification primer binding site located on the identifier oligonucleotide. Table 6 provides the sequences of the forward primers used in this example.

Table 6. Forward primers for templated-primer extension method.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACGCTCAGATATAGCCTACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	5
AATGATACGGCGACCACCGAGATCTACACGCTCAGAATAGAGGCACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	6
AATGATACGGCGACCACCGAGATCTACACGCTCAGACCTATCCTACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	7
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGGCTCTGAACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	8
AATGATACGGCGACCACCGAGATCTACACGCTCAGAAGGCGAAGACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	9
AATGATACGGCGACCACCGAGATCTACACGCTCAGATAATCTTAACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	10
AATGATACGGCGACCACCGAGATCTACACGCTCAGACAGGACGTACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	11
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGTACTGACACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	12

[00369] The reverse primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the second amplification primer binding site located single-stranded nucleic acid template. Table 7 provides the sequences of the reverse primers used.

Table 7. Reverse primers for templated-primer extension method.

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATGTCGGTAAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	57
CAAGCAGAAGACGGCATACGAGATAGGTCACTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	58
CAAGCAGAAGACGGCATACGAGATGAATCCGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	59
CAAGCAGAAGACGGCATACGAGATGTACCTTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	60
CAAGCAGAAGACGGCATACGAGATCATGAGGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	61
CAAGCAGAAGACGGCATACGAGATTGACTGACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	62
CAAGCAGAAGACGGCATACGAGATCGTATTCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	63
CAAGCAGAAGACGGCATACGAGATCTCCTAGAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	64
CAAGCAGAAGACGGCATACGAGATTAGTTGCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	65

CAAGCAGAAGACGGCATACGAGATGAGATACGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	66
CAAGCAGAAGACGGCATACGAGATAGGTGTACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	67
CAAGCAGAAGACGGCATACGAGATTAATGCCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	68
CAAGCAGAAGACGGCATACGAGATTCAGACGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	69
CAAGCAGAAGACGGCATACGAGATGATAGGCTGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	70
CAAGCAGAAGACGGCATACGAGATTGGTACAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	71
CAAGCAGAAGACGGCATACGAGATCAAGGTCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	72
CAAGCAGAAGACGGCATACGAGATGCTATCCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	73
CAAGCAGAAGACGGCATACGAGATATGGAAGGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	74
CAAGCAGAAGACGGCATACGAGATTCAAGGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	75
CAAGCAGAAGACGGCATACGAGATGTTACGCAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	76
CAAGCAGAAGACGGCATACGAGATAGTCTGTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	77
CAAGCAGAAGACGGCATACGAGATGCACGTAAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	78
CAAGCAGAAGACGGCATACGAGATAACCTTGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	79
CAAGCAGAAGACGGCATACGAGATATTGCGTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	80
CAAGCAGAAGACGGCATACGAGATACCTGGAAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	81
CAAGCAGAAGACGGCATACGAGATGGAGATGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	82
CAAGCAGAAGACGGCATACGAGATGTACTCTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	83
CAAGCAGAAGACGGCATACGAGATGTAACGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	84
CAAGCAGAAGACGGCATACGAGATATTCCTCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	85
CAAGCAGAAGACGGCATACGAGATGTGTTCCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	86
CAAGCAGAAGACGGCATACGAGATAAGCACTGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	87
CAAGCAGAAGACGGCATACGAGATCTAGCAAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	88
CAAGCAGAAGACGGCATACGAGATTGCTTCCAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	89
CAAGCAGAAGACGGCATACGAGATGCTTAGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	90

CAAGCAGAAGACGGCATACGAGATAACCGTTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	91
CAAGCAGAAGACGGCATACGAGATGACATTCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	92
CAAGCAGAAGACGGCATACGAGATAGACCGTAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	93
CAAGCAGAAGACGGCATACGAGATGATACTGGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	94
CAAGCAGAAGACGGCATACGAGATTGCGTAGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	95
CAAGCAGAAGACGGCATACGAGATTCGGTTACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	96
CAAGCAGAAGACGGCATACGAGATATGACGTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	97
CAAGCAGAAGACGGCATACGAGATGCTGTAAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	98
CAAGCAGAAGACGGCATACGAGATGCAATGGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	99
CAAGCAGAAGACGGCATACGAGATATCTCGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	100
CAAGCAGAAGACGGCATACGAGATGGCTATTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	101
CAAGCAGAAGACGGCATACGAGATGGTGTCTTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	102
CAAGCAGAAGACGGCATACGAGATTCAACTGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	103
CAAGCAGAAGACGGCATACGAGATCTTCACCAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	104
CAAGCAGAAGACGGCATACGAGATACGGTCTTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	105
CAAGCAGAAGACGGCATACGAGATTCTCGCAAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	106
CAAGCAGAAGACGGCATACGAGATGGAATTGCGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	107
CAAGCAGAAGACGGCATACGAGATACGGATTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	108
CAAGCAGAAGACGGCATACGAGATTTAAGCGGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	109
CAAGCAGAAGACGGCATACGAGATTGCAGGTAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	110
CAAGCAGAAGACGGCATACGAGATCAATCGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	111
CAAGCAGAAGACGGCATACGAGATGTGCCATAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	112
CAAGCAGAAGACGGCATACGAGATTGTTCGAGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	113
CAAGCAGAAGACGGCATACGAGATTGGAGTTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	114
CAAGCAGAAGACGGCATACGAGATACGATGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	115

CAAGCAGAAGACGGCATACGAGATTGATGTCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	116
CAAGCAGAAGACGGCATACGAGATTGAACCTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	117
CAAGCAGAAGACGGCATACGAGATCTTCGTTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	118
CAAGCAGAAGACGGCATACGAGATCTTCTGAGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	119
CAAGCAGAAGACGGCATACGAGATTGCTCATGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	120
CAAGCAGAAGACGGCATACGAGATAGTTCGTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	121
CAAGCAGAAGACGGCATACGAGATTAGCGTCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	122
CAAGCAGAAGACGGCATACGAGATGGCGTTATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	123
CAAGCAGAAGACGGCATACGAGATGGTGATTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	124
CAAGCAGAAGACGGCATACGAGATAACTTGCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	125
CAAGCAGAAGACGGCATACGAGATGCAAGATCGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	126
CAAGCAGAAGACGGCATACGAGATTCGCATTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	127
CAAGCAGAAGACGGCATACGAGATTGTACACCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	128
CAAGCAGAAGACGGCATACGAGATAGCTCCTAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	129
CAAGCAGAAGACGGCATACGAGATGCAATTCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	130
CAAGCAGAAGACGGCATACGAGATCTTAGGACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	131
CAAGCAGAAGACGGCATACGAGATGTCCTAAGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	132
CAAGCAGAAGACGGCATACGAGATAACGTGGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	133
CAAGCAGAAGACGGCATACGAGATCTGTGTTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	134
CAAGCAGAAGACGGCATACGAGATGTTAAGGCGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	135
CAAGCAGAAGACGGCATACGAGATCACCTTACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	136
CAAGCAGAAGACGGCATACGAGATTGGTAGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	137
CAAGCAGAAGACGGCATACGAGATCAGTGAAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	138
CAAGCAGAAGACGGCATACGAGATGTTCAACCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	139
CAAGCAGAAGACGGCATACGAGATTGGCTATCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	140

CAAGCAGAAGACGGCATACGAGATCTGGAGTAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	141
CAAGCAGAAGACGGCATACGAGATTCTCTTCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	142
CAAGCAGAAGACGGCATACGAGATTCTAACGCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	143
CAAGCAGAAGACGGCATACGAGATGGTCAGATGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	144
CAAGCAGAAGACGGCATACGAGATCTCTGGTTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	145
CAAGCAGAAGACGGCATACGAGATTGTGGTACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	146
CAAGCAGAAGACGGCATACGAGATCCTATACCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	147
CAAGCAGAAGACGGCATACGAGATTTCTCTCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	148
CAAGCAGAAGACGGCATACGAGATGTATGCTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	149
CAAGCAGAAGACGGCATACGAGATAAGTCGAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	150
CAAGCAGAAGACGGCATACGAGATAACCGAAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	151
CAAGCAGAAGACGGCATACGAGATTGTTGTGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	152

[00370] The PCR program used to amplify the purified extension products comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) eighteen times

(6) 2 minutes at 72°C

[00371] The amplified extension products were stored at 4°C. 4 µL of each PCR reaction were combined into four pools, for 24 samples per pool.

[00372] The pooled PCR reactions were purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 100 µL of AMPure XP beads and 400 µL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). Purification was performed with 76.8 µL of diluted AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. Beads were retained and cleanup process was repeated

with 24 μL of AMPure XP buffer and eluted in 20 μL of buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00373] The purified PCR products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The purified PCR products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with a custom spike-in primer comprising the nucleotide sequence ACACTCTTTAAGACGACGTCGCTATGGCCTCTCC (SEQ ID NO: 25).

[00374] Example 4—long probe hybridization method for 96 multiplexed samples

[00375] In this example, a long probe hybridization method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. In this example, the first nucleic acid probe comprises a 5' phosphate moiety, a nucleic acid sequence complementary to a portion of the identifier oligonucleotide, a first amplification primer binding site comprising an i7 sequence (SEQ ID NO: 1), unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a P7 flow cell adapter sequence (SEQ ID NO: 4). The second nucleic acid probe comprises a nucleic acid sequence complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier, a second amplification primer binding site comprising and i5 sequence (SEQ ID NO: 2) and a P5 flow cell adapter sequence (SEQ ID NO: 3).

[00376] The first and second nucleic acid probes were ordered from Integrated DNA Technologies, Inc. and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). Individual nucleic acid probes were normalized to a standard concentration, pooled to be equimolar, and diluted to 0.83 nM in a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. The nucleic acid probes and the identifier oligonucleotides were hybridized by combining 0.5 μL of diluted nucleic acid probe pool with 2 μL of a mixture of identifier oligonucleotides collected from a sample solution in a buffer comprising 50 mM NaCl. This mixture was heated for 2 minutes at 95°C and cooled for 30 minutes at ambient temperature to yield an annealed identifier oligonucleotide-nucleic acid probe mixture.

[00377] In the case in which the first and the second nucleic acid probes hybridized to the identifier oligonucleotide such that the first and the second nucleic acid probes were not adjacent and were not overlapping, a gap extension reaction was performed. To 2.5 μL of each annealed

identifier oligonucleotide-nucleic acid probe mixture, 3.8 µL of NEBNext Ultra II Q5 Master mix and 1.3 µL of PCR-grade H_2O was added. The mixture was then subjected to the following Gap extension temperature cycle:

(1) 30 seconds at 98°C

(2) 1 minute at 98°C

(3) 1 minute at 68°C

(4) 1 minute at 72°C,

(5) Repeat steps (2)-(4) ten times

(6) 2 minutes at 72°C

[00378] The gap extension products were then stored at 4°C. The first and second nucleic acid probes were then ligated together by adding to 1 µL of the gap extension product, 10 µL of 2X rapid ligation buffer (Enzymatics), 1 µL of T4 DNA Rapid Ligase (Enzymatics), and 8 µL of PCR-grade H_2O. These ligation reactions were incubated for 15 minutes 20°C, subsequently quenched with 1 µL 0.5M EDTA, and pooled into a single 15 mL conical tube.

[00379] In the case in which the first and the second nucleic acid probes hybridized to the identifier oligonucleotide such that the first and the second nucleic acid probes were adjacent and were not overlapping, nick repair ligation reaction was performed. To 2.5 µL of each annealed identifier oligonucleotide-nucleic acid probe mixture, 10 µL of 2X rapid ligation buffer (Enzymatics), 1 µL of T4 DNA Rapid Ligase (Enzymatics), and 1 µL of PCR-grade H_2O was added. These ligation reactions were incubated for 15 minutes at 20°C, subsequently quenched with 1 µL of 0.5M EDTA, and pooled into a single 15 mL conical tube.

[00380] The pools of quenched ligation reactions were then purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 350 µL of AMPure XP beads and 3.15 mL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). The purification was performed with 3.5 mL of diluted AMPure XP beads and eluted in 200 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. The AMPure XP bead cleanup was then repeated with 400 µL of AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00381] To amplify the purified ligation products, PCR reactions with purified ligation products and primers were prepared. To 6 µL of purified ligation product, 10 µL of NEBNext Ultra II Q5 Master Mix, 0.2 µL of 100 µM forward primer (CAAGCAGAAGACGGCATACGA, SEQ ID

NO: 153) and reverse primer (AATGATACGGCGACCACCGA, SEQ ID NO: 154) and 3.6 µL of PCR-grade H_2O was added. The PCR program used to amplify the purified extension products comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) eighteen times

(6) 2 minutes at 72°C

[00382] The amplified products were stored at 4°C. 4 µL of each PCR reaction were combined into six pools, for 16 samples per pool. The amplified products were further purified using an AMPure XP bead cleanup with 64 µL of AMPure XP beads and eluting with 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00383] The purified amplified products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The purified amplified products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with either standard sequencing primers or a custom spike-in Read1 primer (SEQ ID NO: 25).

[00384] Example 5—short probe hybridization method for 96 multiplexed samples

[00385] In this example, a short probe hybridization method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. In this example, the first nucleic acid probe comprises a 5' phosphate moiety, a nucleic acid sequence complementary to a portion of the identifier oligonucleotide, a first amplification primer binding site comprising an i7 sequence (SEQ ID NO: 1) and a unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. The second nucleic acid probe comprises a nucleic acid sequence complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier and a second amplification primer binding site comprising an i5 sequence (SEQ ID NO: 2).

[00386] The first and second nucleic acid probes were ordered from Integrated DNA Technologies, Inc. and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher

Scientific). Individual nucleic acid probes were normalized to a standard concentration, pooled to be equimolar, and diluted to 0.83 nM in a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. The nucleic acid probes and the identifier oligonucleotides were hybridized by combining 0.5 μL of diluted nucleic acid probe pool with 2 μL of a mixture of identifier oligonucleotides collected from a sample solution in a buffer comprising 50 mM NaCl. This mixture was heated for 2 minutes at 95°C and cooled for 30 minutes at ambient temperature to yield an annealed identifier oligonucleotide-nucleic acid probe mixture.

[00387] In the case in which the first and the second nucleic acid probes hybridized to the identifier oligonucleotide such that the first and the second nucleic acid probes were not adjacent and were not overlapping, a gap extension reaction was performed. To 2.5 μL of each annealed identifier oligonucleotide-nucleic acid probe mixture, 3.8 μL of NEBNext Ultra II Q5 Master mix and 1.3 μL of PCR-grade H_2O was added. The mixture was then subjected to the following Gap extension temperature cycle:

(1) 30 seconds at 98°C

(2) 1 minute at 98°C

(3) 1 minute at 68°C

(4) 1 minute at 72°C,

(5) Repeat steps (2)-(4) ten times

(6) 2 minutes at 72°C

[00388] The gap extension products were then stored at 4°C. The first and second nucleic acid probes were then ligated together by adding to 1 μL of the gap extension product, 10 μL of 2X rapid ligation buffer (Enzymatics), 1 μL of T4 DNA Rapid Ligase (Enzymatics), and 8 μL of PCR-grade H_2O. These ligation reactions were incubated 15 min 20°C, quenched with 1 μL 0.5M EDTA, and pooled into a single 15 mL conical tube.

[00389] In the case in which the first and the second nucleic acid probes hybridized to the identifier oligonucleotide such that the first and the second nucleic acid probes were adjacent and were not overlapping, nick repair ligation reaction was performed. To 2.5 μL of each annealed identifier oligonucleotide-nucleic acid probe mixture, 10 μL of 2X rapid ligation buffer (Enzymatics), 1 μL of T4 DNA Rapid Ligase (Enzymatics), and 1 μL of PCR-grade H_2O was added. These ligation reactions were incubated for 15 minutes at 20°C, subsequently quenched with 1 μL of 0.5M EDTA, and pooled into a single 15 mL conical tube.

[00390] The pools of quenched ligation reactions were then purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 350 µL of AMPure XP beads and 3.15 mL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). The purification was performed with 3.5 mL of diluted AMPure XP beads and eluted in 200 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. The AMPure XP bead cleanup was then repeated with 400 µL of AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00391] To amplify the purified ligation products, PCR reactions with purified ligation products and primers were prepared. To 6 µL of purified ligation product, 10 µL of NEBNext Ultra II Q5 Master Mix, 0.2 µL of 100 µM forward primer and reverse primer and 3.6 µL of PCR-grade H_2O was added. . The forward primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the first amplification primer binding site located on the first strand of the nucleic acid adapter. Table 8 provides the sequences of the forward primers used.

Table 8. Forward primers for short probe hybridization method.

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	13
CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	14
CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	15
CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	16
CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	17
CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	18
CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	19
CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	20
CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	21
CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	22
CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	23

CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	24

[00392] The reverse primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the second amplification primer binding site located on the second strand of the nucleic acid adapter. Table 9 provides the sequences of the reverse primers used.

Table 9. Reverse primers for short probe hybridization method.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACA CGACGCTCTTCCGATCT	155
AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCCTAC ACGACGCTCTTCCGATCT	156
AATGATACGGCGACCACCGAGATCTACACCCTATCCTACACTCTTTCCCTACA CGACGCTCTTCCGATCT	157
AATGATACGGCGACCACCGAGATCTACACGGCTCTGAACACTCTTTCCCTACA CGACGCTCTTCCGATCT	158
AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTTCCCTAC ACGACGCTCTTCCGATCT	159
AATGATACGGCGACCACCGAGATCTACACTAATCTTAACACTCTTTCCCTACA CGACGCTCTTCCGATCT	160
AATGATACGGCGACCACCGAGATCTACACCAGGACGTACACTCTTTCCCTACA CGACGCTCTTCCGATCT	161
AATGATACGGCGACCACCGAGATCTACACGTACTGACACACTCTTTCCCTACA CGACGCTCTTCCGATCT	162

[00393] The PCR program used to amplify the purified extension products comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) eighteen times

(6) 2 minutes at 72°C

[00394] The amplified products were stored at 4°C. 4 µL of each PCR reaction were combined into six pools, for 16 samples per pool. The amplified products were further purified using an AMPure XP bead cleanup with 64 µL of AMPure XP beads and eluting with 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00395] The purified amplified products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The purified amplified products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with either standard sequencing primers or a custom spike-in Read1 primer (SEQ ID NO: 25).

[00396] Example 6—direct PCR method for 96 multiplexed samples

[00397] In this example, a direct PCR method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. In this example, 8 species of forward amplification primers and 12 species of reverse amplification primers were used. The forward primers comprised a P5 flow cell adapter (SEQ ID NO: 3), a nucleic acid sequence comprising a unique molecular identifier and a region complementary to a first amplification primer binding site present on the identifier oligonucleotide. Table 10 provides the sequences of the forward amplification primers used.

Table 10. Forward amplification primers for direct PCR method.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACGCTCAGATATAGCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	5
AATGATACGGCGACCACCGAGATCTACACGCTCAGAATAGAGGCACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	6
AATGATACGGCGACCACCGAGATCTACACGCTCAGACCTATCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	7
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGGCTCTGAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	8
AATGATACGGCGACCACCGAGATCTACACGCTCAGAAGGCGAAGACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	9
AATGATACGGCGACCACCGAGATCTACACGCTCAGATAATCTTAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	10
AATGATACGGCGACCACCGAGATCTACACGCTCAGACAGGACGTACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	11
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGTACTGACACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	12

[00398] The reverse primers comprised a P7 flow cell adapter (SEQ ID NO: 4), a nucleic acid sequence comprising a unique molecular identifier and a region complementary to a second amplification primer binding site present on the identifier oligonucleotide. Table 11 provides the sequences of the reverse amplification primers used.

Table 11. Reverse amplification primers for direct PCR method.

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	163
CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	164
CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	165
CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	166
CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	167
CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	168
CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	169
CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	170
CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	171
CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	172
CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	173
CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	174

[00399] In the case of 8 forward amplification primers and 12 reverse amplification primers, when the unique molecular identifiers from a pair of forward and reverse primers are combined, a total of 96 unique combinations can be obtained, allowing for the multiplexing of 96 samples.

[00400] To amplify the collected identifier oligonucleotides for sequencing, PCR reactions with collected identifier oligonucleotides and forward and reverse amplification primers were prepared on a 96-well plate with 2 μL of each identifier oligonucleotide sample, 10 μL of NEBNext Ultra II Q5 Master Mix, 2 μL of 10 μM forward amplification primer, 2 μL of reverse amplification primer, and 4 μL of PCR-grade H_2O. Each well in the 96-well plate contained an identifier oligonucleotide sample and a unique combination of forward and reverse amplification primers. The PCR program used comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) six to ten times

(6) 2 minutes at 72°C

[00401] The amplified products were stored at 4°C. 10 µL of each PCR reaction was combined into a single 15 mL conical tube.

[00402] The pooled PCR reactions were purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 115.2 µL of AMPure XP beads and 1036.8 µL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). Purification was performed with 1152 µL of diluted AMPure XP beads and eluted in 60 µL of a buffer comprising 10 mM Tris pH 8. The purification process was repeated with 60 µL of AMPure XP beads and eluted in 70 µL of a buffer comprising 10 mM Tris pH 8.

[00403] Following AMPure XP cleanup, PCR reactions with universal primers were prepared with 9 µL of pooled direct PCR product, 15 µL of NEBNext Ultra II Q5 Master Mix, 3 µL of 10 µM universal P7 primer (SEQ ID NO: 153) and 2 µL of 10 µM universal P5 primer (SEQ ID NO: 154). The PCR program used was:

(1) 30 seconds at 98°C,

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) 15 to 24 times

(6) 2 minutes at 72°C

[00404] Two rounds of AMPure XP bead cleanup was performed. The first round was performed with 30 µL of beads and eluted with 20 µL of a buffer comprising 10 mM Tris pH 8 and the second round was performed with 20 µL beads and eluted with 11 µL of a buffer comprising 10 mM Tris pH 8.

[00405] These purified PCR products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics). The purified PCR products were also diluted for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with a custom spike-in primer (SEQ ID NO: 25).

[00406] Example 7—spatially detecting target analytes in a FFPE sample

[00407] The methods of the present invention were used to spatially detect a plurality of different target analytes, including target proteins and target RNAs, in a sample of inflamed human tonsil tissue FFPE section.

[00408] In one experiment, 30 different target proteins were spatially detected using the methods of the present disclosure in two serial sections cut from the inflamed human tonsil tissue FFPE section. The 30 target proteins are put forth in Table 12. The 30 target proteins included IgG Rabbit isotype and IgG Mouse isotype as negative controls, as these target proteins should not have been present in the inflamed human tonsil sample and therefore should not have been detected.

Table 12. Target Proteins

Target Protein	Target Protein
AKT	FOXP3
B7-H3	GZMB
Bcl-2	Histone H3
Beta-2-microglobulin	Ki67
Beta-catenin	CD20
CD14	P-AKT
CD19	PanCK
CD3	PD1
CD4	PD-L1
CD44	S6
CD45	STAT3
CD45RO	P-STAT3
CD56	VISTA
CD68	IgG Rabbit isotype (negative control)
CD8A	IgG Mouse isotype (control)

[00409] To spatially detect the 30 target proteins, 30 different probes of the present disclosure were used. Each probe comprised a target binding domain comprising an antibody that specifically binds to one of the 30 target proteins in Table 12. The two serial sections were contacted with a plurality of the 30 different probes. Ninety-six regions of interest (ROI) were then identified. For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a short probe hybridization method of the present disclosure, thereby spatially detecting the 30 target proteins in the two serial sections. As shown

in Figures 21A-21D, the number of reads per target protein in each ROI for the two serial sections were well correlated, demonstrating that the method yields reproducible results.

[00410] In a second experiment, 20 different target RNAs were spatially detected using the methods of the present disclosure in two different serial sections cut from the inflamed human tonsil tissue FFPE section. The 20 different target RNAs are put forth in Table 13. The 20 target RNAs included 6 negative controls(Negative Probe) that should not have been detected in the sample.

Table 13. Target RNAs

Target RNA	Target RNA	Target RNA
CD3E	CD40	CTLA4
CD3G	CD45	GAPDH
CD4	CD74	KRT13
CD20	CD79A	PD1
PSA	RP56	Negative Probe #1
Negative Probe #2	Negative Probe #3	Negative Probe #4
Negative Probe #5	Negative Probe #6	

[00411] To spatially detect the 20 Target RNAs, 20 different probes of the present disclosure were used. Each probe comprised a target binding domain comprising a nucleic acid sequence complementary to at least one portion of one of the 20 target RNAs. The two serial sections were contacted with a plurality of the 20 different probes. Ninety-six regions of interest (ROI) were then identified. For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a direct PCR method of the present disclosure, thereby spatially detecting the 20 target RNAs in the two serial sections. As shown in Figures 22A-22D, the number of reads per target RNA in each ROI for the two serial sections were well correlated, demonstrating that the method yields reproducible results.

[00412] Example 8—spatially detecting target proteins in a fluorescently stained FFPE sample

[00413] In another experiment, a 5 μm FFPE section of inflamed human tonsil tissue was stained with 4 fluorescent visualization markers: (1) CD3E, a T-cell marker; (2) PanCK, an epithelial cell marker; (3) Ki-67, a proliferation marker; and (4) SYTO83, a DNA stain, as shown in the left panel of Figure 23. The stained FFPE section was then contacted with the probes directed

against 30 target proteins, as described in Example 7. As shown in the left panel Figure 23, 96 regions of interest (ROIs) were selected. Each ROI was a circle with a 500 μm diameter. For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a short probe hybridization method of the present disclosure, thereby spatially detecting the 30 target proteins in the FFPE section. As shown in the right panel of Figure 23, PanCK, CD3E and Ki67 were spatially detected in ROIs that correlated with their fluorescent visualization markers. Thus, the results generated by the methods of the present disclosure correlate with the results generated using established immunohistochemical method.

[00414] Example 9—spatially detecting target RNAs in a FFPE sample

[00415] In another experiment, a 5 μm section from an inflamed human tonsil tissue FFPE block was contacted with probes directed against 20 target RNAs, as described in Example 7. 96 regions of interest (ROIs) were then selected. Each ROI was a circle with a 500 μm diameter. For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a direct PCR method of the present disclosure, thereby spatially detecting the 20 target RNAs in the two serial sections. The total RNA from a 20 μm section from the same inflamed human tonsil tissue FFPE block was then isolated. The total RNA was analyzed using the NanoString nCounter® system. Figure 24 shows that the average number of counts for 11 different RNA targets recorded using the methods of the present disclosure were well correlated with the average number of counts for the same 11 different RNA targets recorded using the nCounter® system. Thus, the results generated using the methods of the present disclosure correlate with the results generated using established direct detection methods.

[00416] Example 10—spatially detecting target RNAs in specific sub-regions of an ROI

[00417] In another experiment, a 5 μm section from an inflamed human tonsil tissue FFPE block was contacted with probes directed against 30 target proteins, as described in Example 7. The same 5 μm section was also stained with 4 fluorescent visualization markers: (1) CD3E, a T-cell marker; (2) PanCK, an epithelial cell marker; (3) Ki-67, a proliferation marker; and (4) SYTO83, a DNA stain. As shown in Figure 25, 48 regions of interest (ROIs) were identified. For each ROI, two sub-regions were then identified based on the fluorescent staining. Areas of an ROI that were fluorescently stained positive for PanCK (PanCK+) were designated a "tumor" sub-

region and the areas of an ROI that lacked PanCK fluorescent staining were designated a "micro-environment" sub-region, as shown in Figure 25. For each ROI, the tumor sub-region and the micro-environment sub-region were separately illuminated with UV light to release the identifier oligonucleotides from the probes bound within each sub-region by creating a custom mask based on the intensity of PanCK fluorescent staining. The released identifier oligonucleotides were also separately collected. The collected identifier oligonucleotides were then analyzed using the short-probe hybridization method of the present disclosure and the NanoString nCounter® system. As shown in the bottom panel of Figure 25, the results using the NanoString nCounter® system and the short-probe hybridization method of the present disclosure were well correlated. Furthermore, in the tumor sub-regions, PanCK was detected at a significantly higher level as compared to the micro-environment sub-regions. Thus, the spatial detection results provided by the methods of the present disclosure are consistent with established fluorescent immunohistochemical methods and allows for the spatial detection within highly specific regions of a sample.

[00418] Example 11—96-plex human immuno-oncology panel

[00419] A 96-plex human immuno-oncology panel was designed for use in the direct-PCR methods of the present invention. The panel comprised a plurality of probes that could be used to spatially detect 96 different human target RNAs using the direct-PCR methods of the present disclosure. The 96 target RNAs are shown in Table 14.

Table 14. Target RNAs

Target	Target	Target	Target	Target	Target	Target	Target
AKT1	CD3E	CEACAM8	FOXP3	IFNG	KRT1	NCAM1	SOD2
ARG1	CD4	CMKLR1	GZMB	IFNGR1	KRT10	NKG7	SOX10
B2M	CD40	CSF1R	H3F3A	IL10	KRT14	NT5E	STAT1
BATF3	CD40LG	CTLA4	HAVCR2	IL12B	KRT17	PDCD1	STAT2
BCL2	CD44	CTNNB1	HIF1A	IL15	KRT18	PDCD1LG2	STAT3
BCL2L1	CD47	CXCL10	HLA_DQA1	IL1B	KRT19	PECAM1	TBX21
CCL5	CD68	CXCL9	HLA_DRB	IL6	KRT6A	PMEL	TIGIT
CCND1	CD74	CSCR6	HLA_E	ITGAM	KRT7	PSMB10	TNF
CD14	CD86	DKK2	ICAM1	ITGAV	LAG3	PTEN	TNFRSF9
CD27	CD8A	EPCAM	ICOSLG	ITGAX	LY6E	PRTPRC	TNFSF4

CD274	CEACAM1	FAS	IDO1	ITGB2	MKI67	RPS6	VEGFA
CD276	CEACAM6	FASLG	IFNAR1	ITGB8	MS4A1	S100B	VSIR

[00420] In total, the panel comprised 928 different probes. Each of the probes comprised an identifier oligonucleotide comprising a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier, a unique nucleic acid sequence which identified the target RNA bound to the target binding domain and a second amplification primer binding site. Figure 26 shows a schematic of the probes used in the panel. For each of the 96 target RNAs, there was at least one probe within the 928 probe set comprising a target binding domain that directly or indirectly hybridized to that target RNA. For most of the 96 target RNAs, there were 10 different probes that directly or indirectly hybridized to the specific target RNA. These 10 different probes directly or indirectly hybridized to different locations on the target RNA to create a "tiling" effect, as shown in the top panel of Figure 27. Tiling multiple probes onto a target RNA means that each target RNA will be individually detected multiple times, increasing the overall accuracy of the measurement. For example, as shown in the bottom panel of Figure 27, in the case where 10 probes are tiled onto a single target RNA, one of the probes may be incorrectly detected too many times (outlier high count probe), while another probe may be incorrectly detected too few times (outlier low count probe). However, the other 8 probes may be detected at a similar level, indicating that the two outliers should be discarded during analysis and the signals from the 8 probes used to generate a more accurate measurement of the abundance of the target RNA.

[00421] The set of 928 probes also comprised 80 negative control probes. Each of the 80 negative control probes comprised a target binding domain comprising a scrambled, non-specific nucleic acid sequence that was designed using guidelines from the External RNA Controls Consortium such that the target binding domain should not be complementary to RNA molecules present within a human sample. Thus, these 80 negative control probes should not be detected during analysis.

[00422] The 96-plex human immune-oncology panel was used to analyze a tissue microarray. The tissue microarray comprised FFPE samples of 22 common human cell lines, including normal and cancerous cell types. Some of the cell lines are shown in Table 15.

[00423] Table 15. Cell lines

Cell Line	Cell Line	Cell Line	Cell Line
CCRF-CEM	DAUDI	H596	H2228
HT29	HUT78	HUH7	JURKAT
M14	MDA-MB-468	MOLT4	RAJI
SKBR3	SUDHL1	SUDHL4	

[00424] The tissue microarray also comprised one mouse cell line (3T3) as a negative control. Each of the FFPE samples on the microarray was contacted with a plurality of the 928 different probes in the imunno-onocology panel. As shown in Figure 28, for each of the FFPE samples, at least three circular regions of interest (ROIs) with a diameter of 300 μm were selected. As a negative control, ROIs were also selected on regions of the microarray that did not comprise a FFPE sample (glass negative control). For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a direct PCR method of the present disclosure thereby spatially detecting the 96 target RNAs in each of the FFPE samples on the tissue microarray.

[00425] Figure 29 shows that a sufficient read depth was achieved using a MiSeq v3 flowcell. The top panel of Figure 30 shows that none of the target RNAs were spatially detected for the glass negative control ROIs. Likewise, the bottom panel of Figure 30 shows that nearly none of the target RNAs were detected in the negative control mouse 3T3 FFPE sample. Conversely, as shown in Figures 31, 33 and 34, specific target RNAs were successfully detected in the HEK293 (human embryonic kidney) FFPE sample and the Jurkat (human T-cell lymphocyte) FFPE sample. Figures 31, 33 and 34 show that clusters of "tiled" probes were detected for particular target RNAs, including AKT1, B2M, CD3E, HIF1A, PTEN, RPS6, STAT1, STAT2, STAT3, VEGF, PTPRC (CD45), and KRT1/10/18/19. These results indicated that there are certain target RNAs that are differentially transcribed in the two different cells lines. The results of this experiment were also verified using the NanoString nCounter system to identify the collected identifier oligonucleotides. As shown in Figure 27, the results using the direct-PCR method of the present disclosure were consistent with the results obtained using the NanoString nCounter system.

BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

BACKGROUND OF THE INVENTION

[0001] This application claims priority to, and the benefit of, U.S. Provisional Application No. 62/629,180, filed February 12, 2018 and U.S. Provisional Application No. 62/771,212, filed November 26, 2018. The contents of each of the aforementioned patent applications are incorporated herein by reference in their entireties.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on ~~February 1, 2019~~May 19, 2021, is named "NATE-037_001US_Sequence_Listing.txt" and is ~~48.5 KB~~about 50,049 bytes in size.

BACKGROUND OF THE INVENTION

[0003] Standard immunohistochemical and *in situ* hybridization methods allow for simultaneous detection of, at most, six to ten protein or nucleic acid targets, with three to four targets being typical. There exists a need for probes, compositions, methods, and kits for simultaneous, multiplexed detection and quantification of protein and/or nucleic acid expression in a user-defined region of a tissue, user-defined cell, and/or user-defined subcellular structure within a cell. Furthermore, there is a need for such systems to be adaptable for use with existing sequencing technologies that are already available to a large number of end users.

SUMMARY OF THE INVENTION

[0004] The present disclosure relates to probes, compositions, methods, and kits for simultaneous, multiplexed, spatial detection and quantification of protein and/or nucleic acid expression in a user-defined region of a tissue, user-defined cell, and/or user-defined subcellular structure within a cell.

[0005] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, a second amplification primer binding site and optionally, a constant nucleic acid sequence to minimize ligation bias; (5) amplifying the ligation product produced in step (4); and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0006] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises: a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, a second amplification primer binding site and optionally, a constant nucleic acid sequence to minimize ligation bias, and wherein at least one of the first or second amplification primer binding sites identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; (5) amplifying the ligation product produced in step (4); and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0007] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises: a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, a second amplification primer binding site and optionally, a constant nucleic acid sequence to minimize ligation bias; (5) amplifying the extension product produced in step (4) using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprise a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0008] The nucleic acid adapter of step (4) can be a partially double-stranded nucleic acid molecule. A partially double-stranded nucleic acid adapter can comprise a double-stranded annealed region, a first single-stranded mismatched region and a second single-stranded mismatched region. The first single-stranded mismatched region and the second single stranded mismatched region can be present on opposing sides of the double-stranded annealed region.

[0009] The nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be present in the double-stranded annealed region of a partially double-stranded nucleic acid adapter.

[0010] The constant nucleic acid sequence to minimize ligation bias can be present in the double-stranded annealed region of a partially double-stranded nucleic acid adapter.

[0011] A unique molecular identifier can be present in at least one of the first or second single-stranded mismatched regions of a partially double-stranded nucleic acid adapter.

[0012] The first amplification primer binding site can be present in the first single-stranded mismatched region of a partially double-stranded nucleic acid adapter and the second amplification primer binding site can be present in the second single-stranded mismatched region of the same partially double-stranded nucleic acid adapter.

[0013] The methods of the present disclosure described in the preceding can further comprise prior to step (4), performing an end repair reaction. The methods can also further comprise prior to step (4), performing a tailing reaction to attach a single nucleotide overhang to the 3' ends of the identifier oligonucleotide. The methods can further comprise, prior to step (4), performing an end repair reaction and a tailing reaction to attach a single nucleotide overhang to the 3' ends of the identifier oligonucleotide. The tailing reaction and the end repair reaction can be performed sequentially or concurrently.

[0014] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a first amplification primer binding site, and a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier, a second amplification primer binding site and optionally, a constant nucleic acid sequence to minimize ligation bias; (5) amplifying the ligation product produced in step (4); and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0015] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises: a first amplification primer binding site and a unique nucleic acid sequence which identifies the target

analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) ligating to the released identifier oligonucleotide at least one nucleic acid adapter, wherein the nucleic acid adapter comprises a nucleic acid sequence comprising a unique molecular identifier, a second amplification primer binding site, and optionally, a constant nucleic acid sequence to minimize ligation bias; (5) amplifying the extension product produced in step (4) using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprise a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; and (6) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0016] The nucleic acid adapter of step (4) can be a partially double-stranded nucleic acid molecule. A partially double-stranded nucleic acid adapter can comprise a double-stranded annealed region and a single-stranded mismatched region.

[0017] The nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be present in the double-stranded annealed region of a partially double-stranded nucleic acid adapter.

[0018] The constant nucleic acid sequence to minimize ligation bias can be present in the double-stranded annealed region of a partially double-stranded nucleic acid adapter. The constant nucleic acid sequence can also comprise a cleavable moiety. The cleavable moiety can be an enzymatically cleavable moiety. The enzymatically cleavable moiety can be a USER sequence.

[0019] A unique molecular identifier can be present in the single-stranded mismatched region of a partially double-stranded nucleic acid adapter.

[0020] The second amplification primer binding site can be present in the single-stranded mismatched region of a partially double-stranded nucleic acid adapter.

[0021] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a first

amplification primer binding site, and a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a single stranded nucleic acid template, wherein the nucleic acid template comprises a region complementary to the unique nucleic acid sequence of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier, a nucleic acid sequence complementary to a second amplification primer binding site and optionally, an affinity molecule; (5) extending the identifier oligonucleotide of step (4) to form an extension product complementary to the single stranded nucleic acid template, wherein the extension product comprises the identifier oligonucleotide, the nucleic acid sequence complementary to the unique molecular identifier, and the second amplification primer binding site; (6) amplifying the extension product produced in step (5) using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprise a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0022] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a first amplification primer binding site and a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a single stranded nucleic acid template, wherein the nucleic acid template comprises a region complementary to the unique nucleic acid sequence of the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier, a nucleic acid

sequence complementary to a second amplification primer binding site and optionally, an affinity molecule; (5) extending the identifier oligonucleotide of step (4) to form an extension product complementary to the single stranded nucleic acid template, wherein the extension product comprises the identifier oligonucleotide, the nucleic acid sequence complementary to the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, the nucleic acid sequence complementary to the unique molecular identifier and the second amplification primer binding site; (6) amplifying the extension product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0023] The single stranded nucleic acid template can further comprise an affinity molecule. In aspects in which the single stranded nucleic acid template comprises an affinity molecule, the methods of the present disclosure described in the preceding can further comprise an affinity purification step between steps (4) and (5).

[0024] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a first amplification primer binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier and a second amplification primer binding site, and wherein the first the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) performing nick repair such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying

the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0025] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide and a first amplification primer binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide and a second amplification primer binding site, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence comprising a unique molecular identifier, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) performing nick repair such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0026] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier

oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a first amplification primer binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier and a second amplification primer binding site, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping; (5) performing a gap extension and nick repair reaction such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0027] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide and a first amplification primer binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide and a second amplification primer binding site, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence comprising a unique molecular identifier, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein the first and the second nucleic acid probes

hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping; (5) performing a gap extension and nick repair reaction such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0028] In aspects in which the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released is located on the first nucleic acid probe, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be located 5' to the first amplification primer binding site.

[0029] In aspects in which a unique molecular identifier is located on the second nucleic acid probe, the unique molecular identifier can be located 3' to the second amplification primer binding site.

[0030] In aspects in which the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a unique molecular identifier are present in the first nucleic acid probe, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a unique molecular identifier can be located 5' to the first amplification primer binding site.

[0031] In aspects in which the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a unique molecular identifier are present in the second nucleic acid probe, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the unique molecular identifier can be located 3' to the second amplification primer binding site.

[0032] In aspects in which a unique molecular identifier is present in the first nucleic acid probe and the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released is present in the second nucleic acid probe, the unique molecular identifier can be located 5' to the first amplification primer binding site and the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be located 3' to the second amplification binding site.

[0033] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a first amplification primer binding site, a nucleic acid sequence comprising a first unique molecular identifier and a first flow cell binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a second unique molecular identifier, a nucleic acid sequence comprising a third unique molecular identifier and a second flow cell binding site, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) performing nick repair such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0034] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to a portion of the

identifier oligonucleotide, a first amplification primer binding site, a nucleic acid sequence comprising a first unique molecular identifier and a first flow cell binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a second unique molecular identifier and a second flow cell binding site, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence comprising a third unique molecular identifier, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) performing nick repair such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0035] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to the identifier oligonucleotide, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, a first amplification primer binding site, a nucleic acid sequence comprising a first unique molecular identifier and a first flow cell binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to the identifier oligonucleotide, a nucleic acid sequence comprising a second unique molecular identifier, a nucleic acid sequence comprising a third unique molecular identifier and a second flow cell binding site, and wherein the first and the second nucleic acid

probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping; (5) performing a gap extension and nick repair reaction such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0036] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises a nucleic acid complementary to the identifier oligonucleotide, a first amplification primer binding site, a nucleic acid sequence comprising a first unique molecular identifier and a first flow cell binding site, and wherein the second nucleic acid probe comprises a nucleic acid complementary to the identifier oligonucleotide, a nucleic acid sequence comprising a second unique molecular identifier and a second flow cell binding site, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence comprising a third unique molecular identifier, and wherein at least one of the first or second nucleic acid probes comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping; (5) performing a gap extension and nick repair reaction such that the hybridized first and second nucleic acid probes are ligated together; (6) amplifying the ligation product produced in step (5); and (7) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0037] In aspects in which the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the first unique molecular identifier are present in the first nucleic acid probe, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the first unique molecular identifier can be located 5' to the first flow cell binding site.

[0038] In aspects in which the second and the third unique molecular identifiers are present in the second nucleic acid probe, the second and the third unique molecular identifiers can be located 3' to the second flow cell binding site.

[0039] In some aspects, the first unique molecular identifier can be present in the first nucleic acid probe and can be located 5' to the first flow cell binding site. In other aspects, the second unique molecular identifier can be present in the second nucleic acid probe and can be located 3' to the second flow cell binding site.

[0040] In some aspects, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the third unique molecular identifier can be present in the first nucleic acid probe and can be located 5' to the first flow cell binding site.

[0041] In some aspects, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and the third unique molecular identifier can be present in the second nucleic acid probe and can be located 3' to the second flow cell binding site.

[0042] In some aspect, the third unique molecular identifier can be present in the first nucleic acid probe and can be located 5' to the first flow cell binding site. In this same aspect, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can be present in the second nucleic acid probe and can be located 3' to the second flow cell binding site.

[0043] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding

domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) amplifying the released identifier oligonucleotide using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released; (5) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0044] The present disclosure also provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a first amplification primer binding site and a second amplification primer binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) amplifying the released identifier oligonucleotide using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site, wherein at least one of the amplification primers comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released, and wherein at least one of the amplification primers comprises a nucleic acid sequence comprising a unique molecular identifier; and (5) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0045] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding

domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain and a capture probe binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a capture probe, wherein the capture probe comprises an affinity molecule and a region complementary to the capture probe binding site; and (5) Identifying the released identifier oligonucleotide by sequencing the amplified hybridized product produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0046] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a capture probe binding site and a multiplexing probe binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the released identifier oligonucleotide a capture probe and a multiplexing probe, wherein the capture probe comprises an affinity molecule and a region complementary to the capture probe binding site, and wherein the multiplexing probe comprises a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released and a region complementary to the multiplexing probe binding site; and (5) Identifying the released identifier oligonucleotide by sequencing the hybridized product produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0047] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) hybridizing to the

released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe, wherein the first nucleic acid probe comprises: a nucleic acid complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and wherein the second nucleic acid probe comprises: a nucleic acid complementary to a portion of the identifier oligonucleotide, and a second amplification primer binding site, and wherein the first and the second nucleic acid probes hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent but not overlapping; (5) ligating the hybridized first and second nucleic acid probes together; (6) amplifying the ligation product produced in step (5); and (7) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0048] The present disclosure provides a method for spatially detecting at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least one probe comprising a target binding domain and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises: a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotide; (3) collecting the released identifier oligonucleotide; (4) amplifying the collected identifier oligonucleotide; (5) Identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[0049] A In all methods of the present disclosure, the ligation process can be a nick ligation process. The nick ligation process can be a nick repair process.

[0050] In all methods of the present disclosure, the sequencing can be an enzyme free sequencing method.

[0051] In all methods of the present disclosure, the identifier oligonucleotide can be double-stranded. In aspects in which the identifier oligonucleotide is double-stranded, at least one of the two strands of the identifier oligonucleotide can comprise at least two separate nucleic acid

molecules. Alternatively, at least one 3' end of an identifier oligonucleotide can comprise a single nucleotide overhang.

[0052] In all methods of the present disclosure, the identifier oligonucleotide can be single-stranded.

[0053] In all methods of the present disclosure, the unique nucleic acid sequence which identifies the target analyte bound to a target binding domain can comprise between about 5 nucleotides and about 40 nucleotides preferably about 35 nucleotides, preferably still about 10 nucleotides.

[0054] In all methods of the present disclosure, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can comprise between about 6 nucleotides and about 15 nucleotides, preferably about 12 nucleotides, preferably still about 10 nucleotides.

[0055] In all methods of the present disclosure, at least one of a first nucleic acid probe or a second nucleic acid probe can comprise an affinity molecule. For example, at least one of a first nucleic acid probe or a second nucleic acid probe can comprise a biotin.

[0056] In all methods of the present disclosure, an amplification primer binding site can comprise between about 18 nucleotides and about 40 nucleotides, preferably about 32 nucleotides, preferably still about 25 nucleotides. An amplification primer binding site can comprise an i7 sequence, wherein the i7 sequence comprises the sequence set forth in SEQ ID NO: 1. An amplification primer binding site can comprise an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2.

[0057] In all methods of the present disclosure, an amplification primer can comprise a flow cell adapter sequence, wherein the flow cell adapter sequence is suitable for sequencing. An amplification primer can comprise a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3. An amplification primer can comprise a P7 flow cell adapter sequence, wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4.

[0058] In all methods of the present disclosure, a flow cell binding site can comprise a flow cell adapter sequence, wherein the flow cell adapter sequence is suitable for sequencing. A flow cell binding site can comprise a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3. A flow cell binding site can

comprise a P7 flow cell adapter sequence, wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4.

[0059] In all methods of the present invention, at least one of the amplification primers can comprise an affinity molecule. For example, at least one of the amplification primers cam comprise a biotin.

[0060] In all methods of the present disclosure, amplification can comprise performing PCR. Performing PCR can comprise an amplification primer.

[0061] An amplification primer can comprise a flow cell binding site. An amplification primer can comprise a nucleic sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released. An amplification primer can comprise a nucleic acid sequence complementary to an amplification primer binding site.

[0062] Any of the above aspects can be combined with any other aspect.

[0063] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. In the Specification, the singular forms also include the plural unless the context clearly dictates otherwise; as examples, the terms “a,” “an,” and “the” are understood to be singular or plural and the term “or” is understood to be inclusive. By way of example, “an element” means one or more element. Throughout the specification the word “comprising,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from the context, all numerical values provided herein are modified by the term “about.”

[0064] Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. The references cited herein are not admitted to be prior art to the claimed invention. In the case of conflict, the present Specification, including definitions, will control. In addition, the materials, methods, and examples are

illustrative only and are not intended to be limiting. Other features and advantages of the disclosure will be apparent from the following detailed description and claim.

BRIEF DESCRIPTION OF THE DRAWINGS

[0065] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

[0066] The above and further features will be more clearly appreciated from the following detailed description when taken in conjunction with the accompanying drawings.

[0067] Figure 1 is a schematic of a two-ended adapter ligation method of the present disclosure.

[0068] Figure 2 is a schematic of a one-ended adapter ligation method of the present disclosure.

[0069] Figure 3 is a schematic of a templated-primer extension method of the present disclosure.

[0070] Figure 4 is a schematic of a template-extended identifier oligonucleotide of the present disclosure.

[0071] Figure 5 is a schematic of a short probe hybridization method of the present disclosure.

[0072] Figure 6 is a schematic of a short probe hybridization method of the present disclosure.

[0073] Figure 7 is a schematic of a short probe hybridization method of the present disclosure.

[0074] Figure 8 is a schematic of a long probe hybridization method of the present disclosure.

[0075] Figure 9 is a schematic of a long probe hybridization method of the present disclosure.

[0076] Figure 10 is a schematic of a long probe hybridization method of the present disclosure.

[0077] Figure 11 is a schematic of a direct-PCR method of the present disclosure.

[0078] Figure 12 is a schematic of an enzyme free method of the present disclosure.

[0079] Figure 13 is a schematic of a multiplexed enzyme free method of the present disclosure.

[0080] Figure 14 is a schematic of a probe of the present disclosure indirectly binding to a target nucleic acid.

[0081] Figure 15 is a schematic of an identifier oligonucleotide-short nucleic acid probe complex of the present disclosure.

[0082] Figure 16 is a schematic of a short probe hybridization method of the present disclosure.

[0083] Figure 17 is a schematic of an identifier oligonucleotide-short nucleic acid probe complex of the present disclosure.

[0084] Figure 18 is a schematic of a short probe hybridization method of the present disclosure.

[0085] Figure 19 is a schematic of a direct-PCR method of the present disclosure.

[0086] Figure 20 is a schematic overview of the methods of the present disclosure.

[0087] Figure 21A, Figure 21B, Figure 21C and Figure 21D ~~shows~~ show the spatial detection of protein target analytes using the methods of the present disclosure.

[0088] Figure 22A, Figure 22B, Figure 22C and Figure 22D ~~shows~~ show the spatial detection of RNA target analytes using the methods of the present disclosure.

[0089] Figure 23 shows the spatial detection of protein target analytes using the methods of the present disclosure.

[0090] Figure 24 shows the spatial detection of RNA target analytes using the methods of the present disclosure.

[0091] Figure 25 shows the spatial detection of protein target analytes using the methods of the present disclosure.

[0092] Figure 26 is a schematic of a probe of the present disclosure. The nucleic acid sequence shown in Figure 26 corresponds to SEQ ID NO: 175.

[0093] Figure 27 shows the use of probe tiling in the methods of the present disclosure.

[0094] Figure 28 shows the regions of interest selected on a tissue microarray.

[0095] Figure 29 is a series of graphs showing the read depth achieved using the methods of the present disclosure.

[0096] Figure 30 is a series of graphs showing the spatial detection of RNA target analytes in negative control samples using the methods of the present disclosure.

[0097] Figure 31 is a series of graphs showing the spatial detection of RNA target analytes in a HEK293 sample (top panel) and a Jurkat cell sample (bottom panel) using the methods of the present disclosure.

[0098] Figure 32 is a series of graphs showing the spatial detection of RNA target analytes in sixteen FFPE samples using the methods of the present disclosure.

[0099] Figure 33 is a graph showing the spatial detection of RNA target analytes in a HEK293 sample using the methods of the present disclosure.

[00100] Figure 34 is a graph showing the spatial detection of RNA target analytes in a Jurkat cell sample using the methods of the present disclosure.

DETAILED DESCRIPTION OF THE INVENTION

[00101] The present disclosure is based in part on probes, compositions, methods, and kits for simultaneous, multiplexed spatial detection and quantification of protein and/or nucleic acid expression in a user-defined region of a tissue, user-defined cell, and/or user-defined subcellular structure within a cell using existing sequencing methods.

[00102] The present disclosure provides a comparison of the identity and abundance of target proteins and/or target nucleic acids present in a first region of interest (*e.g.*, tissue type, a cell (including normal and abnormal cells), and a subcellular structure within a cell) and the identity and abundance of target proteins and/or target nucleic acids present in a second region of interest. There is no pre-defined upper limit to the number of regions of interest and comparisons that can be made; the upper limit relates to the size of the region of interest relative the size of the sample. As examples, when a single cell represents a region of interest, then a section may have hundreds to thousands of regions of interest; however, if a tissue section includes only two cell types, then the section may have only two regions of interest (each including only one cell type).

[00103] The present disclosure provides a higher degree of multiplexing than is possible with standard immunohistochemical or *in situ* hybridization methods. Standard immunohistochemical methods allow for maximal simultaneous detection of six to ten protein targets, with three to four protein targets being more typical. Similarly, *in situ* hybridization methods are limited to simultaneous detection of fewer than ten nucleic acid targets. The present disclosure provides detection of large combinations of nucleic acid targets and/or protein targets from a defined region of a sample. The present disclosure provides an increase in objective measurements by digital quantification and increased reliability and consistency, thereby enabling comparison of results among multiple centers.

[00104] Various compositions and methods of the present disclosure are described in full detail herein.

[00105] In one aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "two-ended adapter ligation method".

[00106] A two ended-adapter ligation method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The probes and samples of the present disclosure are described in further detail herein. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain.

[00107] Following contacting the at least on target analyte with the at least one probe, a two-ended adapter ligation method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00108] Following release of the identifier oligonucleotide, a two-ended adapter ligation method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location, thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00109] Following collection of the released identifier oligonucleotide, a two-ended adapter ligation method can further comprise: (4) ligating to the released identifier oligonucleotide collected in step (3) at least one nucleic acid adapter.

[00110] The nucleic acid adapter can comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated

"ROI #1", the nucleic acid adapter would comprise a nucleic acid sequence that corresponds to "ROI #1".

[00111] The nucleic acid adapter can also comprise a unique molecular identifier.

[00112] The nucleic acid adapter can also comprise a first amplification primer binding site. In other aspects, the nucleic acid adapter can also comprise a second amplification primer binding site.

[00113] In some aspects, the nucleic acid adapter can also comprise a constant nucleic acid sequence to minimize ligation bias caused by differences in sequences of particular identifier oligonucleotides.

[00114] The nucleic acid adapter can be a partially double-stranded nucleic acid molecule. In aspects in which the nucleic acid adapter is partially double-stranded, the nucleic acid adapter comprises a double-stranded annealed region, a first single-stranded mismatched region, and a second single-stranded mismatched region. The first single-stranded mismatched region and the second single stranded mismatched region can be present on opposing sides of the double-stranded annealed region.

[00115] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released can be present in the double-stranded annealed region of the nucleic acid adapter.

[00116] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a constant nucleic acid sequence to minimize ligation bias, the constant nucleic acid sequence to minimize ligation bias can be present in the double-stranded annealed region of the nucleic acid adapter.

[00117] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a unique molecular identifier, the unique molecular identifier can be present in at least one of the first or second single-stranded mismatched regions of the nucleic acid adapter.

[00118] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a first and a second amplification primer binding site, the first amplification primer binding site can be present in the first single-stranded mismatched region of the nucleic acid adapter and the

second amplification primer binding site can be present in the second single-stranded mismatched region of the nucleic acid adapter.

[00119] After ligation of the at least one nucleic acid adapter, a two-ended adapter ligation method can further comprise: (5) amplifying the ligation product produced in step (4) using amplification primers that bind to the first and second amplification primer binding sites; and (6) identifying the released oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the sample.

[00120] A two-ended adapter ligation method of the present disclosure can further comprise, prior to step (4), performing an end repair reaction using methods known in the art. The method can also further comprise, prior to step (4), performing a tailing reaction to attach a single nucleotide overhang to the 3' ends of the identifier oligonucleotide using methods known in the art. In aspects, the end repair reaction and the tailing reaction can be performed sequentially or concurrently.

[00121] In preferred aspects of a two-ended adapter ligation method, a nucleic acid adapter is ligated to both ends of the released and collected identifier oligonucleotide.

[00122] In other aspects of a two-ended adapter ligation method, at least one of the amplification primers used in step (5) to amplify the ligation product produced in step (4) comprises a nucleic acid sequence which identifies the specific location of the tissue sample form which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated "ROI #1", at least one of the amplification primers would comprise a nucleic acid sequence that corresponds to "ROI #1".

[00123] Figure 1 shows a schematic of a preferred aspect of a two-ended adapter ligation method of the present disclosure. In this aspect, the probe comprises a target binding domain comprising an antibody that binds to a target protein. In the upper left panel, the probe binds to the target protein. In the upper right panel, a UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target protein bound to the target binding domain. In the bottom panel, a nucleic acid adapter is ligated to both ends of the identifier oligonucleotide. In this non-limiting example, the nucleic acid adapter is partially double-stranded and comprises a double-stranded annealed region, a first single-stranded mismatched region, and a second single-stranded mismatched region. Present in

the double-stranded annealed region is a constant nucleic acid sequence to minimize ligation bias and a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. Present in the first single-stranded mismatched region is a first amplification primer binding site. Present in the second single-stranded mismatched region is a unique molecular identifier and a second amplification primer binding site. Following ligation of the nucleic acid adapters to the identifier oligonucleotide, the product is amplified using amplification primers that bind the first and the second amplification primer binding sites and sequenced to identify the target protein bound by the probe.

[00124] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide dually ligated to two nucleic acid adapters for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide dually ligated to two nucleic acid adapters comprises an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which is capable of identifying a target analyte in a sample. Each end of the identifier oligonucleotide is attached to a nucleic acid adapter molecule, wherein the nucleic acid adapter molecule is partially double-stranded and comprises a double-stranded annealed region, a first single-stranded mismatched region and a second single-stranded mismatched region. The first single-stranded mismatched region and the second single stranded mismatched region are present on opposing sides of the double-stranded annealed region. The double-stranded mismatch region comprises a constant nucleic acid sequence to minimize ligation bias and a nucleic acid sequence nucleic acid sequence which is capable of identifying a specific location of a sample. The first single-stranded mismatched region comprises a first amplification primer binding site. The second single-stranded mismatched region comprises a second amplification primer binding site and a nucleic acid sequence comprising a unique molecular identifier. A schematic of an identifier oligonucleotide dually ligated to two nucleic acid adapters is shown in the bottom panel of Figure 1.

[00125] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "one-ended adapter ligation method".

[00126] A one-ended adapter ligation method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that

the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. The identifier oligonucleotide can also comprise a first amplification primer binding site. In some aspects, the identifier oligonucleotide also comprises at least one 3' end with a single nucleotide overhang.

[00127] Following contacting the at least one target analyte with the at least one probe, a one-ended adapter ligation method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00128] Following release of the identifier oligonucleotide, a one-ended adapter ligation method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00129] Following collection of the released identifier oligonucleotide, a one-ended adapter ligation method can further comprise: (4) ligating to the released oligonucleotide collected in step (3) at least one nucleic acid adapter;

[00130] The nucleic acid adapter can comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated "ROI #1", the nucleic acid adapter would comprise a nucleic acid sequence that corresponds to "ROI #1". The nucleic acid adapter can also comprise a unique molecular identifier. The nucleic acid adapter can also comprise a second amplification primer binding site.

[00131] In some aspects, the nucleic acid adapter can also comprise a constant nucleic acid sequence to minimize ligation bias caused by differences in sequences of particular identifier oligonucleotides. The constant nucleic acid sequence can comprise a cleavable moiety. The cleavable moiety can be enzymatically cleavable. In a non-limiting example, the enzymatically cleavable moiety can be a USER sequence, wherein the USER sequence comprises the sequence GUGUATUG.

[00132] The nucleic acid adapter can comprise any combination of the features described above.

[00133] The nucleic acid adapter can be a partially double-stranded nucleic acid molecule. In aspects in which the nucleic acid adapter is partially double-stranded, the nucleic acid adapter comprises a double-stranded annealed region and a single-stranded mismatched region.

[00134] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released can be present in the double-stranded annealed region of the nucleic acid adapter.

[00135] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a constant nucleic acid sequence to minimize ligation bias, the constant nucleic acid sequence to minimize ligation bias can be present in the double-stranded annealed region of the nucleic acid adapter.

[00136] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a unique molecular identifier, the unique molecular identifier can be present in the single-stranded mismatched region.

[00137] In aspects in which the nucleic acid adapter is partially double-stranded and comprises a second amplification primer binding site, the second amplification primer binding site can be present in the single-stranded mismatched region of the nucleic acid adapter.

[00138] After ligation of the at least one nucleic acid adapter, a two-ended adapter ligation method can further comprise: (5) amplifying the ligation product produced in step (4); and (6) identifying the released oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the sample.

[00139] In other aspects of a one-ended adapter ligation method of the present disclosure, at least one of the amplification primers used in step (5) to amplify the ligation product produced in step

(4) comprises a nucleic acid sequence which identifies the specific location of the tissue sample form which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated "ROI #1", at least one of the amplification primers would comprise a nucleic acid sequence that corresponds to "ROI #1".

[00140] Figure 2 shows a schematic of a preferred aspect of a one-ended adapter ligation method of the present disclosure. In this aspect, the probe comprises a target binding domain that is an antibody that binds to a target protein. In the upper left panel, the probe binds to the target protein. In upper right panel, the UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target protein and a first amplification primer binding site. In this non-limiting example, the identifier oligonucleotide is double-stranded with one strand that comprises three separate nucleic acid molecules. The identifier oligonucleotide also comprises one 3' end with a single nucleotide overhang.

[00141] In the bottom panel of Figure 2, a nucleic acid adapter is ligated to the end of the identifier oligonucleotide that comprises the 3' single nucleotide overhang. In this non-limiting example, the nucleic acid adapter is partially double-stranded and comprises a double-stranded annealed region and a single-stranded mismatched region. Present in the double-stranded annealed region is a constant nucleic acid sequence to minimize ligation bias and a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. Present in the single-stranded mismatched region is a unique molecular identifier and a second amplification primer binding site. Following ligation of the nucleic acid adapter to the identifier oligonucleotide, the product is amplified using amplification primers that bind to the first and the second amplification primer binding sites and sequenced to identify the target protein bound by the probe.

[00142] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide ligated to one nucleic acid adapter for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide ligated to one nucleic acid adapter comprises an identifier oligonucleotide, wherein the identifier oligonucleotide comprises a unique nucleic acid sequence which is capable of identifying a target analyte in a sample and a first amplification primer binding site. One end of the identifier oligonucleotide is attached to a nucleic acid adapter molecule,

wherein the nucleic acid adapter molecule is partially double-stranded and comprises a double-stranded annealed region and a single-stranded mismatched region and a second single-stranded mismatched region. The double-stranded mismatch region comprises a constant nucleic acid sequence to minimize ligation bias and a nucleic acid sequence nucleic acid sequence which is capable of identifying a specific location of a sample. The single-stranded mismatched region comprises a second amplification primer binding site. A schematic of an identifier oligonucleotide ligated to one nucleic acid adapter is shown in the bottom panel of Figure 2.

[00143] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "templated-primer extension method".

[00144] A templated-primer extension method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. The identifier oligonucleotide can also comprise a first amplification primer binding site.

[00145] Following contacting the at least one target analyte with the at least one probe, a templated-primer extension method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00146] Following release of the identifier oligonucleotide, a templated-primer extension method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier

oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00147] Following collection of the released identifier oligonucleotide, a templated-primer extension method can further comprise: (4) hybridizing to the released identifier oligonucleotide collected in step (3) a single stranded nucleic acid template.

[00148] The single stranded nucleic acid template can comprise a region complementary to the unique nucleic acid sequence of the identifier oligonucleotide, thereby allowing for the hybridization of the single stranded nucleic acid template and the collected identifier oligonucleotide.

[00149] The single stranded nucleic acid template can also comprise a nucleic acid sequence comprising a unique molecular identifier.

[00150] The single stranded nucleic acid template can also comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released.

[00151] The single stranded nucleic acid template can also comprise a nucleic acid sequence that is complementary to a second amplification primer binding site.

[00152] The single stranded nucleic acid template can comprise any combination of the features described above.

[00153] Following hybridization of the identifier oligonucleotide to the single stranded nucleic acid template, a templated-primer extension method can further comprise: (5) extending the identifier oligonucleotide of step (4) to form an extension produce complementary to the single stranded nucleic acid template, wherein the extension product comprises the identifier oligonucleotide and the sequence complementary to the single stranded nucleic acid template; (6) amplifying the extension product of step (6) using amplification primers that hybridize to the first and second amplification primer binding sites; and (7) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the sample.

[00154] In some aspects, the single stranded nucleic acid template can comprise an affinity molecule. In aspects in which the single stranded nucleic acid template comprises an affinity

molecule, a templated-primer extension method can further comprise an affinity purification step between steps (4) and (5).

[00155] Figure 3 shows a schematic of a preferred aspect of a templated-primer extension method of the present disclosure. In this aspect, the probe comprises a target binding domain that is an antibody that binds to a target protein. In the upper left panel, the probe binds to the target protein. In upper right panel, the UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target protein and a first amplification primer binding site. In the lower right panel, the identifier oligonucleotide is hybridized to a single stranded nucleic acid template. In this non-limiting example, the single-stranded nucleic acid template comprises an affinity molecule, a nucleic acid sequence complementary to the unique nucleic acid sequence of the identifier oligonucleotide, a first unique molecular identifier, and a sequence complementary to a second amplification primer biding site. The identifier oligonucleotide is extended to form an extension product complementary to the single stranded nucleic acid template. As shown in the lower left panel, the extension product comprises the identifier oligonucleotide, the nucleic acid sequence complementary to the first unique molecular identifier, and the second amplification primer binding site. Following the extension reaction, the primer extension product is amplified using amplification primers that bind to the first and the second amplification primer binding sites. In this non-limiting example, one of the amplification primers comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. The amplified product is then sequenced to identify the target protein bound by the probe.

[00156] In one aspect, the present disclosure provides a composition of a template-extended identifier oligonucleotide for spatially detecting at least one target analyte in a sample. The template-extended identifier oligonucleotide comprises a first flow cell adapter sequence suitable for sequencing, followed by a first unique molecular identifier, followed by an identifier oligonucleotide, followed by a second unique molecular identifier, followed by a second amplification primer binding site, followed by a third unique molecular identifier, followed by a second flow cell adapter sequence suitable for sequencing. The identifier oligonucleotide comprises a first amplification primer binding site and a unique nucleic acid sequence which is

capable of identifying a target analyte in a sample. A schematic of a template-extended identifier oligonucleotide is shown in the bottom panel of Figure 4.

[00157] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "short probe hybridization method".

[00158] A short probe hybridization method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain.

[00159] Following contacting the at least one target analyte with the at least one probe, a short probe hybridization method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00160] Following release of the identifier oligonucleotide, a short probe hybridization method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00161] Following collection of the released identifier oligonucleotide, a short probe hybridization method can further comprise: (4) hybridizing to the release identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe.

[00162] The first or the second nucleic acid probe can comprise a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. The first or the second nucleic acid probe can also comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. The first or the second nucleic acid probe can also comprise a nucleic acid sequence comprising unique molecular identifier. The first nucleic acid probe can comprise a first amplification primer binding site. The second nucleic acid probe can comprise a second amplification primer binding site.

[00163] The first or the second nucleic acid probe can comprise any combination of the features described above. In a preferred aspect depicted in Figure 5, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released is located 5' to the first amplification primer binding site and the unique molecular identifier is located 3' to the second amplification primer binding site.

[00164] In another preferred aspect depicted in Figure 6, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this same preferred aspect, the second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and the unique molecular identifier are located 5' to the first amplification primer binding site.

[00165] In another preferred aspect depicted in Figure 7, the first nucleic acid probe comprises a first amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this same preferred aspect, the second nucleic acid probe

comprises a second amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and the unique molecular identifier are located 3' to the second amplification primer binding site.

[00166] In another preferred aspect depicted in Figure 15, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence comprising a unique molecular identifier is located 3' to the first amplification binding site.

[00167] In another preferred aspect depicted in Figure 17, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence comprising a unique molecular identifier is located 5' to the first amplification binding site.

[00168] The first nucleic acid probe and the second nucleic acid probe can hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping. Alternatively, the first nucleic acid probe and the second nucleic acid probe can hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping.

[00169] Following hybridization of the first and the second nucleic acid probe to the identifier oligonucleotide, a short probe hybridization method can further comprise: (5) in the aspect in which the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping, ligating the first and the second nucleic acid probes together, for example, by performing a nick repair

reaction. Alternatively, in the aspect in which the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping, the method comprises ligating the first and the second nucleic acid probes together, for example, by performing a gap extension reaction and a nick repair reaction, such that the first and the second nucleic acid probes are ligated together.

[00170] Following ligation of the first and the second nucleic acid probe, a short probe hybridization method can further comprise: (6) amplifying the ligation product produced in step (5) using amplification primers that hybridize to the first and second amplification primer binding sites; and (7) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the sample.

[00171] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide-short nucleic acid probe complex for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide-short nucleic acid probe complex comprises an identifier oligonucleotide hybridized to a first nucleic acid probe and a second nucleic acid probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a target analyte in a sample. The first nucleic acid probe comprises a first amplification primer binding site, followed by a unique nucleic acid sequence capable of identifying a specific location in a sample, followed by a region complementary to the identifier oligonucleotide. The second nucleic acid probe comprises a second amplification primer binding site, followed by a nucleic acid sequence comprising a unique molecular identifier, followed by a region complementary to the identifier oligonucleotide. A schematic of an identifier oligonucleotide-short nucleic acid probe complex is depicted in Figure 5.

[00172] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide-short nucleic acid probe complex for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide-short nucleic acid probe complex comprises an identifier oligonucleotide hybridized to a first nucleic acid probe and a second nucleic acid probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a target analyte in a sample. The first nucleic acid probe comprises a first amplification primer binding site, followed by a nucleic acid sequence comprising a unique molecular identifier, followed by a region complementary to the identifier oligonucleotide,

wherein the nucleic acid sequence comprising a unique molecular identifier is located 3' to the first amplification primer binding site. The second nucleic acid probe comprises a second amplification primer binding site followed by a region complementary to the identifier oligonucleotide. A schematic of an identifier oligonucleotide-short nucleic acid probe complex is depicted in Figure 15.

[00173] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide-short nucleic acid probe complex for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide-short nucleic acid probe complex comprises an identifier oligonucleotide hybridized to a first nucleic acid probe and a second nucleic acid probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a target analyte in a sample. The first nucleic acid probe comprises a first amplification primer binding site, followed by a nucleic acid sequence comprising a unique molecular identifier, followed by a region complementary to the identifier oligonucleotide, wherein the nucleic acid sequence comprising a unique molecular identifier is located 5' to the first amplification primer binding site. The second nucleic acid probe comprises a second amplification primer binding site followed by a region complementary to the identifier oligonucleotide. A schematic of an identifier oligonucleotide-short nucleic acid probe complex is depicted in Figure 17.

[00174] Figure 16 shows a schematic overview of an exemplary short probe hybridization method of the present disclosure. First at least one target analyte in a sample is contacted with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target domain.

[00175] Following contacting the at least one target analyte with the at least one probe, a force is then provided to a location of the sample sufficient to release the identifier oligonucleotide. The identifier oligonucleotide is collected following release, as shown in the top panel of Figure 16.

[00176] As shown in the second panel from the top of Figure 16, the released identifier oligonucleotide is then hybridized to a first nucleic acid probe and a second nucleic acid probe. In this non-limiting example, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. The nucleic acid sequence comprising the unique molecular identifier is located 3' to the first amplification primer binding site. The second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this non-limiting example, the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping. Following hybridization to the identifier oligonucleotide, the first and second probe are ligated together, for example, by performing a nick repair reaction.

[00177] Following ligation of the first and second nucleic acid probes, the ligation product is amplified via PCR using amplification primers that hybridize to the first and second amplification primer binding sites. As shown in the second panel from the bottom of Figure 16, the amplification primer that hybridizes to the second amplification primer binding site comprises a first flow cell binding site, a first nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the second amplification primer binding site. The amplification primer that hybridizes to the first amplification primer binding site comprises a second flow cell binding site, a second nucleic acid sequence which identifies the specific location from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the first amplification primer binding site. The PCR product shown in the bottom panel of Figure 16 is then sequenced to identify the released oligonucleotide, thereby spatially detecting the at least one target analyte in the sample.

[00178] Figure 18 shows a schematic overview of an exemplary short probe hybridization method of the present disclosure. First, at least one target analyte in a sample is contacted with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target

nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target domain.

[00179] Following contacting the at least one target analyte with the at least one probe, a force is then provided to a location of the sample sufficient to release the identifier oligonucleotide. The identifier oligonucleotide is collected following release, as shown in the top panel of Figure 18.

[00180] As shown in the second panel from the top of Figure 18, the released identifier oligonucleotide is then hybridized to a first nucleic acid probe and a second nucleic acid probe. In this non-limiting example, the first nucleic acid probe comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. The nucleic acid sequence comprising the unique molecular identifier is located 5' to the first amplification primer binding site. The second nucleic acid probe comprises a second amplification primer binding site and a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. In this non-limiting example, the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping. Following hybridization to the identifier oligonucleotide, the first and second probe are ligated together, for example, by performing a nick repair reaction.

[00181] Following ligation of the first and second nucleic acid probes, the ligation product is amplified via PCR using amplification primers that hybridize to the first and second amplification primer binding sites. As shown in the second panel from the bottom of Figure 18, the amplification primer that hybridizes to the second amplification primer binding site comprises a first flow cell binding site, a first nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the second amplification primer binding site. The amplification primer that hybridizes to the first amplification primer binding site comprises a second flow cell binding site, a second nucleic acid sequence which identifies the specific location from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the first amplification primer binding site. The PCR product shown in the bottom panel Figure 18 is then

sequenced to identify the released oligonucleotide, thereby spatially detecting the at least one target analyte in the sample.

[00182] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "long probe hybridization method".

[00183] A long probe hybridization method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain.

[00184] Following contacting the at least one target analyte with the at least one probe, a long probe hybridization method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00185] Following release of the identifier oligonucleotide, a long probe hybridization method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00186] Following collection of the released identifier oligonucleotide(s), a long probe hybridization method can further comprise: (4) hybridizing to the released identifier oligonucleotide a first nucleic acid probe and a second nucleic acid probe.

[00187] The first or the second nucleic acid probe can comprise a nucleic acid sequence complementary to a portion of the identifier oligonucleotide. The first or the second nucleic acid probe can also comprise a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released.

[00188] The first or the second nucleic acid probe can also comprise a first unique molecular identifier. The first or the second nucleic acid probe can also comprise a second unique molecular identifier. The first or the second nucleic acid probe can also comprise a third unique molecular identifier.

[00189] The first nucleic acid probe can comprise a first amplification primer binding site.

[00190] The first nucleic acid probe can also comprise a first flow cell binding site. The second nucleic acid probe can comprise a second flow cell binding site.

[00191] The first and the second nucleic acid probes can comprise any combination of the features described above. In a preferred aspect depicted in Figure 8, the first nucleic acid probe comprises a first flow cell binding site, a first unique molecular identifier, a first amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second flow cell binding site, a second unique molecular identifier, a third unique molecular identifier and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and the first unique molecular identifier are located 5' to the first flow cell binding site and the second and the third unique molecular identifiers are located 3' to the second flow cell binding site.

[00192] In another preferred aspect depicted in Figure 9, the first nucleic acid probe comprises a first flow cell binding site, a first unique molecular identifier, a second unique molecular identifier, a first amplification primer binding site, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the identifier oligonucleotide. In the same preferred aspect, the second nucleic acid probe comprises a second flow cell binding site, a third unique molecular identifier and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the first unique molecular identifier, the second unique molecular

identifier and the nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released are located 5' to the first flow cell binding site and the third unique molecular identifier is located 3' to the second flow cell binding site.

[00193] In another preferred aspect depicted in Figure 10, the first nucleic acid probe comprises a first flow cell binding site, a first unique molecular identifier, a first amplification primer binding site and a nucleic acid sequence complementary to the identifier oligonucleotide. In this same preferred aspect, the second nucleic acid probe comprises a second flow cell binding site, a second unique molecular identifier, a third unique molecular identifier, a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the identifier oligonucleotide. In this preferred aspect, the first unique molecular identifier is located 5' to the first flow cell binding site and the second unique molecular identifier, the third unique molecular identifier, and the nucleic acid sequence which identifies the specific location of the sample form which the identifier oligonucleotide was released are located 3' to the second amplification primer binding site.

[00194] The first nucleic acid probe and the second nucleic acid probe can hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping. Alternatively, the first nucleic acid probe and the second nucleic acid probe can hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping.

[00195] Following hybridization of the first and the second nucleic acid probe to the identifier oligonucleotide, a long probe hybridization method can further comprise: (5) in the aspect in which the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are adjacent and are not overlapping, performing a nick repair reaction such that the first and the second nucleic acid probes are ligated together. Alternatively, in the aspect in which the first and the second nucleic acid probe hybridize to the identifier oligonucleotide such that the first and the second nucleic acid probes are not adjacent and are not overlapping, the method comprises performing a gap extension and a nick repair reaction such that the first and the second nucleic acid probes are ligated together.

[00196] The name method can further comprise: (6) amplifying the ligation product produced in step (5) using amplification primers that hybridize to the first and second amplification primer

binding sites; and (7) identifying the released identifier oligonucleotide by sequencing the amplified products produced in step (6), thereby spatially detecting the at least one target analyte in the sample.

[00197] In one aspect, the present disclosure provides a composition of an identifier oligonucleotide-long nucleic acid probe complex for spatially detecting at least one target analyte in a sample. An identifier oligonucleotide-long nucleic acid probe complex comprises an identifier oligonucleotide hybridized to a first nucleic acid probe and a second nucleic acid probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a target analyte in a sample. The first nucleic acid probe comprises a first flow cell binding site suitable for sequencing, followed by a first unique molecular identifier, followed by a first amplification primer binding site, followed by a unique nucleic acid sequence capable of identifying a specific location in a sample, followed by a region complementary to the identifier oligonucleotide. The second nucleic acid probe comprises a second flow cell binding site, followed by a second unique molecular identifier, followed by a third unique molecular identifier, followed by a region complementary to the identifier oligonucleotide. A schematic of an identifier oligonucleotide-short nucleic acid probe complex is depicted in Figure 8.

[00198] In another aspect, the present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "direct PCR method".

[00199] A direct PCR method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. The identifier oligonucleotide can also comprise a first amplification primer binding site, a second amplification primer binding site, or a unique molecular identifier. The identifier oligonucleotide can comprise any combination of these features. Any of these features

can also be flanked by regions comprising constant nucleic acid sequences of about 1 nucleotide to about 10 nucleotides.

[00200] Following contacting the at least one analyte with the at least one probe, a direct PCR method further comprises: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00201] A direct PCR method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00202] Following release of the identifier oligonucleotide, a direct PCR method can further comprise: (4) amplifying the released identifier oligonucleotide using a first amplification primer capable of binding to the first amplification primer binding site and a second amplification primer capable of binding to the second amplification primer binding site. In some aspects, at least one of the amplification primers comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. For example, if the identifier oligonucleotide was released from location of the sample designated "ROI #1", at least one of the amplification primers would comprise a nucleic acid sequence that corresponds to "ROI #1". In still other aspects, at least one of the amplification primers comprises a unique molecular identifier.

[00203] Following amplification, a direct PCR method of the present disclosure can further comprise: (5) identifying the released oligonucleotide by sequencing the amplified products produced in step (5), thereby spatially detecting the at least one target analyte in the sample.

[00204] Figure 11 shows a schematic of a preferred aspect of a direct PCR method of the present disclosure. In this aspect, the probe comprises a target binding domain comprising a nucleic acid sequence that is complementary to a target nucleic acid. In the upper panel, the probe hybridizes to the target nucleic acid. In the lower panel, a UV photo-cleavable linker located between the

target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a first amplification primer binding site, a second amplification primer binding site, a unique molecular identifier, and a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. Located between these four features are constant spacer regions that are 3 nucleotides in length. The identifier oligonucleotide is double-stranded and comprises a strand that comprises 3 separate nucleic acid molecules. After release, the identifier oligonucleotide is amplified using a first amplification primer that hybridizes to the first amplification primer binding site and comprises a nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a second amplification primer that hybridizes to the second amplification primer binding site. The amplified product is then sequenced to identify the target nucleic acid bound by the probe.

[00205] Figure 19 shows a schematic of a preferred aspect of a direct PCR method of the present disclosure. In this aspect, the identifier oligonucleotide comprises a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier, a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain and a second amplification primer binding site, as shown in the top panel of Figure 19. The identifier oligonucleotide is amplified using amplification primers that hybridize to the first and second amplification primer binding sites. As shown in the middle panel of Figure 19, the amplification primer that hybridizes to the second amplification primer binding site comprises a first flow cell binding site, a first nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the second amplification primer binding site. The amplification primer that hybridizes to the first amplification primer binding site comprises a second flow cell binding site, a second nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a nucleic acid sequence complementary to the first amplification primer binding site. The PCR product shown in the bottom panel of Figure 19 is sequenced to identify the released oligonucleotide, thereby spatially detecting the at least one target analyte in the sample.

[00206] In one aspect, the present disclosure provides a composition of a direct-PCR compatible identifier oligonucleotide for spatially detecting at least one target analyte in a sample. A direct-

PCR compatible identifier oligonucleotide comprises a first amplification primer binding site, followed a unique nucleic acid sequence which is capable of identifying a target analyte in a sample, followed by a unique molecular identifier, followed by a second amplification primer binding site. A schematic of a direct-PCR compatible identifier oligonucleotide is depicted in the lower panel of Figure 11.

[00207] In one aspect, the present disclosure provides a composition of a direct-PCR compatible identifier oligonucleotide for spatially detecting at least one target analyte in a sample. A direct-PCR compatible identifier oligonucleotide comprises a first amplification primer binding site, followed a nucleic acid sequence comprising a unique molecular identifier, followed by a unique nucleic acid sequence which is capable of identifying a target analyte in a sample, followed by a second amplification primer binding site. A schematic of a direct-PCR compatible identifier oligonucleotide is shown in the top panel of Figure 19.

[00208] In some aspects of the methods of the present disclosure at least one, or at least two, or at least three, or at least four, or at least five, or at least six, or at least seven, or at least eight, or at least nine, or at least ten, or at least eleven, or at least twelve, or at least thirteen, or at least fourteen, or at least fifteen, or at least sixteen, or at least seventeen, or at least eighteen, or at least nineteen, or at least twenty, or at least thirty, or at least forty, or at least fifty, or at least sixty, or at least seventy, or at least eighty, or at least ninety, or at least one hundred probes can been to a single target analyte. As used herein, the term "tiling" is used to describe when more than one probe of the present disclosure is bound to a target analyte. The top panel of Figure 27 shows the tiling of probes onto a target RNA. Tiling multiple probes onto a target analyte means that each target analyte will be individually detected multiple times, increasing the overall accuracy of the measurement. In a non-limiting example, as shown in the bottom panel of Figure 27, in the case where 10 probes are tiled onto a single target RNA, one of the probes may be incorrectly detected too many times (outlier high count probe), while another probe may be incorrectly detected too few times (outlier low count probe). However, the other 8 probes may be detected at a similar level, indicating that the two outliers should be discarded during analysis and the signals from the 8 probes used to generate a more accurate measurement of the abundance of the target RNA.

[00209] The present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as an "enzyme free method".

[00210] An enzyme free method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. An identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. An identifier oligonucleotide can also comprise a capture probe binding site.

[00211] Following contacting at least one target analyte with at least one probe, an enzyme free method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00212] Following release of an identifier oligonucleotide, an enzyme free method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00213] Following collection of a released identifier oligonucleotide, an enzyme free method can further comprise: (4) hybridizing to a released identifier oligonucleotide a capture probe.

[00214] A capture probe can comprise a region complementary to the capture probe binding site. A capture probe can also comprise an affinity molecule.

[00215] Following hybridization of a capture probe, an enzyme free method can further comprise: (5) Identifying a released identifier oligonucleotide by sequencing the hybridized product produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[00216] A hybridized product produced in step (4) can be sequenced using an enzyme free method of sequencing. Enzyme-free methods of sequencing have been described in, e.g., US2014946386 and US15/819,151 (US Patent No. 10,415,080), each of which is incorporated herein by reference in its entirety.

[00217] Figure 12 shows a schematic of a preferred aspect of an enzyme free method of the present disclosure. In this aspect, the probe comprises a target binding domain comprising a nucleic acid sequence that is complementary to a target nucleic acid. In the top panel, the probe hybridizes to the target nucleic acid. In the middle panel, a UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target nucleic acid bound to the target binding domain and a capture probe binding site. After release, the identifier oligonucleotide is hybridized to a capture probe, as depicted in the bottom panel. The capture probe comprises a nucleic acid sequence complementary to the capture probe binding site and an affinity molecule. The hybridized product is then sequenced using enzyme free sequencing methods to identify the target nucleic acid bound by the probe.

[00218] In one aspect, the present disclosure provides a composition of a hybridized identifier oligonucleotide-capture probe complex for spatially detecting at least one target analyte in a sample. A hybridized identifier oligonucleotide-capture probe complex comprises an identifier oligonucleotide hybridized to a capture probe. An identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a specific a target analyte in a sample and a capture probe binding site. A capture probe comprises an affinity molecule and a region complementary to the capture probe binding site. A schematic of a hybridized identifier oligonucleotide-capture probe complex is depicted in the bottom of panel Figure 12.

[00219] The present disclosure provides compositions and methods for spatially detecting at least one target analyte in a sample using the probes of the present disclosure in a method herein referred to as a "multiplexed enzyme free method".

[00220] A multiplexed enzyme free method of the present disclosure can comprise: (1) contacting at least one target analyte in a sample with at least one probe of the present disclosure. The at least one target analyte can be a target protein or a target nucleic acid. In the aspect that the at least one target analyte is a target protein, the probe can comprise a target binding domain that is a target protein-binding region that can specifically bind to a target protein of interest. In the aspect that the at least one target analyte is a target nucleic acid, the probe can comprise a target nucleic acid-binding region that can directly or indirectly hybridize to a target nucleic acid of interest. The probe further comprises an identifier oligonucleotide. The identifier oligonucleotide can comprise a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain. The identifier oligonucleotide can also comprise a capture probe binding site. The identifier oligonucleotide can also comprise a multiplexing probe binding site.

[00221] Following contacting the at least one target analyte with the at least one probe, a multiplexed enzyme free method can further comprise: (2) providing a force to a location of the sample sufficient to release the identifier oligonucleotide. In a non-limiting example, in aspects in which the probe comprises a photo-cleavable linker between the identifier oligonucleotide and the target binding domain, a region-of-interest (ROI) is excited with light of a sufficient wavelength capable of cleaving the photo-cleavable linker.

[00222] Following release of the identifier oligonucleotide, a multiplexed enzyme free method can further comprise: (3) collecting the released identifier oligonucleotide. By directing the force only to a specific location in step (2), identifier oligonucleotides are only released from probes within that location and not from probes located outside that location. Thus, identifier oligonucleotides are collected only for probes that are bound to targets within that location in step (3), thereby permitting detection of the identities and quantities of the targets (proteins and/or nucleic acids) located only within that location.

[00223] Following collection of the released identifier oligonucleotide, a multiplexed enzyme free method can further comprise: (4) hybridizing to the released identifier oligonucleotide a capture probe and a multiplexing probe.

[00224] A capture probe can comprise a region complementary to the capture probe binding site. A capture probe can also comprise an affinity molecule.

[00225] A multiplexing probe can comprise a region complementary to the multiplexing probe binding site. A multiplexing probe can also comprise a nucleic acid sequence which identifies

the specific location of the sample from which the identifier oligonucleotide was released and a region complementary to the multiplexing probe binding site.

[00226] Following hybridization of a capture probe and a multiplexing probe, a multiplexed enzyme free method can further comprise: (5) Identifying the released identifier oligonucleotide by sequencing the hybridized product produced in step (4), thereby spatially detecting the at least one target analyte in the at least one cell in a tissue sample.

[00227] A hybridized product produced in step (4) can be sequenced using an enzyme free method of sequencing. Enzyme-free methods of sequencing have been described in, e.g., US2014946386 and US15/819,151, each of which is incorporated herein by reference in its entirety.

[00228] Figure 13 shows a schematic of a preferred aspect of a multiplexed enzyme free method of the present disclosure. In this aspect, the probe comprises a target binding domain comprising a nucleic acid sequence that is complementary to a target nucleic acid. In the top panel, the probe hybridizes to the target nucleic acid. In the middle panel, a UV photo-cleavable linker located between the target binding domain and the identifier oligonucleotide is cleaved, releasing the identifier oligonucleotide. The identifier oligonucleotide comprises a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a capture probe binding site, and a multiplexing probe binding site, as shown in the middle panel. After release, the identifier oligonucleotide is hybridized to a capture probe and a multiplexing probe as shown in the lower panel. The capture probe comprises a nucleic acid sequence complementary to the capture probe binding site and an affinity molecule. The multiplexing probe comprises a nucleic acid sequence complementary to the multiplexing probe binding site and a nucleic acid sequence which identifies the specific location of a sample from which the identifier oligonucleotide was released. The hybridized product is then sequenced using enzyme free sequencing methods to identify the target nucleic acid bound by the probe.

[00229] In one aspect, the present disclosure provides a composition of a hybridized identifier oligonucleotide-capture probe-multiplex probe complex for spatially detecting at least one target analyte in a sample. A hybridized identifier oligonucleotide-capture probe-multiplex probe complex comprises an identifier oligonucleotide hybridized to a capture probe and a multiplexing probe. The identifier oligonucleotide comprises a unique nucleic acid sequence capable of identifying a specific a target analyte in a sample, a capture probe binding site and a

multiplexing probe binding site. The capture probe comprises an affinity molecule and a region complementary to the capture probe binding site. The multiplexing probe comprises a nucleic acid sequence which identifies the specific location of a sample from which the identifier oligonucleotide was released and a region complementary to the multiplexing probe binding site. A schematic of a hybridized identifier oligonucleotide-capture probe-multiplex probe complex is depicted in the bottom panel of Figure 13.

[00230] Figure 20 is an exemplary schematic of overview of the methods of the present disclosure. First, a sample on a microscope slide is contacted with a plurality of probes of the present disclosure (step 1 in Figure 20). The slide is then imaged and particular regions of interest (ROIs) are selected (step 2 in Figure 20). A specific ROI is then illuminated by UV light to release identifier oligonucleotides from probes bound within the ROI. The released identifier oligonucleotides are then collected via aspiration with a microcapillary. Following aspiration, the identifier oligonucleotides are transferred to a particular well within a 96 well plate. Steps 4 and 5 are then repeated for each ROI identified in step 2. After all ROIs have been illuminated and all released identifier oligonucleotides collected, the identifier oligonucleotides are sequenced using next generation sequencing methods to spatially detect at least one target analyte in the sample.

[00231] As described in the preceding, the present disclosure provides probes for the compositions and methods of spatially detecting at least one target analyte in a sample. The present disclosure provides probes comprising a target binding domain and an identifier oligonucleotide. The target binding domain is a region of the probe that specifically binds to at least one target analyte in a sample.

[00232] Probes of the present disclosure can be used for spatially detecting a target nucleic acid. In this aspect, the target binding domain can be a target nucleic acid-binding region. The target nucleic acid-binding region is preferably at least 15 nucleotides in length, and more preferably is at least 20 nucleotides in length. In specific aspects, the target nucleic acid-binding region is approximately 10 to 500, 20 to 400, 25, 30 to 300, 35, 40 to 200, or 50 to 100 nucleotides in length. Probes and methods for binding and identifying a target nucleic acid have been described in, *e.g.*, US2003/0013091, US2007/0166708, US2010/0015607, US2010/0261026, US2010/0262374, US2010/0112710, US2010/0047924, and US2014/0371088, each of which is incorporated herein by reference in its entirety.

[00233] The target nucleic acid-binding region can directly hybridize to a target nucleic acid present in a sample. Alternatively, the probes of the present disclosure can indirectly hybridize to a target nucleic acid present in a sample (via an intermediary oligonucleotide). Figure 14 illustrates a probe (or composition) of this aspect. The probe includes a target nucleic-acid binding domain which binds to a synthetic oligonucleotide (the intermediary oligonucleotide) that in turn binds to a target nucleic acid in a biological sample. It could be said that the intermediary oligonucleotide is a probe, as defined herein, since it comprises a nucleic acid backbone and is capable of binding a target nucleic acid. In these aspects, a probe's target nucleic acid-binding region hybridizes to a region of an intermediary oligonucleotide (*i.e.*, a synthetic oligonucleotide) which is different from the target nucleic acid present in a sample. Thus, the probe's target binding region is independent of the ultimate target nucleic acid in the sample. This allows economical and rapid flexibility in an assay design, as the target (present in a sample)-specific components of the assay are included in inexpensive and widely-available synthetic DNA oligonucleotides rather than the more expensive probes. Such synthetic oligonucleotides are simply designed by including a region that hybridizes to the target nucleic acid present in a sample and a region that hybridizes to a probe. Therefore, a single set of indirectly-binding probes can be used to detect an infinite variety of target nucleic acids (present in a sample) in different experiments simply by replacing the target-specific (synthetic) oligonucleotide portion of the assay.

[00234] A target nucleic acid may be DNA or RNA and preferably messenger RNA (mRNA) or miRNA.

[00235] Probes of the present disclosure can be used for detecting a target protein. In this aspect, the target binding domain can be a target protein-binding region. A target protein-binding region includes molecules or assembles that are designed to bind to at least one target protein, at least one target protein surrogate, or both and can, under appropriate conditions, form a molecular complex comprising the probe and the target protein. The target-protein binding region can include an antibody, a peptide, an aptamer, or a peptoid. The antibody can be obtained from a variety of sources, including but not limited to polyclonal antibody, monoclonal antibody, monospecific antibody, recombinantly expressed antibody, humanized antibody, plantibodies, and the like. The terms protein, polypeptide, peptide, and amino acid sequence are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be

linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids or synthetic amino acids. The terms also encompass an amino acid polymer that has been modified, for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term amino acid refers to either natural and/or unnatural or synthetic amino acids, including but not limited to glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. Probes and methods for binding and identifying a target protein have been described, *e.g.*, in US2011/0086774, the contents of which is incorporated herein by reference in its entirety.

[00236] An identifier oligonucleotide is a nucleic acid molecule that identifies the target analyte bound to the target binding domain. The identifier oligonucleotide comprises a unique nucleic acid sequence that identifies the target analyte bound to the target binding domain of the probe. In a non-limiting example, a probe with a target binding domain that binds to the protein P53 comprises an identifier oligonucleotide with a unique nucleic acid sequence that corresponds to P53, while a probe with a target binding domain that binds to the protein P97 comprises an identifier oligonucleotide with a unique nucleic acid sequence that corresponds to P97.

[00237] An identifier oligonucleotide can be DNA, RNA, or a combination of DNA and RNA.

[00238] In some aspects, an identifier oligonucleotide can comprise at least one amplification primer binding site. An amplification primer binding site is a nucleic acid sequence capable of binding to an amplification primer. An amplification primer can be used to amplify the nucleic molecule to which it is bound using methods known in the art, including, but not limited to, polymerase chain reaction (PCR).

[00239] In some aspects, an identifier oligonucleotide can comprise at least one unique molecular identifier.

[00240] An identifier oligonucleotide can be a single-stranded, a double-stranded, or a partially double-stranded nucleic acid molecule. In the aspects in which an identifier oligonucleotide is double-stranded or partially double-stranded, at least one of the two strands can comprise at least two separate nucleic acid molecules which, without being bound by theory, allows for denaturing of the identifier oligonucleotide at lower temperatures.

[00241] An identifier oligonucleotide can also comprise at least one 3' end that comprises a single nucleotide overhang.

[00242] An identifier oligonucleotide can also comprise a capture probe binding site. A capture probe binding site is a nucleic acid sequence to which a capture probe can bind.

[00243] A capture probe of the present disclosure can comprise a nucleic acid sequence complementary to a capture probe binding site. A capture probe can also comprise an affinity molecule.

[00244] An identifier oligonucleotide can also comprise a multiplexing probe binding site. A multiplexing probe binding site is a nucleic acid sequence to which a multiplexing probe can bind.

[00245] A multiplexing probe of the present disclosure can comprise a nucleic acid sequence complementary to a multiplexing probe binding site. A multiplexing probe can also comprise a nucleic acid sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released.

[00246] A probe of the present disclosure can include a region which permits the release of an identifier oligonucleotide following the application of a suitable force. In one non-limited example, the region is a cleavable motif (*e.g.*, a restriction enzyme site or cleavable linker). The cleavable motif allows release of an identifier oligonucleotide from a bound target nucleic acid or protein and the identifier oligonucleotide can then be collected and detected. The region which permits the release of an identifier oligonucleotide can be positioned between the target-binding domain and the identifier oligonucleotide, allowing for the release of the identifier oligonucleotide from the target binding domain. An identifier oligonucleotide is said to be releasable when it can be separated (*i.e.*, cleaved and released) from the remainder of the probe. Examples of cleavable motives include but are not limited to photo-cleavable linkers. Photo-cleavable linkers can be cleaved by light provided by a suitable coherent light source (*e.g.*, a laser and a UV light source) or a suitable incoherent light source (*e.g.*, an arc-lamp and a light-emitting diode (LED)).

[00247] In some aspects, the identifier oligonucleotide is collected from a solution proximal to, e.g., at least immediately above or surrounding, the point at which the identifier oligonucleotide is released or the at least one cell. The proximal solution may be collected by aspirating, e.g., via a pipette, a capillary tube, a microarray pin, a flow cell comprising holes, or another suitable aspirating system known in the art or any combination thereof. The capillary tube may comprise an optical device capable of transmitting a light force, e.g., UV light, to the at least one cell. The

pipette or a microarray pin may be attached to an array comprising a plurality of pipettes or microarray pins. The proximal solution may comprise an anionic polymer, e.g., dextran sulfate, and/or salmon sperm DNA and/or the collected signal oligonucleotide may be added to a solution comprising an anionic polymer, e.g., dextran sulfate, and/or salmon sperm DNA. Other non-specific blocking agents known in the art in addition to or instead of salmon sperm DNA may be used.

[00248] In some aspects, the identifier oligonucleotide is collected from a tissue, at least one cell or proximal to the point at which the identifier oligonucleotide is released via liquid laminar, turbulent, or transitional flow. The flow may be via a channel, e.g., having 25 to 500 m depth between the tissue and a fluidic device or impermeable barrier placed over the tissue.

[00249] In aspects where the target-binding domain of a probe is an antibody, the probe can be prepared using a cysteine bioconjugation method that is stable, site-specific to, preferably, the antibody's hinge-region heavy-chain. This preparation method provides relatively controllable identifier oligonucleotides to antibody stoichiometric ratios. A probe can comprise a plurality (i.e., more than one, e.g., 2, 3, 4, 5, or more) identifier oligonucleotides per antibody. Generally, "heavier" probes, which comprise 3 or 4 identifier oligonucleotides per antibody, are significantly less sensitive than antibodies lacking an identifier oligonucleotide or "lighter" probes, which comprise 1 or 2 identifier oligonucleotides per antibody.

[00250] In aspects, probes are provided to a sample at concentrations typically less than that used for immunohistochemistry (IHC) or for in situ hybridization (ISH). Alternately, the concentration may be significantly less than that used for IHC or ISH. For example, the probe concentration may be 2 fold less, 5 fold less, 10 fold less, 20 fold less, 25 fold less, 30 fold less, 50 fold less, 60 fold less, 70 fold less, 80 fold less, 90 fold less, 100 fold less, 200 fold less, 300 fold less, 400 fold less, 500 fold less, 600 fold less, 700 fold less, 800 fold less, 900 fold less, 1000 fold less, 2000 fold less, or less and any number in between. In aspects, probes are provided at a concentration of 100 nM, 70 nM, 60 nM, 50 nM, 40 nM, 30 nM, 20 nM, 10 nM, 9 nM, 8 nM, 7 nM, 6 nM, 5 nM, 4 nM, 3 nM, 2 nM, 1 nM, 0.9 nM, 0.8 nM, 0.7 nM, 0.6 nM, 0.5 nM, 0.4 nM, 0.3 nM, 0.2 nM, 0.1 nM, 0.09 nM, 0.08 nM, 0.07 nM, 0.06 nM, 0.05 nM, 0.04 nM, 0.03 nM, 0.02 nM, 0.01 nM, and less and any concentration in between.

[00251] Background noise, during protein detection, can be reduced by performing a negative purification of the intact probe molecule. This can be done by conducting an affinity purification

of the antibody or photo-cleavable linker after collection of eluate from a region of interest. Normally, released signal oligonucleotides will not be pulled out of solution. A protein-G or -O mechanism in a pipet tip, tube, or plate can be employed for this step. Such devices and reagents commercially available.

[00252] Background noise, during nucleic acid detection, can be reduced by performing a negative purification of the intact probe molecule. This can be done by conducting an affinity purification of the target binding domain or photo-cleavable linker after collection of eluate from a region of interest. Normally, released signal oligonucleotides will not be pulled out of solution. To assist in the negative purification, a universal purification sequence may be included in a probe, e.g., in the target binding domain.

[00253] Protein-targeting probes and nucleic acid-targeting probes may be applied simultaneously as long as conditions allow for binding of both a protein target and a nucleic acid target. Alternately, protein-targeting probes and nucleic acid-targeting probes may be applied sequentially when conditions allowing for binding of both a protein target and a nucleic acid target are not possible.

[00254] A set of probes is synonymous with a composition of probes. A set of probes includes at least one species of probes, *i.e.*, directed to one target. A set of probes preferably includes at least two, e.g., 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more species of probes. A probe set may include one or multiple copies of each species of probe.

[00255] A first set of probes only may be applied to a sample. Alternately, a second set (or higher number) of probes may be later applied to the sample. The first set and second (or higher number) may target only nucleic acids, only proteins, or a combination thereof.

[00256] In the present disclosure, two or more targets (*i.e.*, proteins, nucleic acids, or a combination thereof) are detected; 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more targets, and any number there between, are detected.

[00257] A set of probes may be pre-defined based upon the cell type or tissue type to be targeted. For example, if the tissue is a breast cancer, then the set of probes will include probes directed to proteins relevant to breast cancer cells (e.g., Her2, EGFR, and PR) and/or probes directed to proteins relevant to normal breast tissues. Additionally, the set of probes may be pre-defined

based upon developmental status of a cell or tissue to be targeted. Alternately, the set of probes may be pre-defined based upon subcellular localizations of interest, e.g., nucleus, cytoplasm, and membrane. For example, antibodies directed to Foxp3, Histone H3, or P-S6 label the nucleus, antibodies directed to CD3, CD4, PD-1, or CD45RO label the cytoplasm, and antibodies directed to PD-L1 label membranes.

[00258] A probe may be chemically synthesized or may be produced biologically using a vector into which a nucleic acid encoding the probe has been cloned.

[00259] Any probe or set of probes described herein may be used in methods and kits of the present disclosure.

[00260] For the herein-described probes, association of a unique nucleic acid sequence to a specific target nucleic acid or target protein is not fixed.

[00261] As described in the preceding, probes of the present disclosure can be used to detect a target nucleic acid or target protein present in any sample, *e.g.*, a biological sample. As will be appreciated by those in the art, the sample may comprise any number of things, including, but not limited to: cells (including both primary cells and cultured cell lines) and tissues (including cultured or explanted). In aspects, a tissue sample (fixed or unfixed) is embedded, serially sectioned, and immobilized onto a microscope slide. As is well known, a pair of serial sections will include at least one cell that is present in both serial sections. Structures and cell types, located on a first serial section will have a similar location on an adjacent serial section. The sample can be cultured cells or dissociated cells (fixed or unfixed) that have been immobilized onto a slide. A sample can be a formalin-fixed paraffin-embedded (FFPE) tissue sample.

[00262] In aspects, a tissue sample is a biopsied tumor or a portion thereof, i.e., a clinically-relevant tissue sample. For example, the tumor may be from a breast cancer. The sample may be an excised lymph node.

[00263] The sample can be obtained from virtually any organism including multicellular organisms, *e.g.*, of the plant, fungus, and animal kingdoms; preferably, the sample is obtained from an animal, *e.g.*, a mammal. Human samples are particularly preferred.

[00264] In some aspects, the probes, compositions, methods, and kits described herein are used in the diagnosis of a condition. As used herein the term diagnose or diagnosis of a condition includes predicting or diagnosing the condition, determining predisposition to the condition, monitoring treatment of the condition, diagnosing a therapeutic response of the disease, and

prognosis of the condition, condition progression, and response to particular treatment of the condition. For example, a tissue sample can be assayed according to any of the probes, methods, or kits described herein to determine the presence and/or quantity of markers of a disease or malignant cell type in the sample (relative to the non-diseased condition), thereby diagnosing or staging a disease or a cancer.

[00265] In general, samples attached to a slide can be first imaged using fluorescence (e.g., fluorescent antibodies or fluorescent stains (e.g., DAPI)) to identify morphology, regions of interest, cell types of interest, and single cells and then expression of proteins and/or nucleic acids can be digitally counted from the sample on the same slide.

[00266] Compositions and kits of the present disclosure can include probes and other reagents, for example, buffers and other reagents known in the art to facilitate binding of a protein and/or a nucleic acid in a sample, *i.e.*, for performing hybridization reactions.

[00267] A kit also will include instructions for using the components of the kit, including, but not limited to, information necessary to hybridize labeled oligonucleotides to a probe, to hybridize a probe to a target-specific oligonucleotide, to hybridize a target-specific oligonucleotide to a target nucleic acid and/or to hybridize a probe to target protein.

[00268] A region of interest may be a tissue type present in a sample, a cell type, a cell, or a subcellular structure within a cell.

[00269] Together, a comparison of the identity and abundance of the target proteins and/or target nucleic acids present in a first region of interest (*e.g.*, tissue type, a cell type (including normal and abnormal cells), and a subcellular structure within a cell) and the identity and abundance of the target proteins and/or target nucleic acids present in second region of interest or more regions of interest can be made using the methods of the present disclosure.

[00270] As described in the preceding, the products produced by the methods of the present disclosure can be used for nucleic acid amplification. In a preferred aspect, the nucleic acid amplification can be solid-phase nucleic acid amplification. Thus, in further aspects the invention provides a method of solid-phase nucleic acid amplification of template polynucleotide molecules which comprises: preparing a library of template polynucleotide molecules which have common sequences at their 5' and 3' ends using the methods of the present disclosure and carrying out a solid-phase nucleic acid amplification reaction wherein said template polynucleotide molecules are amplified. Compositions and methods for nucleic acid

amplification and sequencing have been described in, e.g., US9376678, which is incorporated herein by reference in its entirety.

[00271] The term "solid-phase amplification" as used herein refers to any nucleic acid amplification reaction carried out on or in association with a solid support such that all or a portion of the amplified products are immobilized on the solid support as they are formed. In particular, the term encompasses solid-phase polymerase chain reaction (solid-phase PCR), which is a reaction analogous to standard solution phase PCR, except that one or both of the forward and reverse amplification primers is/are immobilized on the solid support.

[00272] Although the invention encompasses "solid-phase" amplification methods in which only one amplification primer is immobilized (the other primer usually being present in free solution), it is preferred for the solid support to be provided with both the forward and the reverse primers immobilized. In practice, there will be a "plurality" of identical forward primers and/or a "plurality" of identical reverse primers immobilized on the solid support, since the PCR process requires an excess of primers to sustain amplification. References herein to forward and reverse primers are to be interpreted accordingly as encompassing a "plurality" of such primers unless the context indicates otherwise.

[00273] As will be appreciated by the skilled reader, any given PCR reaction requires at least one type of forward primer and at least one type of reverse primer specific for the template to be amplified. However, in certain aspects the forward and reverse primers may comprise template-specific portions of identical sequence, and may have entirely identical nucleotide sequence and structure (including any non-nucleotide modifications). In other words, it is possible to carry out solid-phase amplification using only one type of primer, and such single-primer methods are encompassed within the scope of the invention. Other aspects may use forward and reverse primers which contain identical template-specific sequences but which differ in some other structural features. For example one type of primer may contain a non-nucleotide modification which is not present in the other.

[00274] In other aspects of the invention the forward and reverse primers may contain template-specific portions of different sequence.

[00275] Amplification primers for solid-phase PCR are preferably immobilized by covalent attachment to the solid support at or near the 5' end of the primer, leaving the template-specific portion of the primer free for annealing to its cognate template and the 3' hydroxyl group free for

primer extension. Any suitable covalent attachment means known in the art may be used for this purpose. The chosen attachment chemistry will depend on the nature of the solid support, and any derivatization or functionalization applied to it. The primer itself may include a moiety, which may be a non-nucleotide chemical modification, to facilitate attachment. In one particularly preferred aspect the primer may include a sulphur-containing nucleophile, such as phosphorothioate or thiophosphate, at the 5' end. In the case of solid-supported polyacrylamide hydrogels (as described below), this nucleophile will bind to a "C" group present in the hydrogel. The most preferred means of attaching primers and templates to a solid support is via 5' phosphorothioate attachment to a hydrogel comprised of polymerised acrylamide and N-(5-bromoacetamidylpentyl)acrylamide (BRAPA).

[00276] The terms "cluster" and "colony" are used interchangeably herein to refer to a discrete site on a solid support comprised of a plurality of identical immobilized nucleic acid strands and a plurality of identical immobilized complementary nucleic acid strands. The term "clustered array" refers to an array formed from such clusters or colonies. In this context the term "array" is not to be understood as requiring an ordered arrangement of clusters.

[00277] The invention also encompasses methods of sequencing the amplified nucleic acids generated by solid-phase amplification. Thus, the invention provides a method of nucleic acid sequencing comprising amplifying a library of nucleic acid templates by the methods of the present disclosure described above, using solid-phase amplification as described above to amplify this library on a solid support, and carrying out a nucleic acid sequencing reaction to determine the sequence of the whole or a part of at least one amplified nucleic acid strand produced in the solid-phase amplification reaction.

[00278] Sequencing, as referred to herein, can be carried out using any suitable "sequencing-by-synthesis" technique, wherein nucleotides are added successively to a free 3' hydroxyl group, resulting in synthesis of a polynucleotide chain in the 5' to 3' direction. The nature of the nucleotide added is preferably determined after each nucleotide addition.

[00279] The initiation point for the sequencing reaction may be provided by annealing of a sequencing primer to a product of the whole genome or solid-phase amplification reaction. In this connection, one or both of the adapters added during formation of the template library may include a nucleotide sequence which permits annealing of a sequencing primer to amplified products derived by whole genome or solid-phase amplification of the template library.

[00280] The products of solid-phase amplification reactions wherein both forward and reverse amplification primers are covalently immobilized on the solid surface are so-called "bridged" structures formed by annealing of pairs of immobilized polynucleotide strands and immobilized complementary strands, both strands being attached to the solid support at the 5' end. Arrays comprised of such bridged structures provide inefficient templates for nucleic acid sequencing, since hybridization of a conventional sequencing primer to one of the immobilized strands is not favored compared to annealing of this strand to its immobilized complementary strand under standard conditions for hybridization.

[00281] In order to provide more suitable templates for nucleic acid sequencing it is preferred to remove substantially all or at least a portion of one of the immobilized strands in the "bridged" structure in order to generate a template which is at least partially single-stranded. The portion of the template which is single-stranded will thus be available for hybridization to a sequencing primer. The process of removing all or a portion of one immobilized strand in a "bridged" double-stranded nucleic acid structure may be referred to herein as "linearization".

[00282] Bridged template structures may be linearized by cleavage of one or both strands with a restriction endonuclease or by cleavage of one strand with a nicking endonuclease. Other methods of cleavage can be used as an alternative to restriction enzymes or nicking enzymes, including inter alfa chemical cleavage (e.g. cleavage of a diol linkage with periodate), cleavage of abasic sites by cleavage with endonuclease, or by exposure to heat or alkali, cleavage of ribonucleotides incorporated into amplification products otherwise comprised of deoxyribonucleotides, photochemical cleavage or cleavage of a peptide linker.

[00283] It will be appreciated that a linearization step may not be essential if the solid-phase amplification reaction is performed with only one primer covalently immobilized and the other in free solution.

[00284] In order to generate a linearized template suitable for sequencing it is necessary to remove "unequal" amounts of the complementary strands in the bridged structure formed by amplification so as to leave behind a linearized template for sequencing which is fully or partially single stranded. Most preferably one strand of the bridged structure is substantially or completely removed.

[00285] Following the cleavage step, regardless of the method used for cleavage, the product of the cleavage reaction may be subjected to denaturing conditions in order to remove the portion(s)

of the cleaved strand(s) that are not attached to the solid support. Suitable denaturing conditions will be apparent to the skilled reader with reference to standard molecular biology protocols.

[00286] Denaturation (and subsequent re-annealing of the cleaved strands) results in the production of a sequencing template which is partially or substantially single-stranded. A sequencing reaction may then be initiated by hybridization of a sequencing primer to the single-stranded portion of the template.

[00287] Thus, the nucleic acid sequencing reaction may comprise hybridizing a sequencing primer to a single-stranded region of a linearized amplification product, sequentially incorporating one or more nucleotides into a polynucleotide strand complementary to the region of amplified template strand to be sequenced, identifying the base present in one or more of the incorporated nucleotide(s) and thereby determining the sequence of a region of the template strand.

[00288] One preferred sequencing method which can be used in accordance with the invention relies on the use of modified nucleotides that can act as chain terminators. Once the modified nucleotide has been incorporated into the growing polynucleotide chain complementary to the region of the template being sequenced there is no free 3'-OH group available to direct further sequence extension and therefore the polymerase cannot add further nucleotides. Once the nature of the base incorporated into the growing chain has been determined, the 3' block may be removed to allow addition of the next successive nucleotide. By ordering the products derived using these modified nucleotides it is possible to deduce the DNA sequence of the DNA template. Such reactions can be done in a single experiment if each of the modified nucleotides has attached a different label, known to correspond to the particular base, to facilitate discrimination between the bases added at each incorporation step. Alternatively, a separate reaction may be carried out containing each of the modified nucleotides separately.

[00289] The modified nucleotides may carry a label to facilitate their detection. Preferably this is a fluorescent label. Each nucleotide type may carry a different fluorescent label. However the detectable label need not be a fluorescent label. Any label can be used which allows the detection of an incorporated nucleotide.

[00290] One method for detecting fluorescently labelled nucleotides comprises using laser light of a wavelength specific for the labelled nucleotides, or the use of other suitable sources of

illumination. The fluorescence from the label on the nucleotide may be detected by a CCD camera or other suitable detection means.

[00291] The invention is not intended to be limited to use of the sequencing method outlined above, as essentially any sequencing methodology which relies on successive incorporation of nucleotides into a polynucleotide chain can be used. Suitable alternative techniques include, for example, Pyrosequencing, FISSEQ (fluorescent in situ sequencing), MPSS (massively parallel signature sequencing) and sequencing by ligation-based methods.

[00292] In methods of the present disclosure, the unique nucleic acid sequence present in the identifier oligonucleotide of a probe which identifies the target analyte bound to the target binding domain of the probe can comprise between about 5 nucleotides and about 50 nucleotides. Preferably, the sequence comprises between about 20 nucleotides and about 40 nucleotides. Even more preferably, the sequence comprises about 35 nucleotides. In some preferred aspects, the sequence comprises 10 nucleotides.

[00293] In methods of the present disclosure, the nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released comprises between about 6 nucleotides and about 15 nucleotides. Preferably, the sequence comprises about 12 nucleotides.

[00294] In methods of the present disclosure, an amplification primer binding sites comprises between about 18 nucleotides and about 40 nucleotides. Preferably, an amplification primer binding sites comprises about 32 nucleotides.

[00295] In some aspects of the methods of the present disclosure, an amplification primer binding site can comprise an i7 sequence, wherein the i7 sequence comprises the sequence set forth in SEQ ID NO: 1.

[00296] In some aspects of the methods of the present disclosure, an amplification primer binding site can comprise an i5 sequence, wherein the i5 sequence comprises the sequence set forth in SEQ ID NO: 2.

[00297] In some aspects of the methods of the present disclosure, an amplification primer can comprise a flow cell adapter sequence, wherein the flow cell adapter sequence is suitable for sequencing. Preferably, at least one amplification primer used in the methods of the present disclosure comprises a P5 flow cell adapter sequence, wherein the P5 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 3. Preferably still, at least one amplification

primer used in the methods of the present disclosure comprises a P7 flow cell adapter sequence, wherein the P7 flow cell adapter sequence comprises the sequence set forth in SEQ ID NO: 4.

[00298] In methods of the present disclosure, a unique molecular identifier can comprise between about 6 nucleotides and about 30 nucleotides. Preferably, a unique molecular identifier can comprise about 15 nucleotides. The terms unique molecular identifier and random molecular tags are used interchangeably herein. Using methods known in that art, unique molecular identifiers are random sequences that can be used to correct for biases in amplification prior to sequencing.

[00299] In methods of the present disclosure, a constant nucleic acid sequence to minimize ligation bias comprises between about 1 nucleotide and about 15 nucleotides. Preferably, the constant sequence comprises about 8 nucleotides.

[00300] In some aspects, a flow cell binding site can comprise between about 15 to about 40 nucleotides. A flow cell binding site can comprise about 29 nucleotides. A flow cell binding site can comprise about 24 nucleotides.

[00301] In some aspects, a target binding domain can comprise between about 10 to about 70 nucleotides. A target binding domain can comprise between about 30 to about 55 nucleotides. A target binding domain can comprise between about 35 to about 50 nucleotides.

[00302] In some aspects, a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain can comprise between about 20 to about 40. A unique nucleic acid sequence which identifies the target analyte bound to the target binding domain can comprise about 25 nucleotides, or about 35 nucleotides, or about 12 nucleotides.

[00303] In some aspects, an amplification primer binding site can comprise between about 20 to about 50 nucleotides. An amplification primer binding site can comprise about 33 nucleotides, or about 34 nucleotides.

[00304] In some aspects, a nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can comprise between about 1 to about 20 nucleotides. A nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotide was released can comprise about 8 nucleotides.

[00305] In some aspects, a nucleic acid sequence comprising a unique molecular identifier can comprise between about 5 to about 20 nucleotides. A nucleic acid sequence comprising a unique molecular identifier can comprise about 14 nucleotides.

[00306] As used herein, the terms "region of interest" and "ROI" are used in their broadest sense to refer to a specific location within a sample that is to be analyzed using the methods of the present disclosure.

[00307] As used herein, the term "adjacent" can mean within about 1 nucleotide, or within about 2 nucleotides, or within about 3 nucleotides, or within about 4 nucleotides, or within about 5 nucleotides, or within about 6 nucleotides, or within about 7 nucleotides, or within about 8 nucleotides, or within about 9 nucleotides, or within about 10 nucleotides, or within about 11 nucleotides, or within about 12 nucleotides, or within about 13 nucleotides, or within about 14 nucleotides, or within about 15 nucleotides, or within about 16 nucleotides, or within about 17 nucleotides, or within about 18 nucleotides, or within about 19 nucleotides, or within about 20 nucleotides, or within about 21 nucleotides, or within about 22 nucleotides, or within about 23 nucleotides, or within about 24 nucleotides, or within about 25 nucleotides, or within about 26 nucleotides, or within about 27 nucleotides, or within about 28 nucleotides, or within about 29 nucleotides, or within about 30 nucleotides, or within about 40 nucleotides, or within about nucleotides, or within about 50 nucleotides, or within about 60 nucleotides, or within about 70 nucleotides, or within about 80 nucleotides, or within about 90 nucleotides, or within about 100 nucleotides.

[00308] As used herein, the term "spatially detecting" is used in its broadest sense to refer to the identification of the presence of a specific target analyte within a specific region of interest in a sample. Spatially detecting can comprise quantifying the amount of a specific target analyte present within a specific region of interest in a sample. Spatially detecting can further comprise quantifying the relative amount of a first target analyte within a specific region of interest in a sample as compared to the amount of at least a second target analyte within a specific region of interest in a sample. Spatially detecting can also comprise quantifying the relative amount of a specific target analyte within a first region of interest in a sample compared to the amount of the same target analyte in at least a second region of interest in the same sample or different sample.

[00309] In some aspects of the methods and compositions of the present disclosure, a target analyte can be any molecule within a sample that is to be spatially detected. Target analytes include, but are not limited to, nucleic acid molecules and protein molecules. When the target analyte is a protein, the protein can be referred to as a target protein. When the target analyte is a nucleic acid, the nucleic acid can be referred to as a target nucleic acid. Target nucleic acids can include, but

are not limited to, mRNA molecules, micro RNA (miRNA) molecules, tRNA molecules, rRNA molecules, gDNA or any other nucleic acid present within a sample.

[00310] In some aspects of the methods and compositions of the present disclosure, the term target binding domain is used in its broadest sense to refer to a portion of a probe of the present disclosure that binds to, either directly or indirectly, a target analyte located in a sample. A target binding domain can comprise nucleic acid, protein, at least one antibody, an aptamer, or any combination thereof. A target binding domain can comprise DNA, RNA or any combination thereof. A target binding domain can comprise any number of modified nucleotides and/or nucleic acid analogues.

[00311] In the aspect that the target analyte to be spatially detected is a target protein, a target binding domain can be a protein-target binding domain. A protein-target binding domain can comprise an antibody or antibody fragment that binds to the target protein.

[00312] In the aspect that the target analyte to be spatially detected is a target nucleic acid, a target binding domain can be a target nucleic acid-binding region. A target nucleic acid-binding region can comprise a nucleic acid that is complementary to the target nucleic acid to be spatially detected. A target nucleic acid-binding region can comprise a nucleic acid that hybridizes to the target nucleic acid to be detected.

[00313] As used herein, the term "hybridize" is used in its broadest sense to mean the formation of a stable nucleic acid duplex. In one aspect, "stable duplex" means that a duplex structure is not destroyed by a stringent wash under conditions such as, for example, a temperature of either about 5 °C below or about 5 °C above the Tm of a strand of the duplex and low monovalent salt concentration, *e.g.,* less than 0.2 M, or less than 0.1 M or salt concentrations known to those of skill in the art. A duplex can be "perfectly matched", such that the polynucleotide and/or oligonucleotide strands making up the duplex form a double stranded structure with one another such that every nucleotide in each strand undergoes Watson-Crick base pairing with a nucleotide in the other strand. The term "duplex" comprises, but is not limited to, the pairing of nucleoside analogs, such as deoxyinosine, nucleosides with 2-aminopurine bases, PNAs, and the like, that can be employed. A duplex can comprise at least one mismatch, wherein the term "mismatch" means that a pair of nucleotides in the duplex fail to undergo Watson-Crick bonding.

[00314] As used herein, the term "hybridization conditions," will typically include salt concentrations of less than about 1 M, more usually less than about 500 mM and even more usually less than about 200 mM. Hybridization temperatures can be as low as 5 °C, but are typically greater

than 22 °C, more typically greater than about 30 °C, and often in excess of about 37 °C. Hybridizations are usually performed under stringent conditions, *e.g.,* conditions under which a probe will specifically hybridize to its target analyte. Stringent conditions are sequence-dependent and are different in different circumstances. Longer fragments can require higher hybridization temperatures for specific hybridization. As other factors can affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents and extent of base mismatching, the combination of parameters is more important than the absolute measure of any one alone. Certain hybridization conditions will promote the formation of a duplex between the entire length of a target binding domain and the target analyte. Other hybridization conditions will promote the formation of a duplex only along certain portions of the target binding domain.

[00315] In some aspects of the methods and compositions of the present disclosure, a probe can comprise a target binding domain directly, or indirectly linked to an identifier oligonucleotide. In the context of a probe, an identifier oligonucleotide is a polynucleotide that comprises a nucleic acid sequence that identifies the target analyte bound to the target binding domain of that probe. That is to say, the identifier oligonucleotide comprises a specific nucleic acid sequence that is *a priori* assigned to the specific target analyte bound to the target binding to which the identifier oligonucleotide is attached. In a non-limiting example, a probe designated as "probe *X*" designed to spatially detect "target analyte *X*" comprises a target binding domain designated "target binding domain *X*" linked to an identifier oligonucleotide designated "identifier oligonucleotide *X*". Target binding domain *X* binds to target analyte *X* and identifier oligonucleotide *X* comprises a nucleic acid sequence, designated as "nucleic acid sequence *X*", which corresponds to target analyte *X*. Thus, if a skilled artisan practicing the methods of the present disclosure were to collect identifier oligonucleotides released from a region of interest in sample and obtain nucleic acid sequence *X* after sequencing, the skilled artisan would understand that to indicate that target analyte *X* was present in that region of interest. The amount, or number of sequencing reads, of nucleic acid sequence *X* can be used to determine the quantify, in absolute or relative terms, the amount of target analyte *X* within the region of interest.

[00316] As used herein, the term "amplification primer binding site" is used in its broadest sense to refer to a nucleic acid sequence that is complementary to, or at least partially complementary to at least one amplification primer, wherein the amplification primer is a short single-stranded or

partially single-stranded oligonucleotide that is sufficient to prime DNA and/or RNA synthesis, for example, by PCR.

[00317] In some aspects of the methods and compositions of the present disclosure, a target binding domain can be linked to an identifier oligonucleotide by a cleavable linker. Suitable cleavable linkers include, but are not limited to, chemically cleavable linkers (*e.g.* a linker that is cleaved when exposed to a particular chemical, combination of chemicals or reaction conditions), a photo-cleavable linker (*e.g.* a linker that is cleaved when exposed to light of a sufficient wavelength or light comprising a sufficient range of wavelengths), or an enzymatically cleavable linker (*e.g.* a linker that is cleaved by a specific enzyme or class of enzymes). Thus, as used herein the phrase "providing a force to a location of the sample sufficient to release an identifier oligonucleotide" is used in its broadest sense to describe changing the conditions within a certain region of interest in a sample such that, for any probe bound to a target analyte within that region of interest, the linker between the target binding domain of the probe and the identifier oligonucleotide of the probe is cleaved, thereby separating the identifier oligonucleotide from the target binding domain so that the identifier oligonucleotide can be subsequently collected from solution. For example, in aspects wherein a probe comprises a chemically cleavable linker between the target binding domain and the identifier oligonucleotide, providing a force to a location of the sample sufficient to release an identifier oligonucleotide can comprise exposing that location of the sample to the specific chemical, combination of chemicals or reaction conditions that catalyze the cleavage of the linker. In another non-limiting example, in aspects wherein a probe comprises a photo-cleavable linker between the target binding domain and the identifier oligonucleotide, providing a force to a location of the sample sufficient to release an identifier oligonucleotide can comprise exposing/exciting that location of the sample with light of a sufficient wavelength capable of cleaving the photo-cleavable linker. In another non-limiting example, in aspects wherein a probe comprises an enzymatically cleavable linker between the target binding domain and the identifier oligonucleotide, providing a force to a location of the sample sufficient to release an identifier oligonucleotide can comprise exposing that location of the sample to an amount of enzyme sufficient to catalyze the cleavage of the linker.

[00318] Providing a force to a location of the sample sufficient to release an identifier oligonucleotide can result in at least about 10%, or at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least

about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 99% of probes bound to target analytes within that location of the sample to undergo cleavage of the linker connecting the target binding domain and the identifier oligonucleotide.

[00319] As would be appreciated by one skilled in the art, the term "unique molecular identifier" or "UMI" refer to short nucleic acid sequences that are used to quantify and reduce quantitative bias caused by nucleic acid amplification prior to sequencing reactions.

[00320] In some aspects of the methods and compositions of the present disclosure, an affinity moiety can comprise biotin, avidin, streptavidin, nucleic acid, or any combination thereof.

[00321] In some aspects of the methods and compositions of the present disclosure, a probe can comprise at least at least about 5, about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00322] In some aspects of the methods and compositions of the present disclosure, a target binding domain can comprise at least at least about 5, about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00323] In some aspects of the methods and compositions of the present disclosure, an identifier oligonucleotide can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00324] In some aspects of the methods and compositions of the present disclosure, an amplification primer can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00325] In some aspects of the methods and compositions of the present disclosure, a nucleic acid probe can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95, at least about 100, at least about 110, at least about 120, at least about 130, at least about 140, at least about 150, at least about 160, at least about 170, at least about 180, at least about 190 or at least about 200 nucleotides.

[00326] In some aspects of the methods and compositions of the present disclosure, a nucleic acid complementary to a portion of a identifier oligonucleotide can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95 or at least about 100 nucleotides.

[00327] In some aspects of the methods and compositions of the present disclosure, a nucleic acid sequence comprising a molecular identifier can comprise at least about 5, or at least about 10 nucleotides, or at least about 15, or at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 nucleotides.

[00328] In some aspects of the methods and compositions of the present disclosure, an amplification primer binding site can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35, at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65 or at least about 70 nucleotides.

[00329] In some aspects of the methods and compositions of the present disclosure, a flow cell adapter sequence suitable for sequencing can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35 at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85, at least about 90, at least about 95 or at least about 100 nucleotides.

[00330] In some aspects of the methods and compositions of the present disclosure, a flow cell binding site can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35, at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 90, at least about 95 or at least about 100 nucleotides.

[00331] In some aspects of the methods and compositions of the present disclosure, a nucleic acid sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released can comprise at least about 5, or at least about 10 nucleotides, or at least about 15, or at least about 20, or at least about 25, or at least about 30, or at least about 35, or at least about 40, or at least about 45, or at least about 50 nucleotides.

[00332] In some aspects of the methods and compositions of the present disclosure, a unique nucleic acid sequence which identifies the target analyte bound to a target binding domain can comprise at least about 5, at least about 10, at least about 15, at least about 20, at least about 25, at least about 30, at least about 35, at least about 40, at least about 45, at least about 50, at least about 55, at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 90, at least about 95 or at least about 100 nucleotides.

[00333] In some aspects of the methods and compositions of the present disclosure, a probe, a target binding domain, an identifier oligonucleotide, an amplification primer, a nucleic acid probe, a nucleic acid complementary to a portion of a identifier oligonucleotide, a nucleic acid sequence comprising a molecular identifier, an amplification primer binding site, a flow cell adapter sequence, a flow cell binding site, a nucleic acid sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released, a unique nucleic acid sequence which identifies the target analyte bound to a target binding domain or any combination thereof can comprise at least one natural base, can comprise no natural bases, can comprise at least one modified nucleotide or nucleic acid analog, can comprise no modified

nucleotides or nucleic acid analogs, can comprise at least one universal base, can comprise no universal bases, can comprise at least one degenerate base or can comprise no degenerate bases.

[00334] In some aspects of the methods and compositions of the present disclosure, a probe, a target binding domain, an identifier oligonucleotide, an amplification primer, a nucleic acid probe, a nucleic acid complementary to a portion of a identifier oligonucleotide, a nucleic acid sequence comprising a molecular identifier, an amplification primer binding site, a flow cell adapter sequence, a flow cell binding site, a nucleic acid sequence which identifies the specific location of the tissue sample from which an identifier oligonucleotide was released, a unique nucleic acid sequence which identifies the target analyte bound to a target binding domain or any combination thereof can comprise any combination natural bases (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more natural bases), modified nucleotides or nucleic acid analogs (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more modified or analog nucleotides), universal bases (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more universal bases), or degenerate bases (*e.g.* 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more degenerative bases). When present in a combination, the natural bases, modified nucleotides or nucleic acid analogs, universal bases and degenerate bases can be arranged in any order.

[00335] The terms "modified nucleotides" or "nucleic acid analogues" include, but are not limited to, locked nucleic acids (LNA), bridged nucleic acids (BNA), propyne-modified nucleic acids, zip nucleic acids (ZNA®), isoguanine and isocytosine. Preferably, the modified nucleotides or nucleic acid analogues are locked nucleic acids (LNAs).

[00336] The term "locked nucleic acids (LNA)" as used herein includes, but is not limited to, a modified RNA nucleotide in which the ribose moiety comprises a methylene bridge connecting the 2' oxygen and the 4' carbon. This methylene bridge locks the ribose in the 3'-endo confirmation, also known as the north confirmation, that is found in A-form RNA duplexes. The term inaccessible RNA can be used interchangeably with LNA. The term "bridged nucleic acids (BNA)" as used herein includes, but is not limited to, modified RNA molecules that comprise a five-membered or six-membered bridged structure with a fixed 3'-endo confirmation, also known as the north confirmation. The bridged structure connects the 2' oxygen of the ribose to the 4' carbon of the ribose. Various different bridge structures are possible containing carbon, nitrogen, and hydrogen atoms. The term "propyne-modified nucleic acids" as used herein includes, but is not limited to, pyrimidines, namely cytosine and thymine/uracil, that comprise a propyne modification at the C5 position of the nucleic acid base. The term "zip nucleic acids

(ZNA®)" as used herein includes, but is not limited to, oligonucleotides that are conjugated with cationic spermine moieties.

[00337] The term "universal base" as used herein includes, but is not limited to, a nucleotide base does not follow Watson-Crick base pair rules but rather can bind to any of the four canonical bases (A, T/U, C, G) located on the target nucleic acid. The term "degenerate base" as used herein includes, but is not limited to, a nucleotide base that does not follow Watson-Crick base pair rules but rather can bind to at least two of the four canonical bases A, T/U, C, G), but not all four. A degenerate base can also be termed a Wobble base; these terms are used interchangeably herein.

[00338] As used in this Specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

[00339] Unless specifically stated or obvious from context, as used herein, the term "or" is understood to be inclusive and covers both "or" and "and".

[00340] Unless specifically stated or obvious from context, as used herein, the term "about" is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from the context, all numerical values provided herein are modified by the term "about."

[00341] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although other probes, compositions, methods, and kits similar, or equivalent, to those described herein can be used in the practice of the present disclosure, the preferred materials and methods are described herein. It is to be understood that the terminology used herein is for the purpose of describing particular aspects only, and is not intended to be limiting.

EXAMPLES

[00342] Example 1—two-ended adapter ligation method for 96 multiplexed samples

[00343] In this example, a two-ended adapter ligation method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. The nucleic acid adapters used in this experiment were partially double-stranded. The nucleic acid adapters comprised a first strand and a second strand. The first strand comprised a 5' phosphate moiety for ligation. The first strand also comprised a constant nucleic acid sequence to minimize ligation

bias (GCGTAGTG), a nucleic acid sequence comprising a unique molecular identifier, a unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a first amplification primer binding site (SEQ ID NO: 2). The second strand comprised a single overhanging thymine nucleotide at the 3' end, a sequence complementary to the constant nucleic acid sequence to minimize ligation bias present in the first strand, a sequence complementary to the unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released present in the first strand and a second amplification primer binding site (SEQ ID NO: 1).

[00344] To form the partially double stranded nucleic acid adapters, first strand oligonucleotides and second strand oligonucleotides were combined in equimolar proportion for a final total oligonucleotide concentration of 28 μM in buffer comprising 50 mM NaCl. The oligonucleotide mixture was heated at 95°C for 2 minutes and cooled at ambient temperature for 30 minutes, thereby annealing the first stand and second strand oligonucleotides together to form the partially double-stranded nucleic acid adapters. Annealed nucleic acid adaptors were diluted to final concentration ranging between 0.02 μM to 0.002 μM in a solution of 10 mM Tris pH 8 and 0.05% Tween20.

[00345] Collected identifier oligonucleotides were end repaired and A-tailed using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) with a modified protocol. End repair/A-tail master mix was prepared by combining the following: 627.8 μL of PCR-grade H2O, 143.9 μL of NEBNext Ultra II End Prep Reaction Buffer, and 61.7 μL of NEBNext Ultra II End Prep Enzyme Mix. 8.3 μL of end repair/A-tail master mix was added to 4 μL of each sample of identifier oligonucleotides. The reaction was incubated for 30 minutes at 20°C with a heated lid of >75°C, followed by a second incubation for 30 minutes at 65°C. The repaired/A-tailed identifier oligonucleotide mixtures were then stored at 4°C.

[00346] Following end repair and A-tailing, the nucleic acid adaptors were ligated to the repaired/A-tailed identifier oligonucleotides by adding 6.4 μL of NEBNext Ultra II Ligation Master Mix, 0.2 μL of NEBNext Ligation Enhancer, and 1 μL of the nucleic acid adapter dilution to each repaired/A-tailed identifier oligonucleotide mixture. These reactions were incubated for 15 minutes at 20°C with the heated lid off and subsequently quenched with 1 μL

0.5M EDTA. All of the reactions were then pooled into a single 15 mL conical tube to form a pooled adapter-ligated sample.

[00347] The pooled adaptor-ligated sample was purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 350 µL of AMPure XP beads and 3.15 mL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). AMPure XP bead cleanup was performed with 3.5 mL of diluted AMPure XP beads and eluted in 200 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. AMPure XP bead cleanup was then repeated with 400 µL of AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20 to obtain purified adapter-ligated samples.

[00348] Following AMPure XP cleanup, PCR reactions with purified adaptor-ligated sample were prepared to amplify the adapter-ligated identifier oligonucleotides. To 6 µL of the purified adapter-ligated sample, 10 µL of NEBNext Ultra II Q5 Master Mix, 0.2 µL of 100 µM forward and reverse primers, and 3.6 µL of PCR-grade H2O was added. The forward primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the first amplification primer binding site located on the first strand of the nucleic acid adapter. Table 1 provides the sequences of the forward primers used.

Table 1. Forward primers for two-ended adapter ligation.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACGCTCAGATATAGCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	5
AATGATACGGCGACCACCGAGATCTACACGCTCAGAATAGAGGCACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	6
AATGATACGGCGACCACCGAGATCTACACGCTCAGACCTATCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	7
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGGCTCTGAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	8
AATGATACGGCGACCACCGAGATCTACACGCTCAGAAGGCGAAGACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	9
AATGATACGGCGACCACCGAGATCTACACGCTCAGATAATCTTAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	10
AATGATACGGCGACCACCGAGATCTACACGCTCAGACAGGACGTACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	11
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGTACTGACACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	12

[00349] The reverse primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the second

amplification primer binding site located on the second strand of the nucleic acid adapter. Table 2 provides the sequences of the reverse primers used.

Table 2. Reverse primers for two-ended adapter ligation

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	13
CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	14
CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	15
CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	16
CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	17
CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	18
CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	19
CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	20
CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	21
CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	22
CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	23
CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	24

[00350] The optimal number of PCR cycles was determined empirically with triplicate PCR reactions. Alternately, the optimal number of PCR cycles could have been determined using real-time/qPCR. The PCR program used comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 1 minute at 65°C

(4) Repeating steps (2) and (3) nine times

(5) 5 minutes at 65°C

[00351] The amplified products were purified using 18 µL of AMPure XP beads and eluting with 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00352] The amplified products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The amplified products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with a custom spike-in primer comprising the nucleotide sequence ACACTCTTTAAGACGACGTCGCTATGGCCTCTCC (SEQ ID NO: 25).

[00353] Example 2—one-ended adapter ligation method for 96 multiplexed samples

[00354] In this example, a one-ended adapter ligation method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. The nucleic acid adapters used in this experiment were partially double-stranded. The nucleic acid adapters comprised a first strand and a second strand. The first strand comprised a 5' phosphate moiety for ligation. The first strand also comprised a constant nucleic acid sequence to minimize ligation bias (CACTACGC), a nucleic acid sequence comprising a unique molecular identifier, a unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a first amplification primer binding site (SEQ ID NO: 1). The second strand comprised a single overhanging thymine nucleotide at the 3' end, a sequence complementary to the constant nucleic acid sequence to minimize ligation bias present in the first strand and a sequence complementary to the unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released present in the first strand.

[00355] To form the partially double stranded nucleic acid adapters, first strand oligonucleotides and second strand oligonucleotides were combined in equimolar proportion for a final total oligonucleotide concentration of 28µM in 50 mM NaCl. The oligonucleotide mixture was heated to 95°C for 2 minutes and cooled at ambient temperature for 30 minutes, thereby annealing the first stand and second strand oligonucleotides together to form the partially double-stranded nucleic acid adapters. Annealed nucleic acid adaptors were diluted to a final concentration ranging between 0.02 µM to 0.002 µM in a solution of 10 mM Tris pH 8 and 0.05% Tween20.

[00356] Nucleic acid adapters were ligated to the collected identifier oligonucleotides by addition of 10 µL of 2X rapid ligation buffer (Enzymatics), 1 µL of T4 DNA Rapid Ligase (Enzymatics),

and 1 µL of annealed nucleic acid adapter dilutions to each sample of collected identifier oligonucleotides. Samples were incubated for 15 minutes at 20°C with the heated lid off and subsequently quenched with 1 µL 0.5M EDTA. All of the reactions were then pooled into a single 15 mL conical tube to form a pooled adapter-ligated sample.

[00357] The pooled adaptor-ligated sample was purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 350 µL of AMPure XP beads and 3.15 mL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). AMPure XP bead cleanup was performed with 3.5 mL of diluted AMPure XP beads and eluted in 200 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. AMPure XP bead cleanup was then repeated with 400 µL of AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20 to obtain purified adapter-ligated samples.

[00358] Following AMPure XP cleanup, PCR reactions with purified adaptor-ligated sample were prepared to amplify the adapter-ligated identifier oligonucleotides. To 6 µL of the purified adapter-ligated sample, 10 µL of NEBNext Ultra II Q5 Master Mix, 0.2 µL of 100 µM forward and reverse primers, and 3.6 µL of PCR-grade H_2O was added. The forward primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the first amplification primer binding site located on the first strand of the nucleic acid adapter. Table 3 provides the sequences of the forward primers used.

Table 3. Forward primers for two-ended adapter ligation.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACGCTCAGATATAGCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	5
AATGATACGGCGACCACCGAGATCTACACGCTCAGAATAGAGGCACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	6
AATGATACGGCGACCACCGAGATCTACACGCTCAGACCTATCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	7
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGGCTCTGAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	8
AATGATACGGCGACCACCGAGATCTACACGCTCAGAAGGCGAAGACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	9
AATGATACGGCGACCACCGAGATCTACACGCTCAGATAATCTTAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	10
AATGATACGGCGACCACCGAGATCTACACGCTCAGACAGGACGTACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	11
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGTACTGACACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	12

[00359] The reverse primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the second amplification primer binding site located on the second strand of the nucleic acid adapter. Table 4 provides the sequences of the reverse primers used.

Table 4. Reverse primers for two-ended adapter ligation.

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	13
CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	14
CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	15
CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	16
CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	17
CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	18
CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	19
CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	20
CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	21
CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	22
CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	23
CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	24

[00360] The optimal number of PCR cycles was determined empirically with triplicate PCR reactions. Alternately, the optimal number of PCR cycles could have been determined using real-time/qPCR. The PCR program used comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 1 minute at 65°C

(4) Repeating steps (2) and (3) nine times

(5) 5 minutes at 65°C

[00361] The amplified products were purified using 18 µL of AMPure XP beads and eluting with 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00362] The amplified products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The amplified products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with a custom spike-in primer comprising the nucleotide sequence ACACTCTTTAAGACGACGTCGCTATGGCCTCTCC (SEQ ID NO: 25).

[00363] Example 3—templated-primer extension method for 96 multiplexed samples

[00364] In this example, a templated-primer extension method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. The single stranded nucleic acid templates used in this example comprised a 3' biotin moiety, a region complementary to the unique nucleic acid sequences present in the collected identifier oligonucleotides, a nucleic acid sequence comprising a unique molecular identifier and a second amplification primer binding sequence (GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT, SEQ ID NO: 26). Table 5 provides the sequences of the single-stranded nucleic acid templates used in this example.

Table 5. Single stranded nucleic acid templates for templated-primer extension method

Single-stranded nucleic acid templates	SEQ ID NO
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNTTGAAGCACACCGTTTTTCTTTCTTCTTTCACGG	27
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNACCCACAGGTTATACGGGATTATCCGGTTATCCA	28
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNCGACACCGAGTTCGACCGTTATGTTGGTAGGATC	29
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNCGGTGTGTAAGCGTAACGATGTTGGTGTCGCTCT	30
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNCAGACACTGCGACAACTCACGATCATGACACAGA	31
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNATATTCTGTACTCAGTGCCTATCCACCTAATAGG	32
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNTTCAGTTATAATGTGTCCAGCAGAAGCAGGAATT	33
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNGTCCTTTGTTGGGCGGACCGTAATGAGGAATTTG	34
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNNGATGAGACTTCTACATGTCCGATGTTTTTGTGCT	35

GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNACTC ACACATAGTACTGACACGTAAGATAGGATG	36
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTTAC CCTATCTCGTCTATGTACGTCAGGCTGAAT	37
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNATCA ACGTAGGGTAAGGTCATATTTTTACCTTAC	38
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTTCC CTCTTTCTCCGCTTATGGATGAAAGGACAG	39
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCCTG CACAGTGAGTTTCTTTCACTCTAACTCTCT	40
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTGTC GCTCTAGTGTGACTTTTCCACCTCGCATCT	41
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNATAT CTTTCTCGGGTAAAGATTAGGCGTCCGATA	42
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCGAT TAGCCGTAGACGCAACTCATTGCCGAAGAT	43
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTGTG AGCATTTCAGTACGAGTGATGCAGATAAAC	44
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTATA GTTACCAAGTACTATGGGTTGGTGGAAGCC	45
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCCAA TTATACTGTCTGTTATGTTCTCGGATAAGC	46
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTCAG GTGCTTGTAGGCTCATGATAGGGGTAATGC	47
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCTCT GCTGTAATCTCAGCTCCACTTGTTTCTAAG	48
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNGTGC ATATTGCAGCTGAGCCAGCTCAATTTGAAG	49
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCCGT TGATTTACGCAACAGCGGCTTATATAGCTC	50
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCATC ATCGACAGTTCGCAGCCCTATAACATGATA	51
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNATCG CAGGATGGTACAGCATCATACATGATGAGC	52
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCTGA TAAGTCGTAGGAATGTCGCTTAATACGGAT	53
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNATGG CGGTTTCGGGTCCTGCACTATTCCTAATAA	54
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNCCAG TACGGGTACTAATAAGTGTCATATCTATTG	55
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNTGTT GGAGAGGTTAGAGGTGAGGAGGCGAAGATA	56

[00365] Single stranded nucleic acid templates were ordered from Integrated DNA Technologies, Inc. and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). Individual single stranded nucleic acid templates were normalized to a standard concentration

and then pooled to be equimolar. The pool of single stranded nucleic acid templates was diluted to 0.83 nM in a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00366] The collected identifier oligonucleotides were hybridized to the single stranded nucleic acid templates and extended by addition of 10 µL of NEBNext Ultra II Q5 Master Mix (New England Biolabs), 4 µL of the diluted single stranded nucleic acid template pool and 4 µL of H_2O to 2 µL of each sample of identifier oligonucleotides. The following PCR program was used to extend the identifier oligonucleotides:

(1) 30 seconds at 98°C, 10×

(2) 1 minutes at 98°C,

(3) 1 minutes at 68°C

(4) 1 minutes at 72°C

(5) Repeating steps (2)-(4) ten times

(6) 2 minutes at 72°C

[00367] The extension products were stored at 4°C. Magnetic streptavidin beads (MyOne Streptavidin C1 beads, Thermo Fisher Scientific) were washed in 1X Binding and Washing Buffer (5 mM Tris-HCl, 0.5 mM EDTA, 1M NaCl), and 5 µL of streptavidin beads were added to each extension product sample. The extension product samples were incubated with the beads on an orbital mixer for a minimum of 15 minutes. Following incubation, the samples were heated to 95°C for 3 minutes and transferred to a magnetic plate. Supernatant was extracted immediately after sufficient bead pelleting to yield the purified extension product samples.

[00368] The purified extension product samples were amplified by adding to 7.5 µL of each purified extension product sample, 12.5 µL of NEBNext Ultra II Q5 Master Mix, 0.25 µL of 100 µM forward primer, 1 µL of 25 µM reverse primer and 3.8 µL of PCR-grade H_2O. The forward primer comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the first amplification primer binding site located on the identifier oligonucleotide. Table 6 provides the sequences of the forward primers used in this example.

Table 6. Forward primers for templated-primer extension method.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACGCTCAGATATAGCCTACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	5
AATGATACGGCGACCACCGAGATCTACACGCTCAGAATAGAGGCACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	6
AATGATACGGCGACCACCGAGATCTACACGCTCAGACCTATCCTACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	7
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGGCTCTGAACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	8
AATGATACGGCGACCACCGAGATCTACACGCTCAGAAGGCGAAGACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	9
AATGATACGGCGACCACCGAGATCTACACGCTCAGATAATCTTAACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	10
AATGATACGGCGACCACCGAGATCTACACGCTCAGACAGGACGTACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	11
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGTACTGACACACTCTTTA AGACGACGTCGCTATGGCCTCTCC	12

[00369] The reverse primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the second amplification primer binding site located single-stranded nucleic acid template. Table 7 provides the sequences of the reverse primers used.

Table 7. Reverse primers for templated-primer extension method.

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATGTCGGTAAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	57
CAAGCAGAAGACGGCATACGAGATAGGTCACTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	58
CAAGCAGAAGACGGCATACGAGATGAATCCGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	59
CAAGCAGAAGACGGCATACGAGATGTACCTTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	60
CAAGCAGAAGACGGCATACGAGATCATGAGGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	61
CAAGCAGAAGACGGCATACGAGATTGACTGACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	62
CAAGCAGAAGACGGCATACGAGATCGTATTCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	63
CAAGCAGAAGACGGCATACGAGATCTCCTAGAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	64
CAAGCAGAAGACGGCATACGAGATTAGTTGCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	65

CAAGCAGAAGACGGCATACGAGATGAGATACGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	66
CAAGCAGAAGACGGCATACGAGATAGGTGTACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	67
CAAGCAGAAGACGGCATACGAGATTAATGCCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	68
CAAGCAGAAGACGGCATACGAGATTCAGACGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	69
CAAGCAGAAGACGGCATACGAGATGATAGGCTGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	70
CAAGCAGAAGACGGCATACGAGATTGGTACAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	71
CAAGCAGAAGACGGCATACGAGATCAAGGTCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	72
CAAGCAGAAGACGGCATACGAGATGCTATCCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	73
CAAGCAGAAGACGGCATACGAGATATGGAAGGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	74
CAAGCAGAAGACGGCATACGAGATTCAAGGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	75
CAAGCAGAAGACGGCATACGAGATGTTACGCAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	76
CAAGCAGAAGACGGCATACGAGATAGTCTGTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	77
CAAGCAGAAGACGGCATACGAGATGCACGTAAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	78
CAAGCAGAAGACGGCATACGAGATAACCTTGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	79
CAAGCAGAAGACGGCATACGAGATATTGCGTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	80
CAAGCAGAAGACGGCATACGAGATACCTGGAAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	81
CAAGCAGAAGACGGCATACGAGATGGAGATGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	82
CAAGCAGAAGACGGCATACGAGATGTACTCTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	83
CAAGCAGAAGACGGCATACGAGATGTAACGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	84
CAAGCAGAAGACGGCATACGAGATATTCCTCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	85
CAAGCAGAAGACGGCATACGAGATGTGTTCCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	86
CAAGCAGAAGACGGCATACGAGATAAGCACTGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	87
CAAGCAGAAGACGGCATACGAGATCTAGCAAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	88
CAAGCAGAAGACGGCATACGAGATTGCTTCCAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	89
CAAGCAGAAGACGGCATACGAGATGCTTAGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	90

CAAGCAGAAGACGGCATACGAGATAACCGTTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	91
CAAGCAGAAGACGGCATACGAGATGACATTCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	92
CAAGCAGAAGACGGCATACGAGATAGACCGTAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	93
CAAGCAGAAGACGGCATACGAGATGATACTGGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	94
CAAGCAGAAGACGGCATACGAGATTGCGTAGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	95
CAAGCAGAAGACGGCATACGAGATTCGGTTACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	96
CAAGCAGAAGACGGCATACGAGATATGACGTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	97
CAAGCAGAAGACGGCATACGAGATGCTGTAAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	98
CAAGCAGAAGACGGCATACGAGATGCAATGGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	99
CAAGCAGAAGACGGCATACGAGATATCTCGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	100
CAAGCAGAAGACGGCATACGAGATGGCTATTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	101
CAAGCAGAAGACGGCATACGAGATGGTGTCTTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	102
CAAGCAGAAGACGGCATACGAGATTCAACTGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	103
CAAGCAGAAGACGGCATACGAGATCTTCACCAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	104
CAAGCAGAAGACGGCATACGAGATACGGTCTTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	105
CAAGCAGAAGACGGCATACGAGATTCTCGCAAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	106
CAAGCAGAAGACGGCATACGAGATGGAATTGCGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	107
CAAGCAGAAGACGGCATACGAGATACGGATTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	108
CAAGCAGAAGACGGCATACGAGATTTAAGCGGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	109
CAAGCAGAAGACGGCATACGAGATTGCAGGTAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	110
CAAGCAGAAGACGGCATACGAGATCAATCGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	111
CAAGCAGAAGACGGCATACGAGATGTGCCATAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	112
CAAGCAGAAGACGGCATACGAGATTGTTCGAGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	113
CAAGCAGAAGACGGCATACGAGATTGGAGTTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	114
CAAGCAGAAGACGGCATACGAGATACGATGACGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	115

CAAGCAGAAGACGGCATACGAGATTGATGTCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	116
CAAGCAGAAGACGGCATACGAGATTGAACCTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	117
CAAGCAGAAGACGGCATACGAGATCTTCGTTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	118
CAAGCAGAAGACGGCATACGAGATCTTCTGAGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	119
CAAGCAGAAGACGGCATACGAGATTGCTCATGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	120
CAAGCAGAAGACGGCATACGAGATAGTTCGTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	121
CAAGCAGAAGACGGCATACGAGATTAGCGTCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	122
CAAGCAGAAGACGGCATACGAGATGGCGTTATGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	123
CAAGCAGAAGACGGCATACGAGATGGTGATTCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	124
CAAGCAGAAGACGGCATACGAGATAACTTGCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	125
CAAGCAGAAGACGGCATACGAGATGCAAGATCGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	126
CAAGCAGAAGACGGCATACGAGATTCGCATTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	127
CAAGCAGAAGACGGCATACGAGATTGTACACCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	128
CAAGCAGAAGACGGCATACGAGATAGCTCCTAGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	129
CAAGCAGAAGACGGCATACGAGATGCAATTCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	130
CAAGCAGAAGACGGCATACGAGATCTTAGGACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	131
CAAGCAGAAGACGGCATACGAGATGTCCTAAGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	132
CAAGCAGAAGACGGCATACGAGATAACGTGGAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	133
CAAGCAGAAGACGGCATACGAGATCTGTGTTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	134
CAAGCAGAAGACGGCATACGAGATGTTAAGGCGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	135
CAAGCAGAAGACGGCATACGAGATCACCTTACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	136
CAAGCAGAAGACGGCATACGAGATTGGTAGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	137
CAAGCAGAAGACGGCATACGAGATCAGTGAAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	138
CAAGCAGAAGACGGCATACGAGATGTTCAACCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	139
CAAGCAGAAGACGGCATACGAGATTGGCTATCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	140

CAAGCAGAAGACGGCATACGAGATCTGGAGTAGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	141
CAAGCAGAAGACGGCATACGAGATTCTCTTCCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	142
CAAGCAGAAGACGGCATACGAGATTCTAACGCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	143
CAAGCAGAAGACGGCATACGAGATGGTCAGATGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	144
CAAGCAGAAGACGGCATACGAGATCTCTGGTTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	145
CAAGCAGAAGACGGCATACGAGATTGTGGTACGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	146
CAAGCAGAAGACGGCATACGAGATCCTATACCGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	147
CAAGCAGAAGACGGCATACGAGATTTCTCTCGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	148
CAAGCAGAAGACGGCATACGAGATGTATGCTGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	149
CAAGCAGAAGACGGCATACGAGATAAGTCGAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	150
CAAGCAGAAGACGGCATACGAGATAACCGAAGGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCT	151
CAAGCAGAAGACGGCATACGAGATTGTTGTGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	152

[00370] The PCR program used to amplify the purified extension products comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) eighteen times

(6) 2 minutes at 72°C

[00371] The amplified extension products were stored at 4°C. 4 µL of each PCR reaction were combined into four pools, for 24 samples per pool.

[00372] The pooled PCR reactions were purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 100 µL of AMPure XP beads and 400 µL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). Purification was performed with 76.8 µL of diluted AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. Beads were retained and cleanup process was repeated

with 24 μL of AMPure XP buffer and eluted in 20 μL of buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00373] The purified PCR products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The purified PCR products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with a custom spike-in primer comprising the nucleotide sequence ACACTCTTTAAGACGACGTCGCTATGGCCTCTCC (SEQ ID NO: 25).

[00374] Example 4—long probe hybridization method for 96 multiplexed samples

[00375] In this example, a long probe hybridization method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. In this example, the first nucleic acid probe comprises a 5' phosphate moiety, a nucleic acid sequence complementary to a portion of the identifier oligonucleotide, a first amplification primer binding site comprising an i7 sequence (SEQ ID NO: 1), unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released and a P7 flow cell adapter sequence (SEQ ID NO: 4). The second nucleic acid probe comprises a nucleic acid sequence complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier, a second amplification primer binding site comprising and i5 sequence (SEQ ID NO: 2) and a P5 flow cell adapter sequence (SEQ ID NO: 3).

[00376] The first and second nucleic acid probes were ordered from Integrated DNA Technologies, Inc. and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). Individual nucleic acid probes were normalized to a standard concentration, pooled to be equimolar, and diluted to 0.83 nM in a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. The nucleic acid probes and the identifier oligonucleotides were hybridized by combining 0.5 μL of diluted nucleic acid probe pool with 2 μL of a mixture of identifier oligonucleotides collected from a sample solution in a buffer comprising 50 mM NaCl. This mixture was heated for 2 minutes at 95°C and cooled for 30 minutes at ambient temperature to yield an annealed identifier oligonucleotide-nucleic acid probe mixture.

[00377] In the case in which the first and the second nucleic acid probes hybridized to the identifier oligonucleotide such that the first and the second nucleic acid probes were not adjacent and were not overlapping, a gap extension reaction was performed. To 2.5 μL of each annealed

identifier oligonucleotide-nucleic acid probe mixture, 3.8 µL of NEBNext Ultra II Q5 Master mix and 1.3 µL of PCR-grade H_2O was added. The mixture was then subjected to the following Gap extension temperature cycle:

(1) 30 seconds at 98°C

(2) 1 minute at 98°C

(3) 1 minute at 68°C

(4) 1 minute at 72°C,

(5) Repeat steps (2)-(4) ten times

(6) 2 minutes at 72°C

[00378] The gap extension products were then stored at 4°C. The first and second nucleic acid probes were then ligated together by adding to 1 µL of the gap extension product, 10 µL of 2X rapid ligation buffer (Enzymatics), 1 µL of T4 DNA Rapid Ligase (Enzymatics), and 8 µL of PCR-grade H_2O. These ligation reactions were incubated for 15 minutes 20°C, subsequently quenched with 1 µL 0.5M EDTA, and pooled into a single 15 mL conical tube.

[00379] In the case in which the first and the second nucleic acid probes hybridized to the identifier oligonucleotide such that the first and the second nucleic acid probes were adjacent and were not overlapping, nick repair ligation reaction was performed. To 2.5 µL of each annealed identifier oligonucleotide-nucleic acid probe mixture, 10 µL of 2X rapid ligation buffer (Enzymatics), 1 µL of T4 DNA Rapid Ligase (Enzymatics), and 1 µL of PCR-grade H_2O was added. These ligation reactions were incubated for 15 minutes at 20°C, subsequently quenched with 1 µL of 0.5M EDTA, and pooled into a single 15 mL conical tube.

[00380] The pools of quenched ligation reactions were then purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 350 µL of AMPure XP beads and 3.15 mL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). The purification was performed with 3.5 mL of diluted AMPure XP beads and eluted in 200 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. The AMPure XP bead cleanup was then repeated with 400 µL of AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00381] To amplify the purified ligation products, PCR reactions with purified ligation products and primers were prepared. To 6 µL of purified ligation product, 10 µL of NEBNext Ultra II Q5 Master Mix, 0.2 µL of 100 µM forward primer (CAAGCAGAAGACGGCATACGA, SEQ ID

NO: 153) and reverse primer (AATGATACGGCGACCACCGA, SEQ ID NO: 154) and 3.6 µL of PCR-grade H_2O was added. The PCR program used to amplify the purified extension products comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) eighteen times

(6) 2 minutes at 72°C

[00382] The amplified products were stored at 4°C. 4 µL of each PCR reaction were combined into six pools, for 16 samples per pool. The amplified products were further purified using an AMPure XP bead cleanup with 64 µL of AMPure XP beads and eluting with 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00383] The purified amplified products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The purified amplified products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with either standard sequencing primers or a custom spike-in Read1 primer (SEQ ID NO: 25).

[00384] Example 5—short probe hybridization method for 96 multiplexed samples

[00385] In this example, a short probe hybridization method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. In this example, the first nucleic acid probe comprises a 5' phosphate moiety, a nucleic acid sequence complementary to a portion of the identifier oligonucleotide, a first amplification primer binding site comprising an i7 sequence (SEQ ID NO: 1) and a unique nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotide was released. The second nucleic acid probe comprises a nucleic acid sequence complementary to a portion of the identifier oligonucleotide, a nucleic acid sequence comprising a unique molecular identifier and a second amplification primer binding site comprising an i5 sequence (SEQ ID NO: 2).

[00386] The first and second nucleic acid probes were ordered from Integrated DNA Technologies, Inc. and quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher

Scientific). Individual nucleic acid probes were normalized to a standard concentration, pooled to be equimolar, and diluted to 0.83 nM in a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. The nucleic acid probes and the identifier oligonucleotides were hybridized by combining 0.5 µL of diluted nucleic acid probe pool with 2 µL of a mixture of identifier oligonucleotides collected from a sample solution in a buffer comprising 50 mM NaCl. This mixture was heated for 2 minutes at 95°C and cooled for 30 minutes at ambient temperature to yield an annealed identifier oligonucleotide-nucleic acid probe mixture.

[00387] In the case in which the first and the second nucleic acid probes hybridized to the identifier oligonucleotide such that the first and the second nucleic acid probes were not adjacent and were not overlapping, a gap extension reaction was performed. To 2.5 µL of each annealed identifier oligonucleotide-nucleic acid probe mixture, 3.8 µL of NEBNext Ultra II Q5 Master mix and 1.3 µL of PCR-grade H_2O was added. The mixture was then subjected to the following Gap extension temperature cycle:

(1) 30 seconds at 98°C

(2) 1 minute at 98°C

(3) 1 minute at 68°C

(4) 1 minute at 72°C,

(5) Repeat steps (2)-(4) ten times

(6) 2 minutes at 72°C

[00388] The gap extension products were then stored at 4°C. The first and second nucleic acid probes were then ligated together by adding to 1 µL of the gap extension product, 10 µL of 2X rapid ligation buffer (Enzymatics), 1 µL of T4 DNA Rapid Ligase (Enzymatics), and 8 µL of PCR-grade H_2O. These ligation reactions were incubated 15 min 20°C, quenched with 1 µL 0.5M EDTA, and pooled into a single 15 mL conical tube.

[00389] In the case in which the first and the second nucleic acid probes hybridized to the identifier oligonucleotide such that the first and the second nucleic acid probes were adjacent and were not overlapping, nick repair ligation reaction was performed. To 2.5 µL of each annealed identifier oligonucleotide-nucleic acid probe mixture, 10 µL of 2X rapid ligation buffer (Enzymatics), 1 µL of T4 DNA Rapid Ligase (Enzymatics), and 1 µL of PCR-grade H_2O was added. These ligation reactions were incubated for 15 minutes at 20°C, subsequently quenched with 1 µL of 0.5M EDTA, and pooled into a single 15 mL conical tube.

[00390] The pools of quenched ligation reactions were then purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 350 µL of AMPure XP beads and 3.15 mL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). The purification was performed with 3.5 mL of diluted AMPure XP beads and eluted in 200 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20. The AMPure XP bead cleanup was then repeated with 400 µL of AMPure XP beads and eluted in 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00391] To amplify the purified ligation products, PCR reactions with purified ligation products and primers were prepared. To 6 µL of purified ligation product, 10 µL of NEBNext Ultra II Q5 Master Mix, 0.2 µL of 100 µM forward primer and reverse primer and 3.6 µL of PCR-grade H_2O was added. . The forward primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the first amplification primer binding site located on the first strand of the nucleic acid adapter. Table 8 provides the sequences of the forward primers used.

Table 8. Forward primers for short probe hybridization method.

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	13
CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	14
CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	15
CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	16
CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	17
CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	18
CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	19
CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	20
CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	21
CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	22
CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	23

CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCT	24

[00392] The reverse primers comprised a flow cell adapter sequence suitable for sequencing, a unique molecular identifier and a nucleic acid sequence complementary to the second amplification primer binding site located on the second strand of the nucleic acid adapter. Table 9 provides the sequences of the reverse primers used.

Table 9. Reverse primers for short probe hybridization method.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACA CGACGCTCTTCCGATCT	155
AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCCTAC ACGACGCTCTTCCGATCT	156
AATGATACGGCGACCACCGAGATCTACACCCTATCCTACACTCTTTCCCTACA CGACGCTCTTCCGATCT	157
AATGATACGGCGACCACCGAGATCTACACGGCTCTGAACACTCTTTCCCTACA CGACGCTCTTCCGATCT	158
AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTTCCCTAC ACGACGCTCTTCCGATCT	159
AATGATACGGCGACCACCGAGATCTACACTAATCTTAACACTCTTTCCCTACA CGACGCTCTTCCGATCT	160
AATGATACGGCGACCACCGAGATCTACACCAGGACGTACACTCTTTCCCTACA CGACGCTCTTCCGATCT	161
AATGATACGGCGACCACCGAGATCTACACGTACTGACACACTCTTTCCCTACA CGACGCTCTTCCGATCT	162

[00393] The PCR program used to amplify the purified extension products comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) eighteen times

(6) 2 minutes at 72°C

[00394] The amplified products were stored at 4°C. 4 µL of each PCR reaction were combined into six pools, for 16 samples per pool. The amplified products were further purified using an AMPure XP bead cleanup with 64 µL of AMPure XP beads and eluting with 20 µL of a buffer comprising 10 mM Tris pH 8 and 0.05% Tween20.

[00395] The purified amplified products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics) and KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems). The purified amplified products were also diluted to 15 pM for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with either standard sequencing primers or a custom spike-in Read1 primer (SEQ ID NO: 25).

[00396] Example 6—direct PCR method for 96 multiplexed samples

[00397] In this example, a direct PCR method of the present disclosure was used to sequence identifier oligonucleotides collected from 96 multiplexed samples. In this example, 8 species of forward amplification primers and 12 species of reverse amplification primers were used. The forward primers comprised a P5 flow cell adapter (SEQ ID NO: 3), a nucleic acid sequence comprising a unique molecular identifier and a region complementary to a first amplification primer binding site present on the identifier oligonucleotide. Table 10 provides the sequences of the forward amplification primers used.

Table 10. Forward amplification primers for direct PCR method.

Primer Sequence	SEQ ID NO
AATGATACGGCGACCACCGAGATCTACACGCTCAGATATAGCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	5
AATGATACGGCGACCACCGAGATCTACACGCTCAGAATAGAGGCACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	6
AATGATACGGCGACCACCGAGATCTACACGCTCAGACCTATCCTACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	7
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGGCTCTGAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	8
AATGATACGGCGACCACCGAGATCTACACGCTCAGAAGGCGAAGACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	9
AATGATACGGCGACCACCGAGATCTACACGCTCAGATAATCTTAACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	10
AATGATACGGCGACCACCGAGATCTACACGCTCAGACAGGACGTACACTCTT TAAGACGACGTCGCTATGGCCTCTCC	11
AATGATACGGCGACCACCGAGATCTACACGCTCAGAGTACTGACACACTCTTT AAGACGACGTCGCTATGGCCTCTCC	12

[00398] The reverse primers comprised a P7 flow cell adapter (SEQ ID NO: 4), a nucleic acid sequence comprising a unique molecular identifier and a region complementary to a second amplification primer binding site present on the identifier oligonucleotide. Table 11 provides the sequences of the reverse amplification primers used.

Table 11. Reverse amplification primers for direct PCR method.

Primer Sequence	SEQ ID NO
CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	163
CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	164
CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	165
CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	166
CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	167
CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	168
CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	169
CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	170
CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	171
CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	172
CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	173
CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCTNNNNNNNNNNNNNNAACGGACAGGATGCAGCAAAAT	174

[00399] In the case of 8 forward amplification primers and 12 reverse amplification primers, when the unique molecular identifiers from a pair of forward and reverse primers are combined, a total of 96 unique combinations can be obtained, allowing for the multiplexing of 96 samples.

[00400] To amplify the collected identifier oligonucleotides for sequencing, PCR reactions with collected identifier oligonucleotides and forward and reverse amplification primers were prepared on a 96-well plate with 2 μL of each identifier oligonucleotide sample, 10 μL of NEBNext Ultra II Q5 Master Mix, 2 μL of 10 μM forward amplification primer, 2 μL of reverse amplification primer, and 4 μL of PCR-grade H_2O. Each well in the 96-well plate contained an identifier oligonucleotide sample and a unique combination of forward and reverse amplification primers. The PCR program used comprised the following steps:

(1) 30 seconds at 98°C

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) six to ten times

(6) 2 minutes at 72°C

[00401] The amplified products were stored at 4°C. 10 µL of each PCR reaction was combined into a single 15 mL conical tube.

[00402] The pooled PCR reactions were purified using diluted Agencourt AMPure XP magnetic beads (Beckman Coulter Genomics Inc.), which were prepared by combining 115.2 µL of AMPure XP beads and 1036.8 µL of AMPure XP buffer (2.5M NaCl, 20% PEG8000). Purification was performed with 1152 µL of diluted AMPure XP beads and eluted in 60 µL of a buffer comprising 10 mM Tris pH 8. The purification process was repeated with 60 µL of AMPure XP beads and eluted in 70 µL of a buffer comprising 10 mM Tris pH 8.

[00403] Following AMPure XP cleanup, PCR reactions with universal primers were prepared with 9 µL of pooled direct PCR product, 15 µL of NEBNext Ultra II Q5 Master Mix, 3 µL of 10 µM universal P7 primer (SEQ ID NO: 153) and 2 µL of 10 µM universal P5 primer (SEQ ID NO: 154). The PCR program used was:

(1) 30 seconds at 98°C,

(2) 10 seconds at 98°C

(3) 30 seconds at 65°C

(4) 30 seconds at 72°C

(5) Repeat steps (2)-(4) 15 to 24 times

(6) 2 minutes at 72°C

[00404] Two rounds of AMPure XP bead cleanup was performed. The first round was performed with 30 µL of beads and eluted with 20 µL of a buffer comprising 10 mM Tris pH 8 and the second round was performed with 20 µL beads and eluted with 11 µL of a buffer comprising 10 mM Tris pH 8.

[00405] These purified PCR products were assessed using a High Sensitivity DNA chip on 2100 Bioanalyzer (Agilent Genomics). The purified PCR products were also diluted for sequencing on a MiSeq (Illumina) according to manufacturer's instructions (MiSeq Reagent Kit v3 2×75 bp) with a custom spike-in primer (SEQ ID NO: 25).

[00406] Example 7—spatially detecting target analytes in a FFPE sample

[00407] The methods of the present invention were used to spatially detect a plurality of different target analytes, including target proteins and target RNAs, in a sample of inflamed human tonsil tissue FFPE section.

[00408] In one experiment, 30 different target proteins were spatially detected using the methods of the present disclosure in two serial sections cut from the inflamed human tonsil tissue FFPE section. The 30 target proteins are put forth in Table 12. The 30 target proteins included IgG Rabbit isotype and IgG Mouse isotype as negative controls, as these target proteins should not have been present in the inflamed human tonsil sample and therefore should not have been detected.

Table 12. Target Proteins

Target Protein	Target Protein
AKT	FOXP3
B7-H3	GZMB
Bcl-2	Histone H3
Beta-2-microglobulin	Ki67
Beta-catenin	CD20
CD14	P-AKT
CD19	PanCK
CD3	PD1
CD4	PD-L1
CD44	S6
CD45	STAT3
CD45RO	P-STAT3
CD56	VISTA
CD68	IgG Rabbit isotype (negative control)
CD8A	IgG Mouse isotype (control)

[00409] To spatially detect the 30 target proteins, 30 different probes of the present disclosure were used. Each probe comprised a target binding domain comprising an antibody that specifically binds to one of the 30 target proteins in Table 12. The two serial sections were contacted with a plurality of the 30 different probes. Ninety-six regions of interest (ROI) were then identified. For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a short probe hybridization method of the present disclosure, thereby spatially detecting the 30 target proteins in the two serial sections. As shown

in ~~Figure 24~~Figures 21A-21D, the number of reads per target protein in each ROI for the two serial sections were well correlated, demonstrating that the method yields reproducible results.

[00410] In a second experiment, 20 different target RNAs were spatially detected using the methods of the present disclosure in two different serial sections cut from the inflamed human tonsil tissue FFPE section. The 20 different target RNAs are put forth in Table 13. The 20 target RNAs included 6 negative controls(Negative Probe) that should not have been detected in the sample.

Table 13. Target RNAs

Target RNA	Target RNA	Target RNA
CD3E	CD40	CTLA4
CD3G	CD45	GAPDH
CD4	CD74	KRT13
CD20	CD79A	PD1
PSA	RP56	Negative Probe #1
Negative Probe #2	Negative Probe #3	Negative Probe #4
Negative Probe #5	Negative Probe #6	

[00411] To spatially detect the 20 Target RNAs, 20 different probes of the present disclosure were used. Each probe comprised a target binding domain comprising a nucleic acid sequence complementary to at least one portion of one of the 20 target RNAs. The two serial sections were contacted with a plurality of the 20 different probes. Ninety-six regions of interest (ROI) were then identified. For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a direct PCR method of the present disclosure, thereby spatially detecting the 20 target RNAs in the two serial sections. As shown in ~~Figure 22~~Figures 22A-22D, the number of reads per target RNA in each ROI for the two serial sections were well correlated, demonstrating that the method yields reproducible results.

[00412] Example 8—spatially detecting target proteins in a fluorescently stained FFPE sample

[00413] In another experiment, a 5 μm FFPE section of inflamed human tonsil tissue was stained with 4 fluorescent visualization markers: (1) CD3E, a T-cell marker; (2) PanCK, an epithelial cell marker; (3) Ki-67, a proliferation marker; and (4) SYTO83, a DNA stain, as shown in the left panel of Figure 23. The stained FFPE section was then contacted with the probes directed

against 30 target proteins, as described in Example 7. As shown in the left panel Figure 23, 96 regions of interest (ROIs) were selected. Each ROI was a circle with a 500 μm diameter. For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a short probe hybridization method of the present disclosure, thereby spatially detecting the 30 target proteins in the FFPE section. As shown in the right panel of Figure 23, PanCK, CD3E and Ki67 were spatially detected in ROIs that correlated with their fluorescent visualization markers. Thus, the results generated by the methods of the present disclosure correlate with the results generated using established immunohistochemical method.

[00414] Example 9—spatially detecting target RNAs in a FFPE sample

[00415] In another experiment, a 5 μm section from an inflamed human tonsil tissue FFPE block was contacted with probes directed against 20 target RNAs, as described in Example 7. 96 regions of interest (ROIs) were then selected. Each ROI was a circle with a 500 μm diameter. For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a direct PCR method of the present disclosure, thereby spatially detecting the 20 target RNAs in the two serial sections. The total RNA from a 20 μm section from the same inflamed human tonsil tissue FFPE block was then isolated. The total RNA was analyzed using the NanoString nCounter® system. Figure 24 shows that the average number of counts for 11 different RNA targets recorded using the methods of the present disclosure were well correlated with the average number of counts for the same 11 different RNA targets recorded using the nCounter® system. Thus, the results generated using the methods of the present disclosure correlate with the results generated using established direct detection methods.

[00416] Example 10—spatially detecting target RNAs in specific sub-regions of an ROI

[00417] In another experiment, a 5 μm section from an inflamed human tonsil tissue FFPE block was contacted with probes directed against 30 target proteins, as described in Example 7. The same 5 μm section was also stained with 4 fluorescent visualization markers: (1) CD3E, a T-cell marker; (2) PanCK, an epithelial cell marker; (3) Ki-67, a proliferation marker; and (4) SYTO83, a DNA stain. As shown in Figure 25, 48 regions of interest (ROIs) were identified. For each ROI, two sub-regions were then identified based on the fluorescent staining. Areas of an ROI that were fluorescently stained positive for PanCK (PanCK+) were designated a "tumor" sub-

region and the areas of an ROI that lacked PanCK fluorescent staining were designated a "micro-environment" sub-region, as shown in Figure 25. For each ROI, the tumor sub-region and the micro-environment sub-region were separately illuminated with UV light to release the identifier oligonucleotides from the probes bound within each sub-region by creating a custom mask based on the intensity of PanCK fluorescent staining. The released identifier oligonucleotides were also separately collected. The collected identifier oligonucleotides were then analyzed using the short-probe hybridization method of the present disclosure and the NanoString nCounter® system. As shown in the bottom panel of Figure 25, the results using the NanoString nCounter® system and the short-probe hybridization method of the present disclosure were well correlated. Furthermore, in the tumor sub-regions, PanCK was detected at a significantly higher level as compared to the micro-environment sub-regions. Thus, the spatial detection results provided by the methods of the present disclosure are consistent with established fluorescent immunohistochemical methods and allows for the spatial detection within highly specific regions of a sample.

[00418] Example 11—96-plex human immuno-oncology panel

[00419] A 96-plex human immuno-oncology panel was designed for use in the direct-PCR methods of the present invention. The panel comprised a plurality of probes that could be used to spatially detect 96 different human target RNAs using the direct-PCR methods of the present disclosure. The 96 target RNAs are shown in Table 14.

Table 14. Target RNAs

Target	Target	Target	Target	Target	Target	Target	Target
AKT1	CD3E	CEACAM8	FOXP3	IFNG	KRT1	NCAM1	SOD2
ARG1	CD4	CMKLR1	GZMB	IFNGR1	KRT10	NKG7	SOX10
B2M	CD40	CSF1R	H3F3A	IL10	KRT14	NT5E	STAT1
BATF3	CD40LG	CTLA4	HAVCR2	IL12B	KRT17	PDCD1	STAT2
BCL2	CD44	CTNNB1	HIF1A	IL15	KRT18	PDCD1LG2	STAT3
BCL2L1	CD47	CXCL10	HLA_DQA1	IL1B	KRT19	PECAM1	TBX21
CCL5	CD68	CXCL9	HLA_DRB	IL6	KRT6A	PMEL	TIGIT
CCND1	CD74	CSCR6	HLA_E	ITGAM	KRT7	PSMB10	TNF
CD14	CD86	DKK2	ICAM1	ITGAV	LAG3	PTEN	TNFRSF9
CD27	CD8A	EPCAM	ICOSLG	ITGAX	LY6E	PRTPRC	TNFSF4

CD274	CEACAM1	FAS	IDO1	ITGB2	MKI67	RPS6	VEGFA
CD276	CEACAM6	FASLG	IFNAR1	ITGB8	MS4A1	S100B	VSIR

[00420] In total, the panel comprised 928 different probes. Each of the probes comprised an identifier oligonucleotide comprising a first amplification primer binding site, a nucleic acid sequence comprising a unique molecular identifier, a unique nucleic acid sequence which identified the target RNA bound to the target binding domain and a second amplification primer binding site. Figure 26 shows a schematic of the probes used in the panel. For each of the 96 target RNAs, there was at least one probe within the 928 probe set comprising a target binding domain that directly or indirectly hybridized to that target RNA. For most of the 96 target RNAs, there were 10 different probes that directly or indirectly hybridized to the specific target RNA. These 10 different probes directly or indirectly hybridized to different locations on the target RNA to create a "tiling" effect, as shown in the top panel of Figure 27. Tiling multiple probes onto a target RNA means that each target RNA will be individually detected multiple times, increasing the overall accuracy of the measurement. For example, as shown in the bottom panel of Figure 27, in the case where 10 probes are tiled onto a single target RNA, one of the probes may be incorrectly detected too many times (outlier high count probe), while another probe may be incorrectly detected too few times (outlier low count probe). However, the other 8 probes may be detected at a similar level, indicating that the two outliers should be discarded during analysis and the signals from the 8 probes used to generate a more accurate measurement of the abundance of the target RNA.

[00421] The set of 928 probes also comprised 80 negative control probes. Each of the 80 negative control probes comprised a target binding domain comprising a scrambled, non-specific nucleic acid sequence that was designed using guidelines from the External RNA Controls Consortium such that the target binding domain should not be complementary to RNA molecules present within a human sample. Thus, these 80 negative control probes should not be detected during analysis.

[00422] The 96-plex human immune-oncology panel was used to analyze a tissue microarray. The tissue microarray comprised FFPE samples of 22 common human cell lines, including normal and cancerous cell types. Some of the cell lines are shown in Table 15.

[00423] Table 15. Cell lines

Cell Line	Cell Line	Cell Line	Cell Line
CCRF-CEM	DAUDI	H596	H2228
HT29	HUT78	HUH7	JURKAT
M14	MDA-MB-468	MOLT4	RAJI
SKBR3	SUDHL1	SUDHL4	

[00424] The tissue microarray also comprised one mouse cell line (3T3) as a negative control. Each of the FFPE samples on the microarray was contacted with a plurality of the 928 different probes in the imunno-onocology panel. As shown in Figure 28, for each of the FFPE samples, at least three circular regions of interest (ROIs) with a diameter of 300 μm were selected. As a negative control, ROIs were also selected on regions of the microarray that did not comprise a FFPE sample (glass negative control). For each ROI, the ROI was illuminated with UV light to release the identifier oligonucleotides from the probes bound within the ROI. The released identifier oligonucleotides were then collected and identified using a direct PCR method of the present disclosure thereby spatially detecting the 96 target RNAs in each of the FFPE samples on the tissue microarray.

[00425] Figure 29 shows that a sufficient read depth was achieved using a MiSeq v3 flowcell. The top panel of Figure 30 shows that none of the target RNAs were spatially detected for the glass negative control ROIs. Likewise, the bottom panel of Figure 30 shows that nearly none of the target RNAs were detected in the negative control mouse 3T3 FFPE sample. Conversely, as shown in Figures 31, 33 and 34, specific target RNAs were successfully detected in the HEK293 (human embryonic kidney) FFPE sample and the Jurkat (human T-cell lymphocyte) FFPE sample. Figures 31, 33 and 34 show that clusters of "tiled" probes were detected for particular target RNAs, including AKT1, B2M, CD3E, HIF1A, PTEN, RPS6, STAT1, STAT2, STAT3, VEGF, PTPRC (CD45), and KRT1/10/18/19. These results indicated that there are certain target RNAs that are differentially transcribed in the two different cells lines. The results of this experiment were also verified using the NanoString nCounter system to identify the collected identifier oligonucleotides. As shown in Figure 27, the results using the direct-PCR method of the present disclosure were consistent with the results obtained using the NanoString nCounter system.

Protein Immuno-oncology Panel (30-plex)
AKT
B7-H3
Bcl-2
Beta-2-microglobulin
Beta-Catenin
CD14
CD19
CD3
CD4
CD44
CD45
CD45RO
CD56
CD68
CD8A
FOXP3
GZMB
Histone H3
Ki67
CD20
P-AKT
PanCK
PD1
PD-L1
S6
STAT3
P-STAT3
VISTA
IgG Rabbit isotype control
IgG Mouse isotype control

FIG. 21A

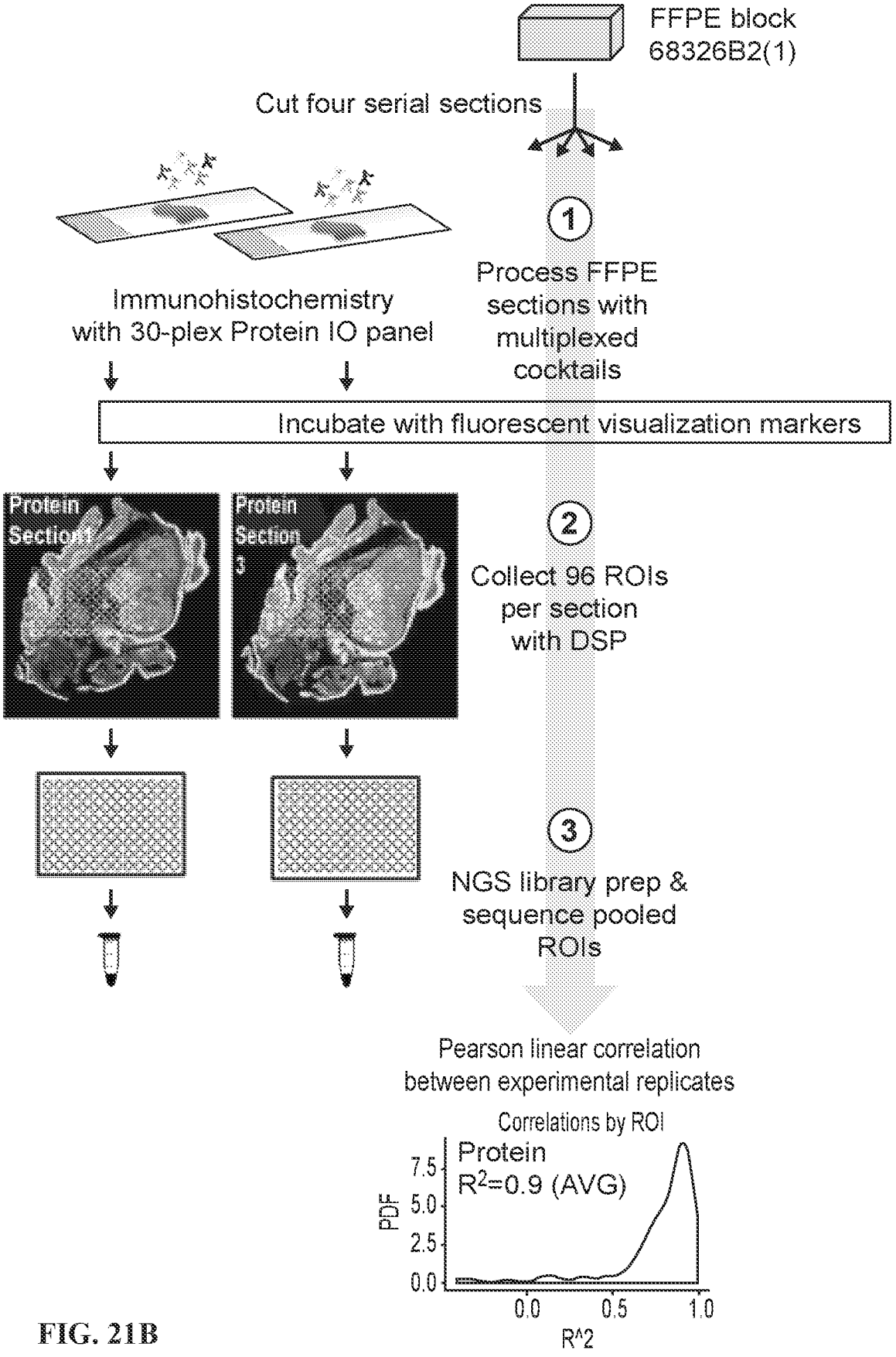

FFPE block
68326B2(1)
Cut four serial sections
1
Process FFPE sections with multiplexed cocktails
Immunohistochemistry with 30-plex Protein IO panel
Incubate with fluorescent visualization markers
Protein Section 1
Protein Section 3
2
Collect 96 ROIs per section with DSP
3
NGS library prep & sequence pooled ROIs
Pearson linear correlation between experimental replicates
Correlations by ROI
Protein
R2=0.9 (AVG)
PDF
7.5
5.0
2.5
0.0
0.0
0.5
1.0
R^2

FIG. 21B

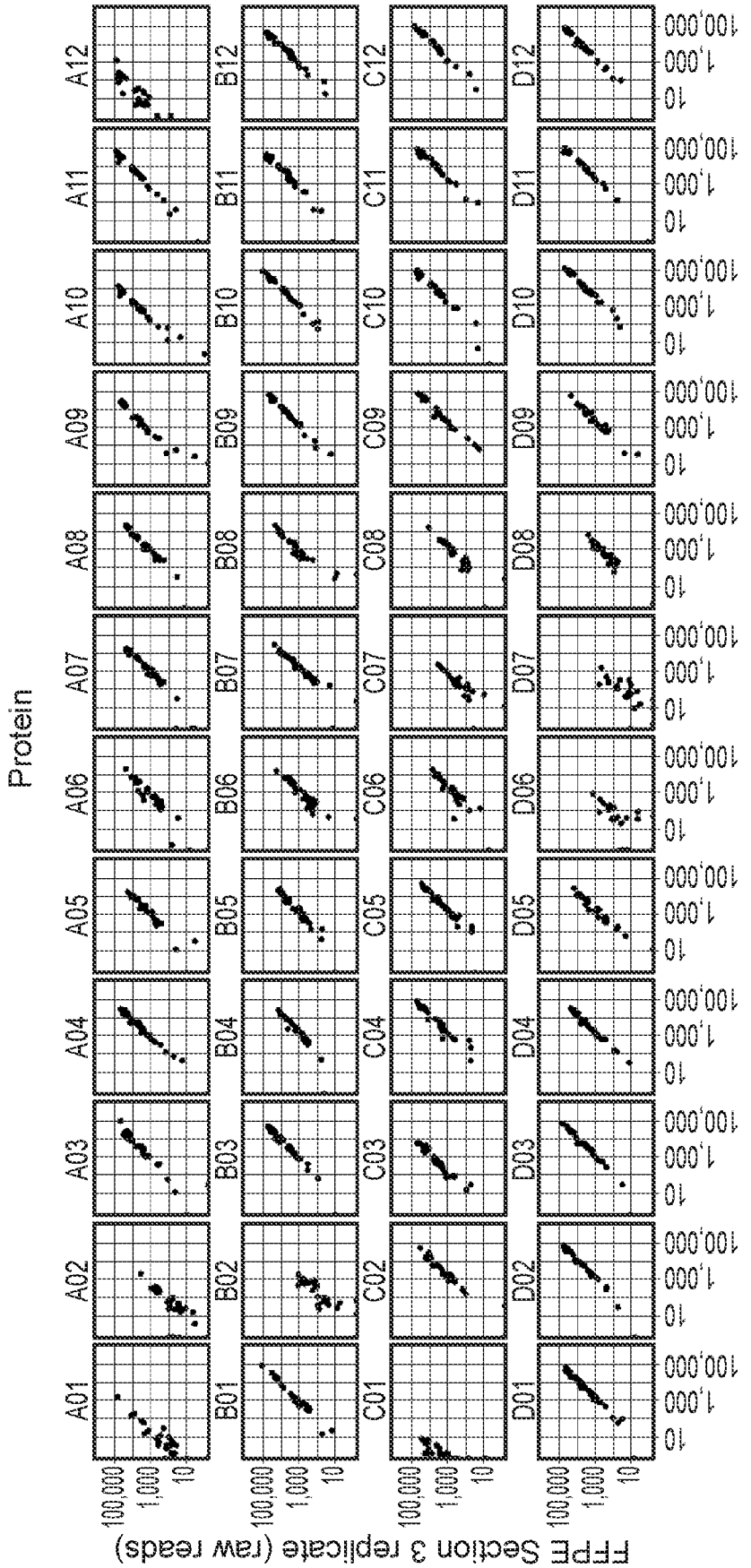

Protein
A01 A02 A03 A04 A05 A06 A07 A08 A09 A10 A11 A12
B01 B02 B03 B04 B05 B06 B07 B08 B09 B10 B11 B12
C01 C02 C03 C04 C05 C06 C07 C08 C09 C10 C11 C12
D01 D02 D03 D04 D05 D06 D07 D08 D09 D10 D11 D12
100,000
1,000
10
FFPE Section 3 replicate (raw reads)
ZZ109-2 Counts
FFPE Section 1 replicate (raw reads)

FIG. 21C

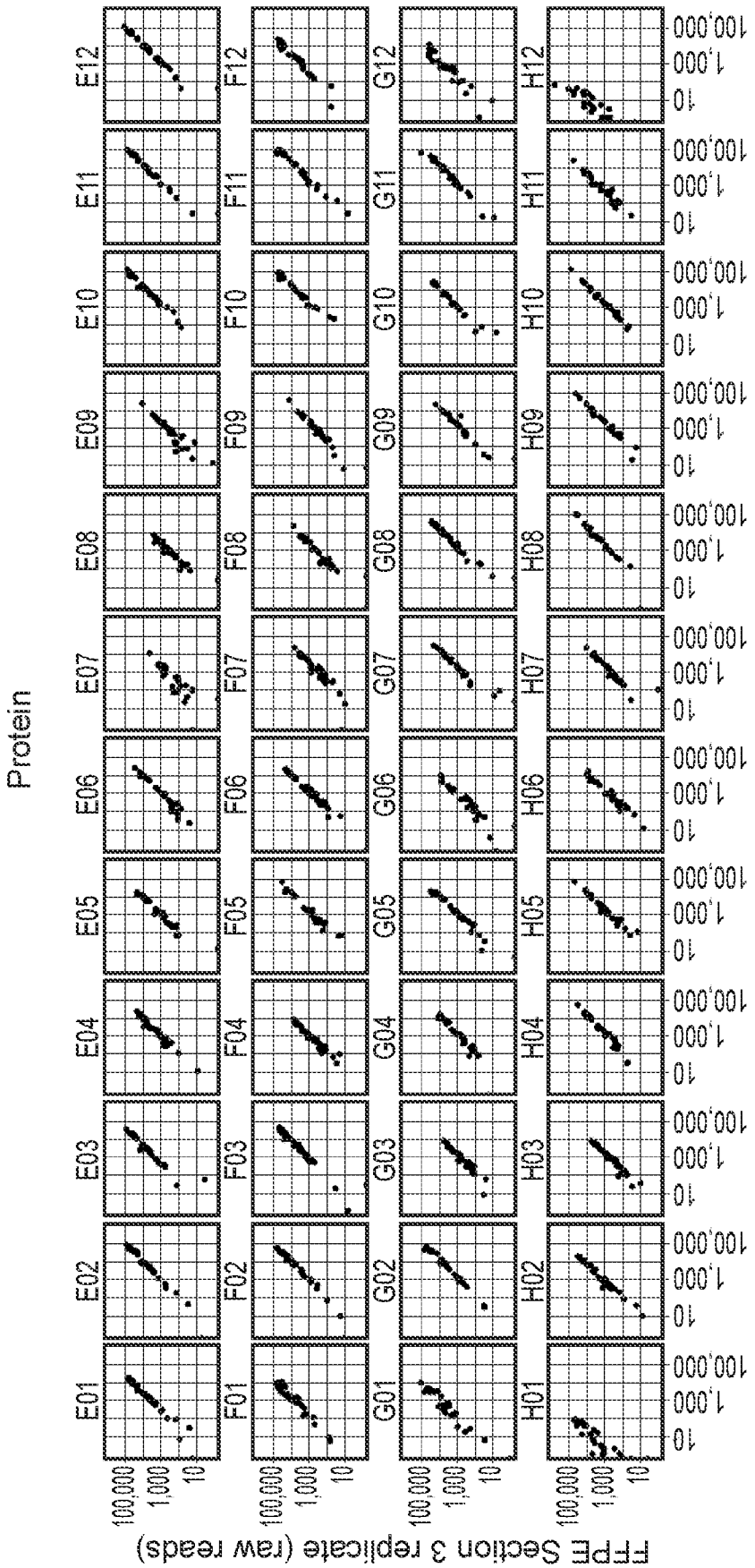

Protein
FFPE Section 3 replicate (raw reads)
FFPE Section 1 replicate (raw reads)
ZZ109-2 Counts
100,000
1,000
10
E01
E02
E03
E04
E05
E06
E07
E08
E09
E10
E11
E12
F01
F02
F03
F04
F05
F06
F07
F08
F09
F10
F11
F12
G01
G02
G03
G04
G05
G06
G07
G08
G09
G10
G11
G12
H01
H02
H03
H04
H05
H06
H07
H08
H09
H10
H11
H12

FIG. 21D

RNA Immuno-oncology Panel (20-plex)
CD3E
CD3G
CD4
CD20
CD40
CD45
CD74
CD79A
CTLA4
GAPDH
KRT13
PD1
PSA
RPS6
Negative probe 1
Negative probe 2
Negative probe 3
Negative probe 4
Negative probe 5
Negative probe 6

FIG. 22A

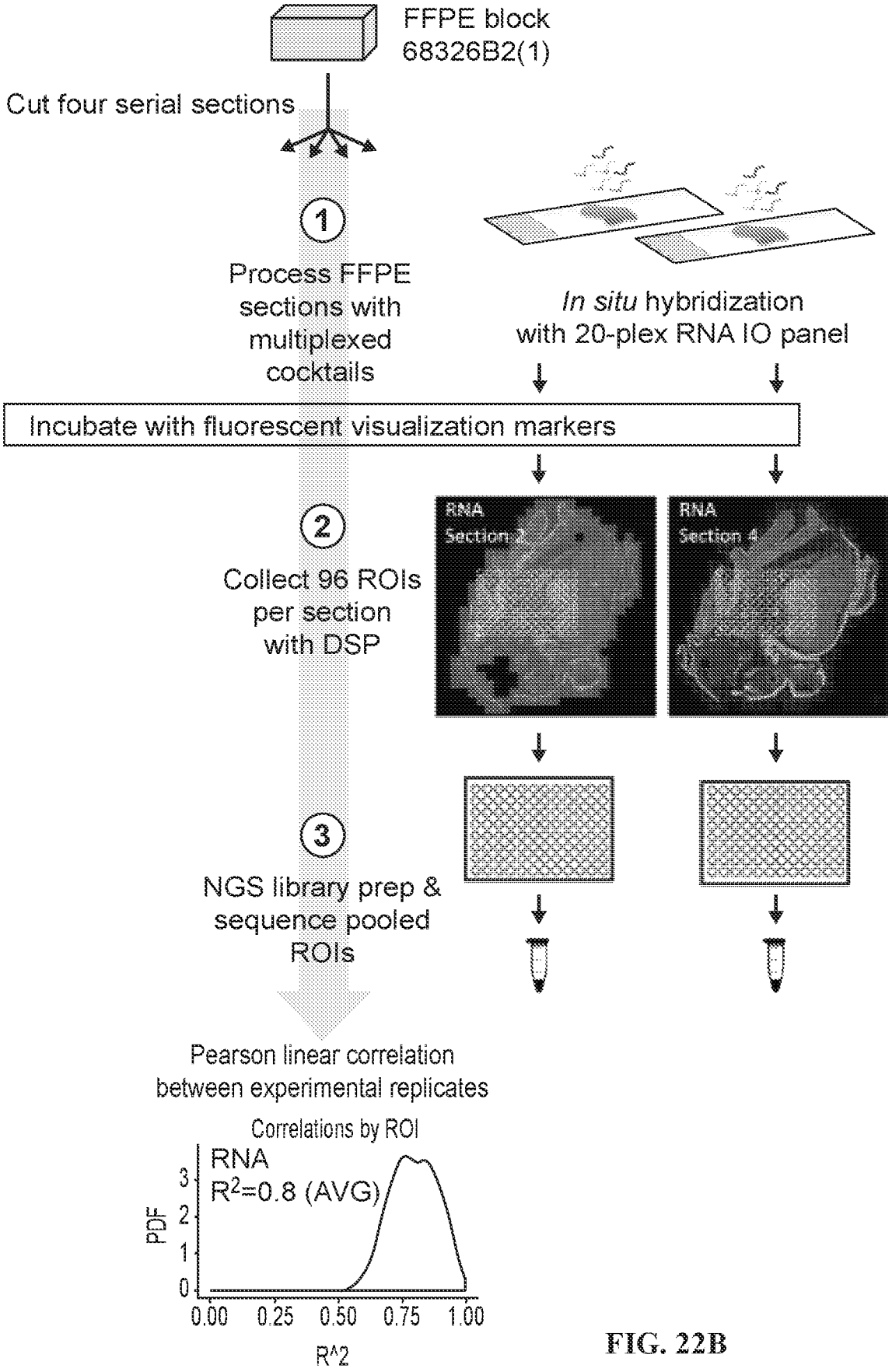

FFPE block
68326B2(1)
Cut four serial sections
1
Process FFPE sections with multiplexed cocktails
In situ hybridization with 20-plex RNA IO panel
Incubate with fluorescent visualization markers
2
Collect 96 ROIs per section with DSP
RNA
Section 3
RNA
Section 4
3
NGS library prep & sequence pooled ROIs
Pearson linear correlation between experimental replicates
Correlations by ROI
RNA
R2=0.8 (AVG)
PDF
0
1
2
3
0.00
0.25
0.50
0.75
1.00
R^2

FIG. 22B

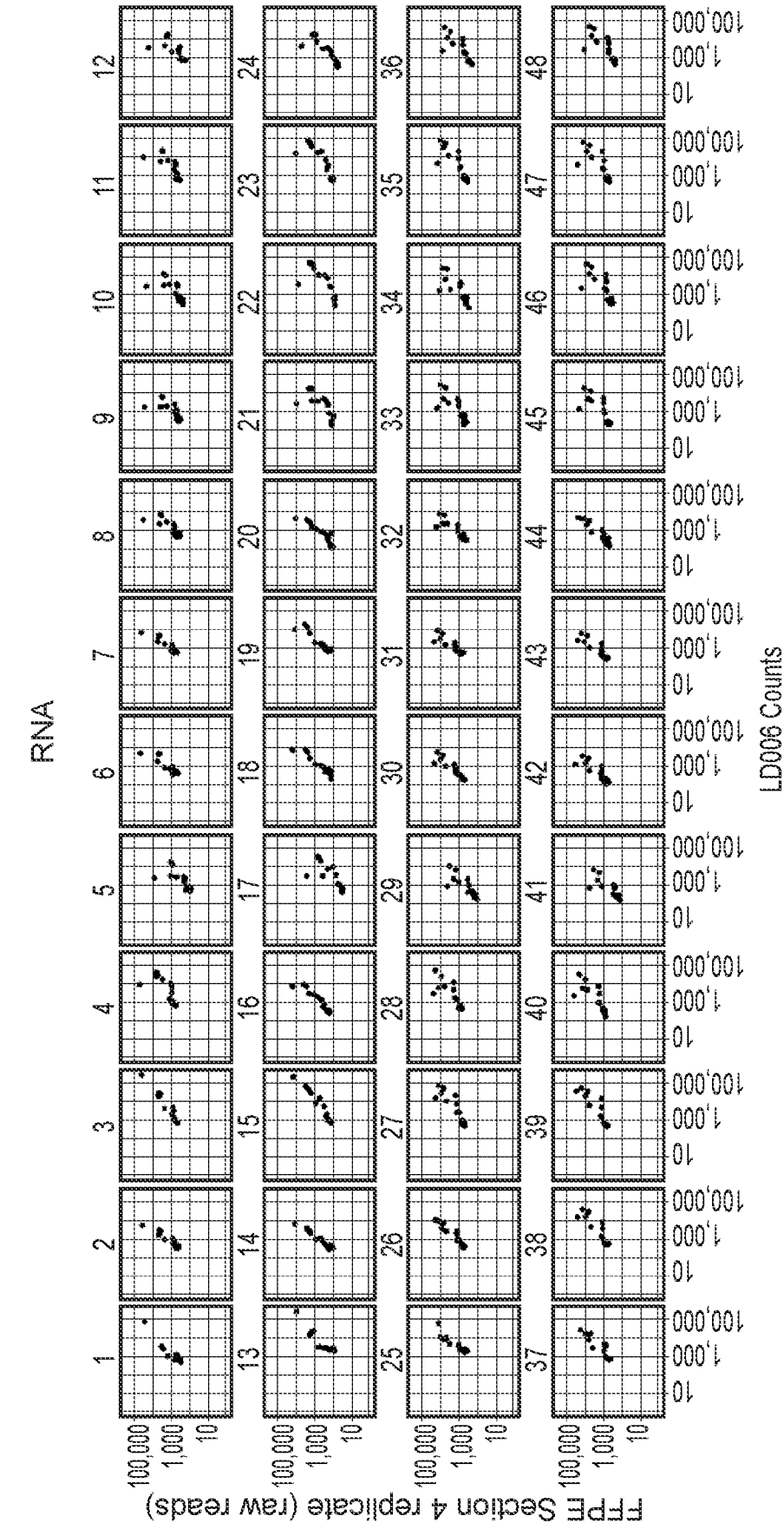

RNA
FFPE Section 4 replicate (raw reads)
LD006 Counts
FFPE Section 2 replicate (raw reads)
100,000
1,000
10
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

FIG. 22C

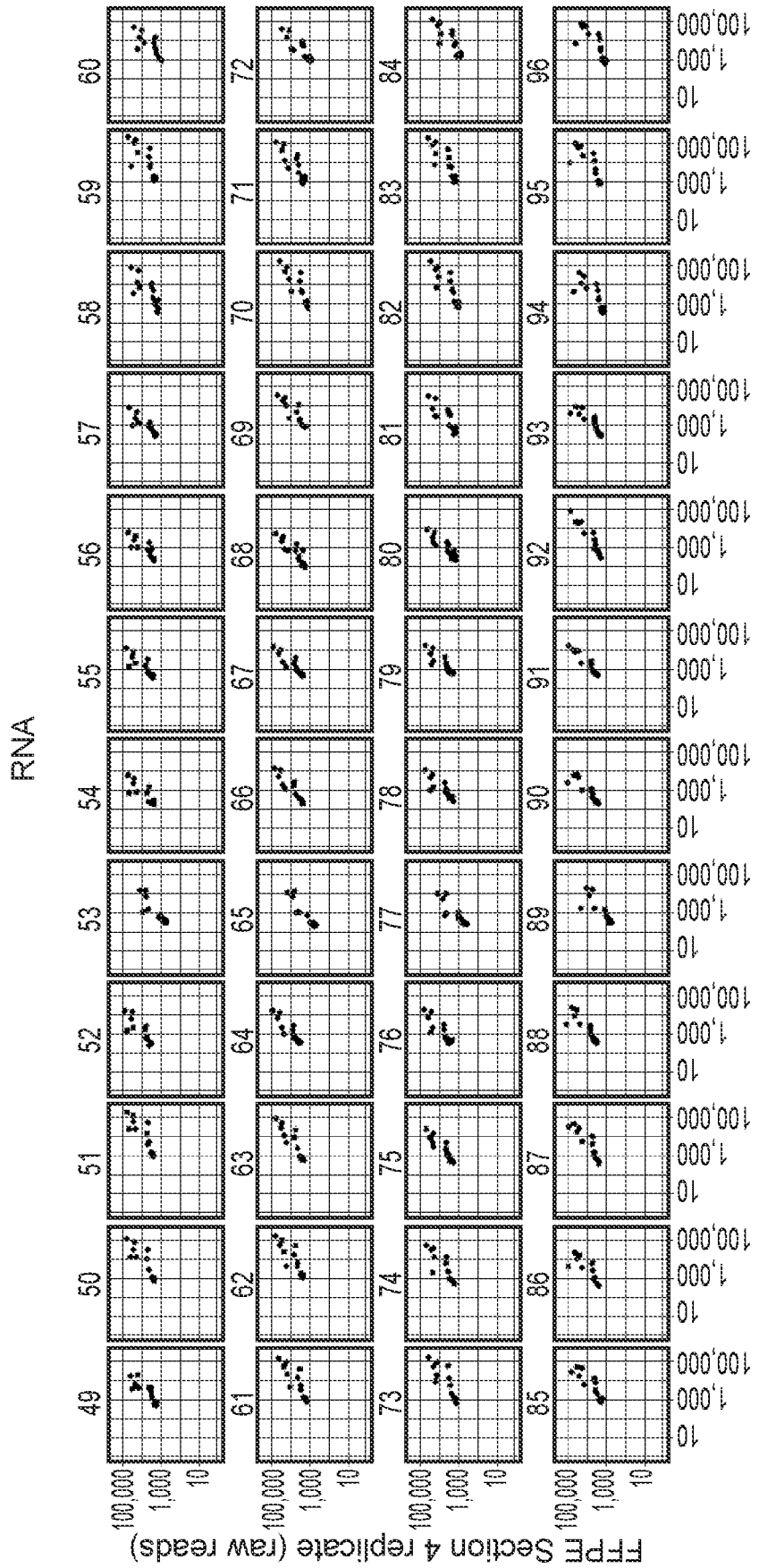

RNA
FFPE Section 4 replicate (raw reads)
FFPE Section 2 replicate (raw reads)
LD006 Counts
100,000
1,000
10
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96

FIG. 22D

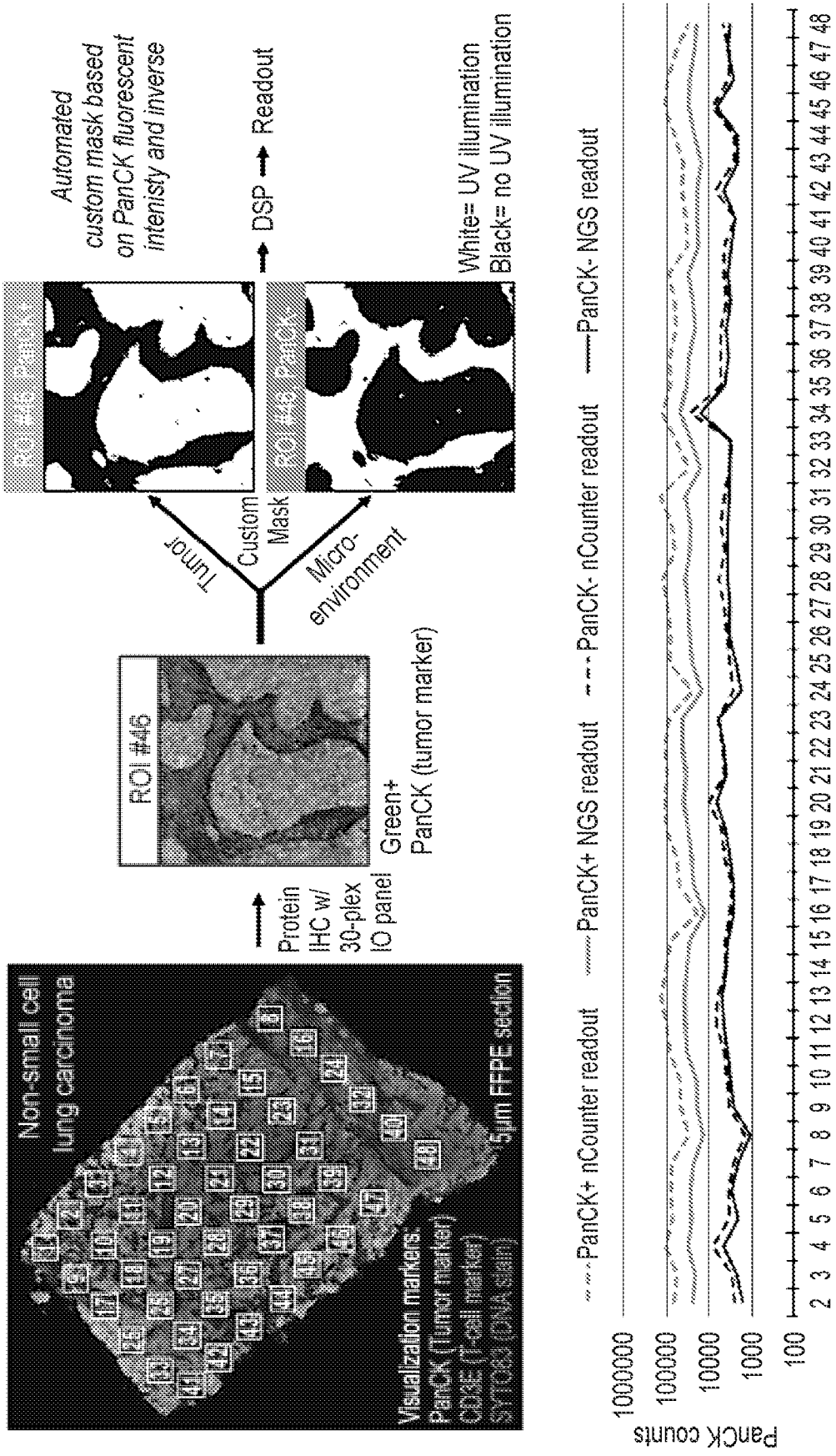

Non-small cell lung carcinoma
Visualization markers:
PanCK (Tumor marker)
CD3E (T-cell marker)
SYTO83 (DNA stain)
5μm FFPE section
Protein IHC w/ 30-plex IO panel
ROI #46
Green+ PanCK (tumor marker)
Tumor
Micro-environment
Custom Mask
ROI #46 Panck+
ROI #46 Panck-
Automated custom mask based on PanCK fluorescent intenisty and inverse
DSP → Readout
White= UV illumination
Black= no UV illumination
PanCK+ nCounter readout
PanCK+ NGS readout
PanCK- nCounter readout
PanCK- NGS readout
PanCK counts
1000000
100000
10000
1000
100
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
ROI

FIG. 25

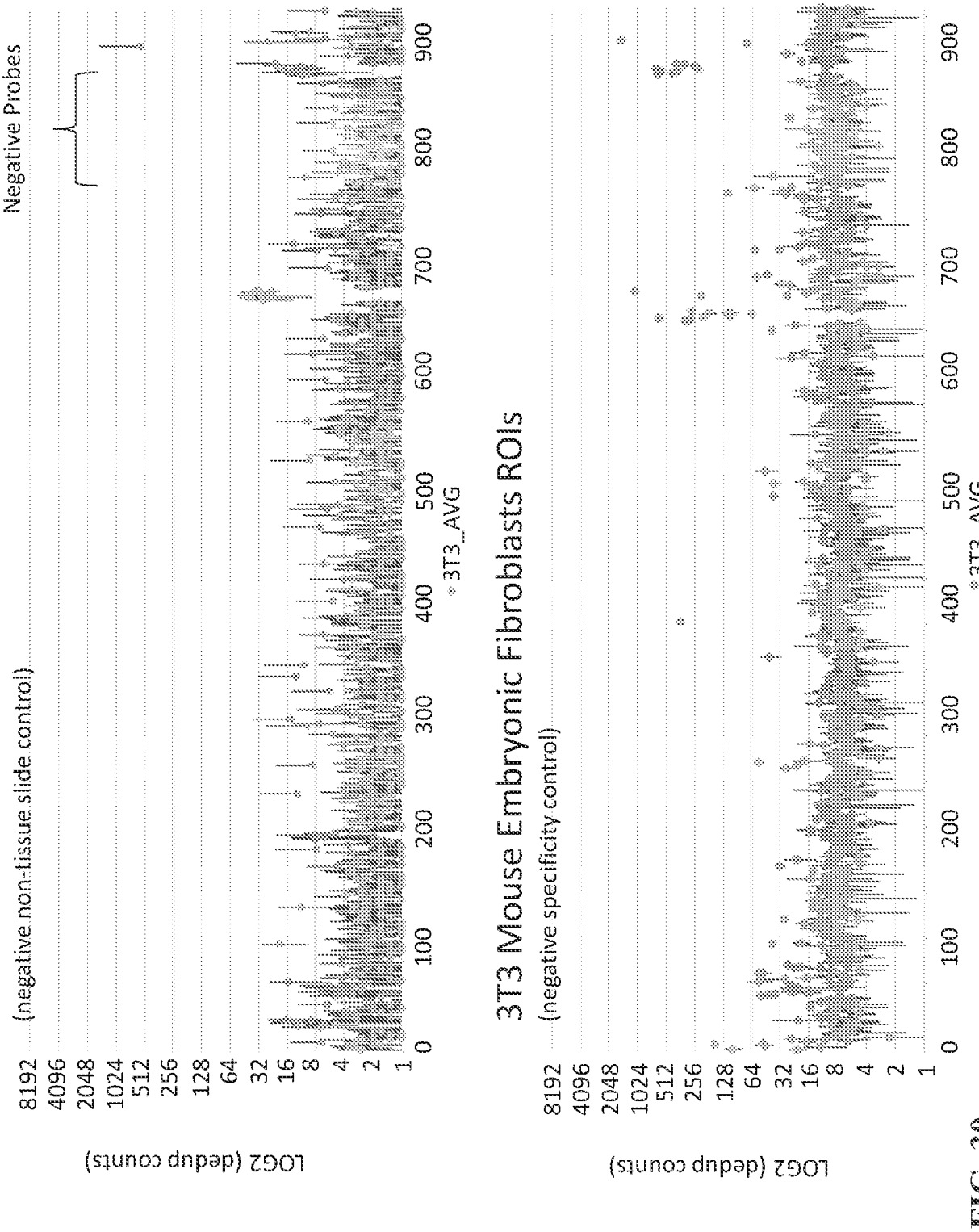

Glass ROIs
(negative non-tissue slide control)
80 ERCC
Negative Probes
LOG2 (dedup counts)
8192
4096
2048
1024
512
256
128
64
32
16
8
4
2
1
0
100
200
300
400
500
600
700
800
900
3T3_AVG
3T3 Mouse Embryonic Fibroblasts ROIs
(negative specificity control)
LOG2 (dedup counts)
3T3_AVG

FIG. 30

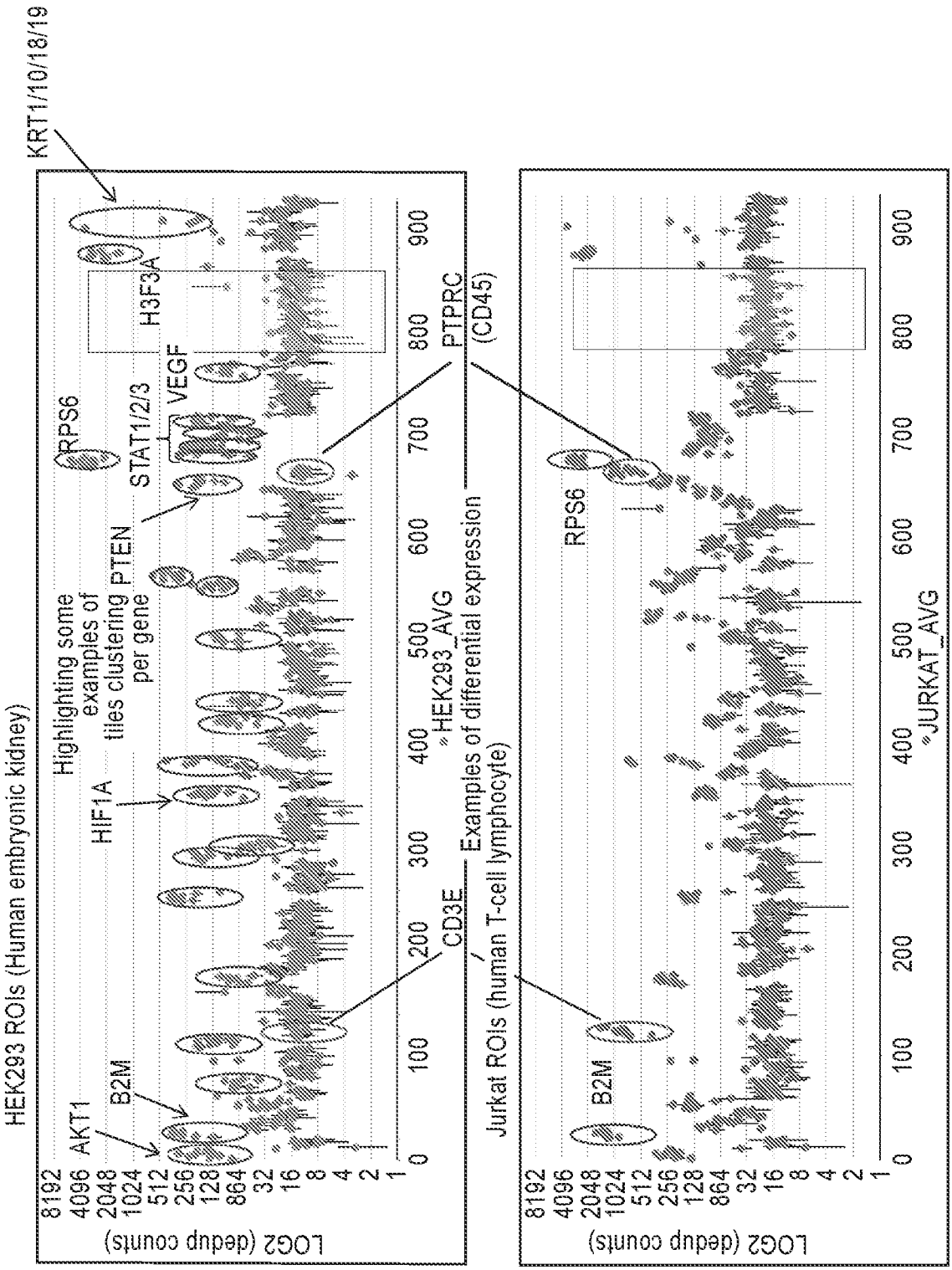

HEK293 ROIs (Human embryonic kidney)
KRT1/10/18/19
Highlighting some examples of tiles clustering per gene
AKT1
B2M
HIF1A
PTEN
RPS6
STAT1/2/3
VEGF
H3F3A
CD3E
PTPRC (CD45)
LOG2 (dedup counts)
8192
4096
2048
1024
512
256
128
64
32
16
8
4
2
1
0
100
200
300
400
500
600
700
800
900
HEK293_AVG
Examples of differential expression
Jurkat ROIs (human T-cell lymphocyte)
JURKAT_AVG

FIG. 31

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	F. W. Lu

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT IN SUPPORT OF COMPUTER READABLE FORM SUBMISSION UNDER 37 C.F.R. § 1.821

I hereby state that the contents of the computer readable form of the Sequence Listing, submitted in the above-identified application in accordance with 37 C.F.R. § 1.821, does not include any new matter that goes beyond the disclosure of the application as filed. The Sequence Listing is supported by the specification and references incorporated therein. Therefore, no new matter is added.

Applicant believes no fee is due with this response. However, if a fee is due, please charge our Deposit Account No. 50-1283, under Order No. NATE-037/001US 321329-2523 from which the undersigned is authorized to draw.

Dated: July 26, 2021

Respectfully submitted,
COOLEY LLP

USPTO CUSTOMER NO. 141429
COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (202) 842-7899

By: /Robert E. Powers/
Robert E. Powers, Ph.D.
Reg. No. 77,802
Matthew Pavao, Ph.D.
Reg. No. 50,572

253971878
253971878 v1

Electronic Patent Application Fee Transmittal

Application Number:	16272487
Filing Date:	11-Feb-2019
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Filer:	Robert E. Powers/Francesca Vandermaelen
Attorney Docket Number:	NATE-037001US 321329-2523

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 3 months with $0 paid	1253	1	1480	1480
Miscellaneous:				
			Total in USD ($)	**1480**

Electronic Acknowledgement Receipt

EFS ID:	43342242
Application Number:	16272487
International Application Number:	
Confirmation Number:	6864
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Customer Number:	141429
Filer:	Robert E. Powers/Francesca Vandermaelen
Filer Authorized By:	Robert E. Powers
Attorney Docket Number:	NATE-037001US 321329-2523
Receipt Date:	26-JUL-2021
Filing Date:	11-FEB-2019
Time Stamp:	14:52:18
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$1480
RAM confirmation Number	E20217PE52364473
Deposit Account	
Authorized User	

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		NATE-037_001US_Response.pdf	190095 2589eda1f78f26bebbeb20ec8a7f35702fdd04f9	yes	15

Multipart Description/PDF files in .zip description

Document Description	Start	End
Amendment/Req. Reconsideration-After Non-Final Reject	1	1
Specification	2	2
Drawings-only black and white line drawings	3	3
Claims	4	7
Applicant Arguments/Remarks Made in an Amendment	8	15

Warnings:					
Information:					
2	Specification	NATE-037_001US_Sub_Spec_Statement.pdf	93942 f6527a1d6e4246c6d2cc0ca7f2c76b295e85da05	no	1
Warnings:					
Information:					
3	Specification	NATE-037_001US_Sub_Spec_Clean.pdf	458016 27d4a92ed4070e1a7f6edbfc6e60ba70a27ac63b	no	102
Warnings:					
Information:					
4	Specification	NATE-037_001US_Sub_Spec_Marked-Up.pdf	460053 d88c3118323ec961a0accba69fc53b7d3a13d445	no	102
Warnings:					
Information:					

5	Drawings-only black and white line drawings	NATE-037_001US_Replacement_Drawings.pdf	2321949 5a86324a6dc0072fe7c27dcd0ce6398053a2e028	no	11
Warnings:					
Information:					
6	Sequence Listing	NATE-037_001US_SL_Statement.pdf	96878 e99a2c202381b382de54e9138b58ce9ead9511cc	no	1
Warnings:					
Information:					
7	Sequence Listing (Text File)	NATE-037_001US_Sequence_Listing.txt	50049	no	-
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	38802 d487a7c8c497b7d2dc34e068762726fa1af40091	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	3709784		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

==========
Sequence Listing was accepted.
If you need help call the Patent Electronic Business Center at (866) 217-9197 (toll free).
Reviewer: Anjum, Durreshwar
Timestamp: [year=2021; month=7; day=26; hr=14; min=54; sec=48; ms=274;]
==========

Validated By CRFValidator v 1.0.5

Application No: 16272487 Version No: 2.0

Input Set:

Output Set:

Started: 2021-07-26 14:52:45.574
Finished: 2021-07-26 14:52:47.249
Elapsed: 0 hr(s) 0 min(s) 1 sec(s) 675 ms
Total Warnings: 807
Total Errors: 0
No. of SeqIDs Defined: 175
Actual SeqID Count: 175

Error code	Error Description
W 213	Artificial or Unknown found in <213> in SEQ ID (1)
W 213	Artificial or Unknown found in <213> in SEQ ID (2)
W 213	Artificial or Unknown found in <213> in SEQ ID (3)
W 213	Artificial or Unknown found in <213> in SEQ ID (4)
W 213	Artificial or Unknown found in <213> in SEQ ID (5)
W 213	Artificial or Unknown found in <213> in SEQ ID (6)
W 213	Artificial or Unknown found in <213> in SEQ ID (7)
W 213	Artificial or Unknown found in <213> in SEQ ID (8)
W 213	Artificial or Unknown found in <213> in SEQ ID (9)
W 213	Artificial or Unknown found in <213> in SEQ ID (10)
W 213	Artificial or Unknown found in <213> in SEQ ID (11)
W 213	Artificial or Unknown found in <213> in SEQ ID (12)
W 213	Artificial or Unknown found in <213> in SEQ ID (13)
W 213	Artificial or Unknown found in <213> in SEQ ID (14)
W 213	Artificial or Unknown found in <213> in SEQ ID (15)
W 213	Artificial or Unknown found in <213> in SEQ ID (16)
W 213	Artificial or Unknown found in <213> in SEQ ID (17)
W 213	Artificial or Unknown found in <213> in SEQ ID (18)
W 213	Artificial or Unknown found in <213> in SEQ ID (19)
W 213	Artificial or Unknown found in <213> in SEQ ID (20)

Input Set:

Output Set:

Started: 2021-07-26 14:52:45.574
Finished: 2021-07-26 14:52:47.249
Elapsed: 0 hr(s) 0 min(s) 1 sec(s) 675 ms
Total Warnings: 807
Total Errors: 0
No. of SeqIDs Defined: 175
Actual SeqID Count: 175

Error code	Error Description
	This error has occured more than 20 times, will not be displayed
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(27) on line number 396
W 447	n or Xaa used, for: SEQID(28) on line number 416
W 447	n or Xaa used, for: SEQID(28) on line number 416
W 447	n or Xaa used, for: SEQID(28) on line number 416
W 447	n or Xaa used, for: SEQID(28) on line number 416
W 447	n or Xaa used, for: SEQID(28) on line number 416
	This error has occured more than 20 times, will not be displayed

SEQUENCE LISTING

<110> NanoString Technologies, Inc.
Beechem, Joseph
Kim, Dae
Hoang, Margaret
Gregory, Mark
Piazza, Erin
Zhou, Denise

<120> BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

<130> NATE-037/001WO

<140> US 16/272,487
<141> 2019-02-11

<150> US 62/629,180
<151> 2018-02-12

<150> US 62/771,212
<151> 2018-11-26

<160> 175

<170> PatentIn version 3.5

<210> 1
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> i7 sequence

<400> 1
agatcggaag agcacacgtc tgaactccag tcac 34

<210> 2
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> i5 sequence

<400> 2
acactctttc cctacacgac gctcttccga 30

<210> 3
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> P5 flow cell adapter

<400> 3
aatgatacgg cgaccaccga gatctacac 29

<210> 4
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> P7 flow cell adapter

<400> 4
atctcgtatg ccgtcttctg cttg 24

<210> 5
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> DSP_NGS_i501 Primer

<400> 5
aatgatacgg cgaccaccga gatctacacg ctcagatata gcctacactc tttaagacga 60

cgtcgctatg gcctctcc 78

<210> 6
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> DSP_NGS_i502 Primer

<400> 6
aatgatacgg cgaccaccga gatctacacg ctcagaatag aggcacactc tttaagacga 60

cgtcgctatg gcctctcc 78

<210> 7
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> DSP_NGS_i503 Primer

<400> 7
aatgatacgg cgaccaccga gatctacacg ctcagaccta tcctacactc tttaagacga 60

cgtcgctatg gcctctcc 78

<210> 8
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> DSP_NGS_i504 Primer

<400> 8
aatgatacgg cgaccaccga gatctacacg ctcagaggct ctgaacactc tttaagacga 60

cgtcgctatg gcctctcc 78

<210> 9
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> DSP_NGS_i505 Primer

<400> 9
aatgatacgg cgaccaccga gatctacacg ctcagaaggc gaagacactc tttaagacga 60

cgtcgctatg gcctctcc 78

<210> 10
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> DSP_NGS_i506 Primer

<400> 10
aatgatacgg cgaccaccga gatctacacg ctcagataat cttaacactc tttaagacga 60

cgtcgctatg gcctctcc 78

<210> 11
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> DSP_NGS_i507 Primer

<400> 11
aatgatacgg cgaccaccga gatctacacg ctcagacagg acgtacactc tttaagacga 60

cgtcgctatg gcctctcc 78

<210> 12
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> DSP_NGS_i508 Primer

<400> 12
aatgatacgg cgaccaccga gatctacacg ctcagagtac tgacacactc tttaagacga 60

cgtcgctatg gcctctcc 78

<210> 13
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i701 Primer

<400> 13
caagcagaag acggcatacg agatcgagta atgtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 14
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i702 Primer

<400> 14
caagcagaag acggcatacg agattctccg gagtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 15
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i703 Primer

<400> 15
caagcagaag acggcatacg agataatgag cggtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 16
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i704 Primer

<400> 16
caagcagaag acggcatacg agatggaatc tcgtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 17
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i705 Primer

<400> 17
caagcagaag acggcatacg agatttctga atgtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 18
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i706 Primer

<400> 18
caagcagaag acggcatacg agatacgaat tcgtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 19
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i707 Primer

<400> 19
caagcagaag acggcatacg agatagcttc aggtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 20
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i708 Primer

<400> 20
caagcagaag acggcatacg agatgcgcat tagtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 21
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i709 Primer

<400> 21
caagcagaag acggcatacg agatcatagc cggtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 22
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i710 Primer

<400> 22
caagcagaag acggcatacg agatttcgcg gagtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 23
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i711 Primer

<400> 23
caagcagaag acggcatacg agatgcgcga gagtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 24
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> NEBNext i712 Primer

<400> 24
caagcagaag acggcatacg agatctatcg ctgtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 25
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Spike-in primer

<400> 25
acactcttta agacgacgtc gctatggcct ctcc 34

<210> 26
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Amplification primer binding site

<400> 26
gtgactggag ttcagacgtg tgctcttccg atct 34

<210> 27
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P01_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 27
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt tgaagcacac 60

cgtttttctt tcttctttca cgg 83

<210> 28
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P03_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 28
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnna cccacaggtt 60

atacgggatt atccggttat cca 83

<210> 29
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P04_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 29
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc gacaccgagt 60

tcgaccgtta tgttggtagg atc 83

<210> 30
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P05_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 30
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc ggtgtgtaag 60

cgtaacgatg ttggtgtcgc tct 83

<210> 31
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P06_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 31
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc agacactgcg 60

acaactcacg atcatgacac aga 83

<210> 32
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P07_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 32
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnna tattctgtac 60

tcagtgccta tccacctaat agg 83

<210> 33
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P09_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 33
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt tcagttataa 60

tgtgtccagc agaagcagga att 83

<210> 34
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P11_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 34
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnng tcctttgttg 60

ggcggaccgt aatgaggaat ttg 83

<210> 35
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P13_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 35
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnng atgagacttc 60

tacatgtccg atgtttttgt gct 83

<210> 36
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P15_PriExtUMI15_tpt

<220>

<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 36
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnna ctcacacata 60

gtactgacac gtaagatagg atg 83

<210> 37
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P16_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 37
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt taccctatct 60

cgtctatgta cgtcaggctg aat 83

<210> 38
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P17_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 38
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnna tcaacgtagg 60

gtaaggtcat atttttacct tac 83

<210> 39
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P18_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 39
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt tccctctttc 60

tccgcttatg gatgaaagga cag 83

<210> 40
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P19_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 40
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc ctgcacagtg 60

agtttctttc actctaactc tct 83

<210> 41
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P20_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 41
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt gtcgctctag 60

tgtgactttt ccacctcgca tct 83

<210> 42
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P21_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 42
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnna tatctttctc 60

gggtaaagat taggcgtccg ata 83

<210> 43
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P23_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 43
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc gattagccgt 60

agacgcaact cattgccgaa gat 83

<210> 44
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P24_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 44
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt gtgagcattt 60

cagtacgagt gatgcagata aac 83

<210> 45

<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P25_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 45
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt atagttacca 60

agtactatgg gttggtggaa gcc 83

<210> 46
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P26_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 46
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc caattatact 60

gtctgttatg ttctcggata agc 83

<210> 47
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P27_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 47
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt caggtgcttg 60

taggctcatg ataggggtaa tgc 83

<210> 48
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P28_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 48
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc tctgctgtaa 60

tctcagctcc acttgtttct aag 83

<210> 49
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P29_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 49
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnng tgcatattgc 60

agctgagcca gctcaatttg aag 83

<210> 50
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P30_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 50

gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc cgttgattta 60

cgcaacagcg gcttatatag ctc 83

<210> 51
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P31_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 51
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc atcatcgaca 60

gttcgcagcc ctataacatg ata 83

<210> 52
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P32_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 52
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnna tcgcaggatg 60

gtacagcatc atacatgatg agc 83

<210> 53
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P33_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)

<223> n is a, c, g, or t

<400> 53
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnna tcgcaggatg 60

gtacagcatc atacatgatg agc 83

<210> 54
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P34_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 54
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnna tggcggtttc 60

gggtcctgca ctattcctaa taa 83

<210> 55
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P35_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 55
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnc cagtacgggt 60

actaataagt gtcatatcta ttg 83

<210> 56
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> P36_PriExtUMI15_tpt

<220>
<221> misc_feature
<222> (35)..(49)
<223> n is a, c, g, or t

<400> 56
gtgactggag ttcagacgtg tgctcttccg atctnnnnnn nnnnnnnnnt gttggagagg 60

ttagaggtga ggaggcgaag ata 83

<210> 57
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> Reverse primer

<400> 57
caagcagaag acggcatacg agatgtcggt aagtgactgg agttcagacg tgtgctcttc 60

cgatct 66

<210> 58
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> Reverse primer

<400> 58
caagcagaag acggcatacg agataggtca ctgtgactgg agttcagacg tgtgctcttc

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1.	7,255,999	08-14-2007	Singh et al.	
	2.	2003/0148335	08-07-2003	Shen et al.	
	3.	2005/0131006	06-16-2005	Mukherjee et al.	
	4.	2005/0170439	08-04-2005	Chan-Hui et al.	
	5.	2005/0196786	09-08-2005	Levy	
	6.	2006/0063196	03-23-2006	Akeson et al.	
	7.	2008/0118934	05-22-2008	Gerdes et al.	
	8.	2010/0075858	03-25-2010	Davis et al.	
	9.	2011/0172115	07-14-2011	Thompson	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (*if known*)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	10.	WO 2009/156725 A1	12-30-2009	Trillion Genomics Limited		
	11.	WO 2010/081114 A2	07-15-2010	20/20 Genesystems, Inc.		
	12.	WO 2012/140224 A1	10-18-2012	Frisen, J.; Stahl, P.; Lundeberg, J.		
	13.	WO 2014/060483 A1	04-24-2014	Spatial Transcriptomics AB		
	14.	WO 2016/162309 A1	10-13-2016	Spatial Transcriptomics AB		
	15.	WO 2018/091676 A1	05-24-2018	Spatial Transcriptomics AB		

Examiner Signature:		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

254069444 v1

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	NATE-037/001US 321329-2523

NON PATENT LITERATURE DOCUMENTS

Examiner's Initials	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title ofthe item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[2]
	16.	ARMANI, M. et al., "2D-PCR: a method of mapping DNA in tissue sections," Lab Chip., 9(24):3526-3534 (2009); doi:10.1039/b910807f.	
	17.	BAILEY, R. C. et al., "DNA-Encoded Antibody Libraries: A Unified Platform for Multiplexed Cell Sorting and Detection of Genes and Proteins," J. Am. Chem. Soc., 129:1959-1967 (2007).	
	18.	CESANO, A. et al., "Abstract 1371: Spatially-resolved, multiplexed digital characterizatio nof protein distribution and abundance in FFPE tissue sections," AACR 107th Annual Meeting, April 16-20, 2016, 5 pages.	
	19.	LEE, H. et al., "Colour-barcoded magnetic microparticles for multiplexed bioassays," Nature Materials, 9:745-749 (2010).	
	20.	LEMAIRE, R. et al., "Tag-Mass: Specific Molecular Imaging of Transcriptome and Proteome by Mass Spectrometry Based on Photocleavable Tag," J Proteome Res., 6(6):2057-2067 (2007); oi:10.1021/pr0700044.	

Examiner Signature:		Date Considered	

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

254069444 v1

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 December 2009 (30.12.2009)

PCT

(10) International Publication Number
WO 2009/156725 A1

(51) International Patent Classification:
G01N 33/68 (2006.01) *C12Q 1/68* (2006.01)

(21) International Application Number:
PCT/GB2009/001583

(22) International Filing Date:
24 June 2009 (24.06.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0811574.3 24 June 2008 (24.06.2008) GB
0821945.3 1 December 2008 (01.12.2008) GB

(71) Applicant *(for all designated States except US)*: **TRILLION GENOMICS LIMITED** [GB/GB]; 30 Canterbury Street, Cambridge CB4 3QF (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **THOMPSON, Andrew, Hugin** [GB/GB]; 30 Canterbury Street, Cambridge CB4 3QF (GB).

(74) Agent: **ROBERTS, Michael, Austin**; Reddie & Grose, 16 Theobalds Road, London WC1X 8PL (GB).

(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, 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, KM, 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, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

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

WO 2009/156725 A1

(54) Title: CHARACTERISING PLANAR SAMPLES BY MASS SPECTROMETRY

(57) Abstract: The present invention is directed to probes and methods for the imaging of specimens such as tissue samples or other biological specimens or arrays using mass spectrometry (MS). In one aspect, there is provided a method for analysing a specimen comprising the steps of contacting the specimen with probe molecules each of which includes one or more mass tags coupled to each probe molecule via a first cleavable linker, contacting the specimen with a Matrix-assisted Laser Desorption Ionisation (MALDI) matrix or similar material, irradiating a portion of the specimen with a laser beam to release ions from the specimen, selecting released ions with mass-to-charge ratios corresponding to the mass tags or derivatives of the mass tags, and recording the amount and type of ions selected together with the location of the portion of the specimen. The irradiation, selection and recording steps are then repeated on a different portion of the specimen.

Characterising Planar Samples by Mass Spectrometry

The present invention relates to probes and methods for the imaging of specimens using mass spectrometry (MS). The probes and methods are particularly suited to imaging of tissue samples or other biological specimens or arrays.

Imaging of gene and protein expression directly in tissues is a highly informative way to determine gene function and activity. Various techniques have been developed for this purpose but most current techniques are limited by the number of tags (also known as "markers") available, restricting the number of species that can be analysed simultaneously.

Autoradiography of "in situ" hybridised probes is a well-established technique for imaging of expression patterns in tissue sections (see Wilson et al., 1997, Brain Res Brain Res Protoc. **1**(2):175-85; Steel et al., 1998, Eur J Histochem. **42**(2):143-50) but typically only one radiolabel can be used at a time and there are safety issues involved with the use of radiolabelled probes. Chemiluminescence and enzyme-linked probes offer safety enhancements compared to radiolabelled techniques but again the number of probes that can be resolved simultaneously is severely limited (Viale et al., 1992, Liver. **12**(4 Pt 2): 243-51).

Imaging mass spectrometry (IMS) is an emerging technique in mass spectrometry with many applications, including the analysis of tissue samples. IMS offers the ability to visualise the spatial distribution of drugs in their correct physiological setting without requiring radio-labelled drugs while also being able to distinguish drug metabolites simultaneously. Drug cocktails can also be analysed by IMS, as each compound can be individually resolved.

IMS also has the potential to simultaneously analyse nucleic acids, peptides and proteins along with endogenous metabolites, opening the possibility of relating drug action to changes in the molecular profiles of tissue components such as mRNA and allowing meaningful mode of action studies to be performed and off-target activities to be determined. Understanding of developmental biology will also be enhanced by the ability to image nucleic acid and polypeptide activities in a spatially resolved manner.

However, identification of large molecular species, such as expressed mRNAs or polypeptides, by mass alone is not trivial as sensitivity of mass spectrometry to large molecules, for example large biological polymers, is currently limited.

A recent publication by the present inventor using electrospray based analysis of mass tags suggests that MS/MS based tag designs can be used to overcome background noise and give sensitive and quantitative measurements of biomolecules linked to the mass tags (Thompson et al., 2007, Nucleic Acids Res. **35**(4):e28).

However, the prior art does not disclose techniques that effectively employ the capability of MS such as Matrix-assisted Laser Desorption Ionisation (MALDI) Tandem Mass Spectrometry (together termed "MALDI MS/MS") to analyse and effectively quantify nucleic acids or polypeptides of interest as a function of their position in a test sample.

In addition, the prior art does not provide probe molecules suitable for imaging tissue samples using MS such as MALDI MS/MS.

Moreover, the prior art does not provide methods to speed up analysis of tissue samples, since it is currently very time-consuming to acquire a large number of mass spectra that are needed to produce an image from a tissue section using a mass spectrometer.

According to the present invention there is provided in one aspect a method for analysing a specimen, comprising the steps of:

a) contacting the specimen with probe molecules, each of the probe molecules including one or more mass tags coupled to each probe molecule via a first cleavable linker, and allowing probe molecules to become bound to the specimen;

b) removing unbound probe molecules from the specimen;

c) contacting the specimen with a MALDI matrix material (or MS-compatible matrix material), or laying the specimen on a surface coated with a MALDI matrix material (or MS-compatible matrix material);

d) irradiating a portion of the specimen with a laser beam to release ions from the specimen;

e) selecting, from the ions released in step (d), ions with mass-to-charge ratios corresponding to the mass tags or derivatives of the mass tags;

f) recording the amount and type of ions selected in step (e), together with the location of the portion of the specimen, as a result;

g) repeating steps (d) to (f) on a different portion of the specimen.

Further aspects and embodiments of the invention are set out in the appended claims and/or described below.

MALDI MS/MS is versatile in its many applications to the analysis of biological samples, especially to peptides and proteins. Typically, samples are mixed with an organic compound, which acts as a matrix to facilitate desorption and ionisation of compounds in the sample. The matrix provides the required sensitivity and specificity for use of laser desorption techniques in the analysis of biological material. The application of thin layers of matrix has special advantages, particularly when very high sensitivity is needed. Methods are also disclosed for the preparation of cellulose membranes precoated with a thin matrix layer for the direct deposition and analysis of aqueous samples. This technique circumvents the problems of mixing and dilution of samples when post addition of matrix is done and effectively allows small (nanoliter) volumes of samples to be applied to the target.

One aspect of the invention is the use of MALDI MS/MS techniques for the imaging of specimens, e.g. tissue sections, where the spatial arrangement of nucleic acid or polypeptide expression is to be determined. Two different approaches may be used: direct targeting of the tissue itself, or analysis of blotted targets previously exposed to the tissue.

The ability to image a specimen in order to obtain the detailed spatial arrangement of compounds in an ordered target specimen such as a slice of tissue using MALDI MS/MS is of enormous value in biological research. Tag daughter ion maps derived from contacting such specimens with tagged probes could provide for example details of drug modes of action, off-target activities of drugs, the roles of genes in developmental biology and mechanisms of disease.

Another aspect of the invention provides probe molecules suitable for the analysis of a specimen, such as a tissue sample, by MS particularly MALDI MS/MS-based imaging.

A further aspect of this invention provides methods for employing said probe molecules to image a specimen using MS (such as MALDI MS/MS) instrumentation.

A further aspect of this invention provides methods to allow for rapid acquisition of high-resolution images of a specimen by MS/MS.

MALDI scanning may be performed with an ultraviolet laser or an infrared laser.

Suitable matrices for UV MALDI include cinnamic acids, sinnapinic acid, hydroxypicolinic acid, nicotinic acid, hydroxybenzoic acid. Matrices for Infrared MALDI include glycerol, urea, succinic acid and water.

MS/MS analysis may be performed in, for example, an ion trap, TOF-TOF, quadrupole-TOF, ion-trap TOF, or Fourier transform ion cyclotron resonance instrument.

A still further aspect of the present invention provides improved techniques which use tandem mass spectrometry to determine, and preferably to visually depict, quantitative information regarding molecules of interest as a function of the spatial arrangement of numerous successive laser spots on a specimen.

A related aspect of the present invention improves the capability of MALDI MS/MS by providing improved methods for quantification of tag ions and for normalisation of the multiple spectra that are used to generate graphic displays of tag ions and consequently the distribution of the molecules of interest that are bound by the corresponding probes from which the tags are released.

A further aspect of the invention improves the speed of imaging mass spectrometry by reducing redundancy in image acquisition thus increasing the throughput of imaging mass spectrometry systems.

It is a further aspect of the present invention that improved techniques are provided for analysing the spatial arrangement of specific molecules within a specimen, such as a tissue sample, by tandem mass spectrometry analysis. A very thin sample layer may be generated and combined with an energy absorbent matrix to form a test specimen, which is then sequentially struck by a laser beam. The sample layer generated is preferably less than 50 microns in thickness.

One feature of the invention is that the intensity or normalised intensity of daughter ion fragments from tag ions may be graphically depicted as a function of the linear distance between successive laser spots.

Another feature of the invention is that the intensity of daughter ion fragments from tag ions may be normalised by comparison with the intensity of a calibrant ion.

It is a further feature of the invention that blotting techniques may be used to blot the sample on a blotting surface. The blotting surface may be a liquid absorbing surface, a chemically prepared surface, or a biologically prepared surface.

Another feature of the present invention is a technique for substantially drying the sample to minimise movement of sample molecules within the sample layer prior to striking the test specimen with laser pulses. The sample layer may be dried in a vacuum dessicator or a hydrolyser for at least two hours.

As used herein, the term "nucleic acid or polypeptide expression" refers to the expression of a variety of molecular species. Nucleic acids include messenger RNA (mRNA), micro-RNA (miRNA), transfer RNA (tRNA) and other endogenously expressed ribonucleic acids or deoxyribosenucleic acids. Similarly, polypeptide expression includes polypeptides, peptides, amino acids, peptide hormones, lipoproteins, carbohydrate-modified proteins, carbohydrates and other products of protein expression.

As used herein, the term "collision" where used on its own encompasses the tem "collision induced dissociation" (CID).

Particular non-limiting examples of the present invention will now be described with reference to the accompanying drawings, in which:

Fig. 1 depicts an example of a probe according this invention;

Fig. 2 is a schematic illustration of a tandem mass spectrometry apparatus for use in a method in accordance with the invention;

Fig. 3 is a schematic representation of a synthesis method for the preparation of singly labelled probes of this invention;

Figs 4(a) and 4(b) are a schematic representation of a synthesis method for the preparation of multiply labelled probes of this invention;

Fig. 5 is a schematic representation of an alternative synthesis method for the preparation of multiply labelled probes of this invention;

Fig. 6 illustrates a series of isobaric mass tag sets according to this invention;

Fig. 7 is a flow diagram showing the method steps carried out in accordance with an embodiment of the invention;

Fig. 8 shows a mass spectrum of a photocleaved and desorbed tag peptide. The x-axis shows mass-to-charge ratio (m/z) and y-axis % intensity;

Figs 9a and 9b show alternative mechanisms for fragmentation of the amide backbone of a peptide;

Fig. 10 shows (a) the structure of an exemplar mass tag Peptide Tag 1 (SEQ ID NO: 12; top) and its photocleavage daughter fragment (bottom), and (b-d) MALDI MS/MS spectra graphs following crystallisation of Peptide Tag 1 (SEQ ID NO: 12) in three different matrices (4-hydroxy-alpha-cyano-cinnamic acid [HCCA], 2,5-dihydroxybenzoic acid [DHB] and sinapinic acid [SA], respectively). In Figs 10(b)-(d), the x-axis shows laser power and the y-axis intensity, and -●- indicates the peptide while -■- indicates the fragment;

Fig. 11 shows further MALDI MS/MS spectra graphs of Peptide Tag 1 (SEQ ID NO: 12; -♦-) and its photocleavage daughter fragment (-■-) using (a-c) Post Source Delay (PSD) and (d-f) Collision Induced Dissociation (CID) analysis. Fig. 11(a) and (d) show HCCA, Fig. 11(b) and (e) show DHB, and Fig. 11(c) and (f) show SA. In each graph, the x-axis shows laser power and the y-axis intensity;

Fig. 12 is a graph showing optimal SA concentration for PSD MALDI MS/MS analysis of Peptide Tag 1 (SEQ ID NO: 12). The x-axis shows SA concentration in mg/ml, the y-axis intensity. Parent (-♦-) and daughter fragment (-■-) results are shown;

Fig. 13 shows(a) the structure of further exemplar mass tags Peptide Tag 2 (SEQ ID NO: 13; top left), Peptide Tag 3 (SEQ ID NO: 14; top right) and Peptide Tag 4 (SEQ ID NO: 15; bottom), (b) a MALDI MS/MS spectrum of desorbed and photocleaved oligonucleotide-conjugated Peptide Tag 2 (CONJUGATE2) and Peptide Tag 3 (CONJUGATE3), and (c) an intensity-adjusted spectra for the three tags. In Fig. 13(b) and (c), the x-axis shows mass-to-charge ratio (m/z) while the y-axis shows % intensity;

Fig. 14 shows (a) various MALDI MS/MS images detecting verapamil in the presence of different background ions ("DPPE" = dipalmitoyl phosphatidylethanolamine, "DPPC" = Dipalmitoyl Phosphatidylcholine, "C" = cholesterol, "N" = no lipid), MALDI MS/MS images of CONJUGATE2 (b-i; DPPE and DPPC as above, "BSA" = bovine serum albumin, "T" = tag alone) and CONJUGATE3 (b-ii; wells as b-i) in the presence of different backgrounds ions, and (b-iii) normalisation of the intensity of CONJUGATE2 with CONJUGATE3;

Fig. 15(a) and (b) shows results of a dilution series of CONJUGATE2 (left side) and CONJUGATE3 (right side), with Fig. 15(a) middle showing the intensities (y-axis) of the daughter ions at m/z (x-axis) 257 and 263. In each graph of Fig. 15(b), the x-axis shows pmol of probe molecule while the y-axis shows average intensity;

Fig. 16 shows images of a pair of mouse kidney tissue sections (left side) spiked with a synthetic Her2 target detected (right side) by MALDI MS/MS using CONJUGATE2;

Fig. 17 shows the structure of two further peptide tags according to the present invention, viz. Peptide Tag 5 (SEQ ID NO: 19; top) and Peptide Tag 6 (SEQ ID NO: 20; bottom); and

Fig. 18 shows a compound mass spectrum (x-axis in m/z; y-axis in intensity for each sub-spectrum) following photocleavage of the following four samples (1) CONGUGATE4 alone, (2) CONGUGATE4 with CONJUGATE_1TAG, (3) CONGUGATE4 with CONJUGATE_2TAG, and (4) CONGUGATE4 with CONJUGATE_4TAG.

Imaging of gene and protein expression directly in tissues is a highly informative way to determine gene function and activity while relating the expression to physical features of the tissue. Typically, radiolabels, fluorescent labels or other optical labels are used, but all of these standard techniques are limited by the number of tags available for use with these techniques. The present invention allows for simultaneous detection of expression of many, if not all, genes in a sample. This allows for better use of precious samples, ensures that changes in one gene can be compared with others directly and ensures that all genes are subjected to the same experimental conditions, thus improving data quality.

Another issue with optical tags is that most tags are quite different chemically from each other, which introduces variation in labelling and detection efficiencies. These label biases can distort quantification of gene expression quite significantly. The present invention addresses this issue.

In one aspect of the invention, a probe molecule is provided comprising an oligonucleotide coupled to one or more mass tags by photocleavable linkers:

$[\text{mass tag} - \text{photocleavable linker}]_n - \text{oligonucleotide}$,

where "n" is any positive integer, for example n = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or higher integer.

In a further embodiment of this first aspect of the invention the mass tag has the form:

mass modifier 1 – collision cleavable linker – mass modifier 2.

The mass tag additionally comprises at least one charge-carrying group. If more than one charge-carrying group is used, these may all be on one side of the collision cleavable linker. The charge-carrying group may be a separate component from the mass modifier or the mass modifier may itself comprise a charge-carrying group.

Alternatively the collision cleavable linker may comprise a charge-carrying group as long as the charge ends up consistently on only one of the fragments generated by collision.

The mass tag may comprise or consist of a peptide. For the purposes of analysis by MS/MS, the collision cleavable linker may comprise a group that is readily cleaved upon low energy collision in the mass spectrometer. The collision sensitive group may be selected from a proline linkage, a piperazine linkage, a piperidine linkage, an aspartic acid linkage or a linkage comprising aspartic acid and proline adjacent to each other with the aspartic acid on the N-terminal side of the proline residue, as is described below in more detail.

A set of two or more oligonucleotides conjugated to mass tags by photocleavable linkers may be provided where the mass tags are isobaric but may be resolved from each other by tandem mass spectrometric methods. Isobaric means that the total mass of each mass tag is the same for every tag in a set. This is achieved by ensuring that the sum of the masses of mass modifier 1 and mass modifier 2 are the same for each mass tag in the set but the actual masses of mass modifier 1 and mass modifier 2 are different from each other. Tags are resolved by fragmentation of the collision cleavable linker according to the methods of the invention to release tag fragments that have uniquely resolvable masses.

The isobaric mass tags may be isotopes of each other. This may be achieved by synthesising peptides that have the same sequence where mass modifier 1 and mass modifier 2 comprise amino acids that have been mass modified using isotopes, preferably stable isotopes. Preferred stable isotopes include ^{13}C, ^{15}N, ^{17}O, ^{18}O, ^{34}S and deuterium (^{2}H).

The tags of this invention can be produced in large numbers and they can be constructed as isotopic sets so that their properties are as similar to each as can possibly be achieved. Thus, imaging of tissue sections, microarrays or other planar samples with the tags of this invention can offer significant advantages over the prior art.

In a typical imaging experiment labelled probes are contacted with the target specimens, for example a tissue or array. Probes that find their target will bind, allowing unbound probes to be removed, typically by washing or rinsing the sample with a suitable buffer. Binding buffers and washing buffers typically contain only volatile salts such as ammonium carbonate, ammonium acetate and similar compounds. After removal of unbound probes the planar sample is typically coated with a MALDI matrix material or laid on a surface

coated with a MALDI matrix. The planar sample may then be introduced into a suitable laser desorption mass spectrometer where it is scanned. A raster scan may be performed or a directed analysis of specific locations may be carried out. In a raster scan, the laser is focussed successively on adjacent locations in a line and the whole planar sample is scanned line-by-line so that the whole sample is imaged. At each location a mass spectrum is obtained. In the context of this invention, the mass spectrum is preferably an MS/MS spectrum. The desorbed ions from each location enter the mass analyser. Ions with the mass-to-charge ratio that corresponds to ratios expected for tags ion are gated into a collision cell (or in the case of ion traps and FT-ICR instruments the ions are selectively retained in the trap). In the collision cell or trap, the ions are subjected to collision with a bath gas such as argon or helium or neon. The fragments are then analysed to detect the characteristic daughter ions that are produced by the tags of this invention. This process of selection and detection enhances selectivity and specificity enabling tags to be distinguished from other ions in the sample. This is an important feature of this invention as tissues comprise complex mixtures of molecules that will generate complex mass spectra. The intensities of the tag daughter ions may be recorded for each location that is analysed on the planar sample.

Fig. 1 illustrates an example of a labelled probe molecule of the form "[mass tag – photocleavable linker]$_1$ – oligonucleotide" in accordance with the invention, where the mass tag is a peptide. As a result of the scan, an ionised mass tag (with $[M+H]^+ = 328.2$ in this example) is photocleaved (shown as "P" in Fig. 1) from the probe molecule. Then, in the collision cell or trap, the mass tag is fragmented by collision with a bath gas (a process known as collision induced dissociation [CID]) to produce a fragment (in this example with $[M+H]^+ = 215.15$). It is these resulting fragments that are analysed to provide an indication of the amounts target molecules in different portions of the specimen.

A map or image of the relative amounts of the target molecule identified by each tag can then be constructed. Each tag can be displayed in a different colour on a single image or each tag can be viewed as a distinct image.

The area or spot irradiated by the desorption laser can be varied on some instruments, and the spot area is an important variable in the imaging process particularly for tissue imaging. The larger the spot size, the fewer spots needed to cover a given area and the

lower the amount of data acquired, thereby reducing data burden, and scanning time. However, the larger the spot size, the poorer the ability to spatially resolve adjacent locations of molecules of interest. With very small spot sizes, i.e. 1 μm diameter or less, the resolution of the resulting image improves, but this is at the expense of time (more spots needed to cover a given area), and at the expense of increased data load. Varying spot size can be used as a way to gather data quickly before focussing on an area of interest. For example, the sample may be exposed to a small number of larger spot sizes initially to provide a "survey scan". If the molecules of interest are present at specific location, then smaller spot sizes can be used to increase the resolution of the image so produced.

The tissue sections may be dried before analysis, such as in a vacuum desiccator. Alternatively, tissue sections may be fixed, for example using paraformaldehyde.

In the context of tissue sections, the flatness of the sample is not critical to the methods of this invention but most tissue sections will start flat during the sectioning process. Thus, to ensure the resulting image matches the original structure of the tissue, it may be important to maintain the flatness of section during the many preparation steps that are needed prior to analysis. For thick sections (80-100 μm), it may be necessary to dry the section slowly over several days to get best results as fast drying tends to warp the specimen making it complicated to mount. Thin sections (1-10 μm) dry much faster and can be kept flat on the mounting membrane more easily. Surface levels that differ significantly in height can produce different mass shifts in simple Time-of-Flight analysis, although with Trap-TOF geometries this is not a major issue as the ions are collected in the trap and subjected to collisional cooling prior to further analysis.

The invention can be practiced with various different kinds of mass spectrometer and some of the more favourable geometries are described below.

Mass spectrometers typically comprise the following components:

Interface – Mass Analyser – Detector – Data Capture and Analysis System.

For the purposes of Imaging Mass Spectrometry, it is preferred that Laser Desorption Interfaces are used. Typically these require that the sample is introduced into a chamber in the instrument, which is evacuated and where the sample can be exposed to a laser beam. The laser source typically generates a laser beam that passes through a beam adjusting mechanism or mask and/or suitable optics that then strikes a sample target containing the test sample, thereby releasing ions of interest with various mass-to-charge ratios, which are then analysed by the mass analyser. Ion optics deliver the ions to the mass analyser. Most MALDI instruments have a capability to focus the laser at successive locations on a planar sample to allow a sample to be scanned. Atmospheric Pressure Laser Desorption interfaces are also commercially available and are suitable for the practice of this invention (Schneider et al., 2005, J. Am. Soc. Mass Spectrom **16**: 176-182).

Most mass analysers separate ions according to mass-to-charge ratios. They may also isolate or gate ions with specific mass-to-charge ratios and subsequently fragment these selected ions. Ions, whether trapped, gated and/or fragmented are then delivered to a detector, which counts the ions. These counts are then converted into mass spectra by the data capture and analysis system.

Fourier Transform instruments detect radio waves that are emitted by moving ions trapped in the instrument. These instruments are operated so that the frequencies of the radio waves can be used to determine the mass-to-charge ratios of trapped ions. In these systems ions are not detected directly but most of the same principles apply in the subsequent data analysis as the signals are typically converted into ion counts.

Imaging by MALDI MS/MS Mass Spectrometry

MALDI mass spectrometry has become a widely used tool for the analysis of many types of biological molecules, especially peptides and proteins. MALDI requires that the biomolecule solution be embedded in a large molar excess of a photo-excitable "matrix". The application of laser light of the appropriate frequency results in the excitation of the matrix, which in turn leads to rapid evaporation of the matrix along with its entrapped biomolecule. Proton transfer from the acidic matrix to the biomolecule gives rise to protonated forms of the biomolecule, which can be detected by positive ion mass spectrometry, particularly by Time-Of-Flight (TOF) mass spectrometry. Negative ion mass

spectrometry is also possible by MALDI TOF. MALDI imparts a significant quantity of translational energy to ions, but tends not to induce excessive fragmentation despite this. Accelerating voltages can again be used to control fragmentation with this technique though. This technique is highly favoured for the determination of peptide mass fingerprints due to its large mass range, due to the prevalence of singly charged ions in its spectra and due to the ability to analyse multiple peptides simultaneously.

The photo-excitable matrix comprises a "dye", i.e. a compound that strongly absorbs light of a particular frequency, and which preferably does not radiate that energy by fluorescence or phosphorescence but rather dissipates the energy thermally, i.e. through vibrational modes. It is the vibration of the matrix caused by laser excitation that results in rapid sublimation of the dye, which simultaneously takes the embedded analyte into the gas phase.

Mass Analysers

Time-of-Flight Mass Analysers

As the name implies, Time-of-flight mass analysers measure the time it takes for ions to travel a predetermined distance under the influence of a predetermined potential difference. The time-of-flight measurement allows the mass-to-charge ratio of ions striking a detector to be calculated. These instruments measure the arrival of almost all of the ions in a sample and as a result can be quite sensitive, although selectivity with this technique is more difficult to achieve. This technique can also detect ions with higher mass-to-charge ratios than can be typically measured in an ion trap or quadrupole mass spectrometer. TOF mass analysers are presently widely used with MALDI. Hybrid instruments that enable MS/MS analysis using TOF analysers are particularly well suited to the methods of this invention, including quadrupole-TOF geometries, trap-TOF geometries and TOF-TOF geometries.

Fig. 2 show an embodiment of an instrument 100 suitable for imaging mass spectrometry in accordance with the methods of this invention. The MALDI MS/MS instrument, 100, comprises a vacuum housing, 101, that contains the instrument components and which has a pumping device and other standard vacuum components (not shown).

The instrument 100 comprises a laser desorption ionisation (LDI) source 110. This is shown contained within the vacuum housing 101 although atmospheric pressure LDI sources are also applicable with this invention. The instrument 100 further comprises a quadrupole ion trap 120 and a Time-Of-Flight analyser 130.

The LDI source 110 comprises a laser 111, optionally a mirror 112 and optionally other optical elements (not shown). The mirror may be movable by servo motors or other means to allow the laser (dotted line leaving the laser 111) to be scanned across the sample stage 113, where planar samples for analysis are deposited. The ions that exit the LDI source 110 may be steered into a quadrupole ion trap 120 by electrostatic focusing elements 114. The quadrupole ion trap 120 traps ions created within the LDI source 110. The quadrupole ion trap 120 typically includes a ring electrode 121 that is separated from a pair of endcap electrodes 122 and 123 by dielectric material 124. The elements 114 and endcap 122 may function as an einsel lens. Typically, the first and last element of the einsel lens (the first element of 114 and electrode 122) can be biased at the same voltage, such as ground potential, and the middle element (the second element of 114) is biased at a different potential. Alternatively, the lenses can have progressively decreasing potential to accelerate the ions. Other focusing elements have been tested and may include multipole ion guides, such as quadrupole, hexapole, or octapole ion guides.

The ions enter the quadrupole ion trap 120 and are stabilised and stored within the trap by the application of an alternating current to the ring electrode 121 in a manner known in the art. The endcaps 122 and 123 are usually held at a constant voltage, such as ground potential, however, auxiliary oscillating current may be applied. The range of ion masses that are stored efficiently depends on the frequency and amplitude of the current applied to the ring electrode, it is typically of radio frequency (such as 1 MHz) and a few hundred to a few thousand volts peak to peak, but can have other values. The ions may continuously accumulate within the trap 120. Waveforms can be applied to one or both endcap electrodes 122 and 123 and/or to the ring electrode 121 to excite specific ion masses in the trap 120 in order to eject them from stable orbits, to prevent them from accumulating, or to excite them to more energetic orbits to cause them to dissociate with background gas in order to produce fragment ions of the selected ions. Each ion mass has a distinct resonance condition. Many different ion masses may be excited simultaneously by applying a superposition of many frequencies. The frequency spectrum may be generated

by a variety of prior art methods. In this embodiment, the arbitrary waveform is formed by superimposing the sum of individual periodic waveforms corresponding to the frequency and amplitude most suited for exciting each ion mass to the desired effect. In one embodiment this waveform may be applied to the exit endcap 123, although it is to be understood that effective excitation may be achieved by application of the waveform to other electrodes in the ion trap.

Following accumulation of ions and the optional manipulation of ions consisting of selective ion ejection and selective collision-induced ion dissociation, the remaining ions in the quadrupole ion trap 120 are mass analysed by ejecting all the ions into a Time-Of-Flight mass analyser 130. Ion ejection to the mass analyser 130 may be accomplished by applying a high voltage pulse to the ion trap exit endcap 123. Alternatively, a high voltage pulse may be applied to the entrance endcap 122 to "push" the ions into the detector 130, or two oppositely-phased pulses may be applied to both endcaps 122 and 123 in a "push-pull" manner to extract ions into the TOF mass analyser 130.

The extracted ions can be accelerated to a higher energy by an acceleration grid 131. The accelerated ion pulse may be focused and collimated by a electrostatic lens assembly 132. This is shown as a three-element einsel lens, however other configurations may be used, such as a two-element assembly. The third element in the einsel lens configuration can make use of the back plate of the detector 133.

The accelerated and collimated ion packet passes through a hole 134 in the coaxial detector 133. A cylinder 135 may be provided in the detector hole 134 to keep a uniform voltage potential for the traversing ions and is electrically isolated from the detector plates themselves (described below). The ions travel through a drift tube 136 under field-free conditions where ions of different mass travel at different speeds and spread out in space. The ions may then reach a reflectron section 137 of the mass detector where they are reversed in direction. This operation acts to focus ions of different initial kinetic energies. The ions then travel back toward the front of the detector 133 where they impact and are recorded as a signal. The resulting signal from the detector is measured with electronics that can distinguish the different arrival times of different ion masses. Although a reflectron 137 is shown with a coaxial detector in the mass spectrometer (100), it is to be understood

that the reflectron 137 may also be of an off-axis design, or the TOF mass analyser 130 may be of a linear design with the detector plate 133 at the end of the drift tube 136.

The instrument 100 is very well suited for the practice of this invention. A tissue section or other planar sample can be scanned by a laser in the ion source 110. Tag ions of the present invention can then be collected in the quadrupole ion trap 120. Non-tag ions, i.e. ions with different mass-to-charge ratios from the tags can be selectively ejected from the trap by application of appropriate waveforms to the trap electrodes. The tag ions can then be fragmented in the trap by application of suitable waveforms. Then, by applying appropriate potentials to the electrodes 122 and 123, fragment ions can be pushed and/or pulled into the TOF analyser 130 for mass determination and detection.

In the TOF-TOF geometry, the ion source is linked to a TOF drift tube in which ions can separate according to flight time. This drift region is separated from a full TOF analyser by a gate and collision cell arrangement that allows ions from the first drift region to be selected and fragmented prior to mass spectrometry analysis in the second TOF analyser where the mass-to-charge ratios of the gated and fragmented ions can be determined. When the gate and collision region are not in use the instrument acts as a single MS-mode TOF analyser. This geometry is quite advantageous for the purposes of this invention due to its high repetition rate, i.e. the rate at which spectra can be obtained.

Ion Traps

Ion Trap mass analysers are related to the quadrupole mass analysers. An ion trap generally has a 3 electrode construction - a cylindrical electrode with "cap" electrodes at each end forming a cavity. A sinusoidal radio frequency potential is applied to the cylindrical electrode while the cap electrodes are biased with DC or AC potentials. Ions injected into the cavity are constrained to a stable circular trajectory by the oscillating electric field of the cylindrical electrode. However, for a given amplitude of the oscillating potential, certain ions will have an unstable trajectory and will be ejected from the trap. A sample of ions injected into the trap can be sequentially ejected from the trap according to their mass/charge ratio by altering the oscillating radio frequency potential. The ejected ions can then be detected allowing a mass spectrum to be produced.

Ion traps are generally operated with a small quantity of a "bath gas", such as helium, present in the ion trap cavity. This increases both the resolution and the sensitivity of the device as the ions entering the trap are essentially cooled to the ambient temperature of the bath gas through collision with the bath gas. Collisions both increase ionisation when a sample is introduced into the trap and dampen the amplitude and velocity of ion trajectories keeping them nearer the centre of the trap. This means that when the oscillating potential is changed, ions whose trajectories become unstable gain energy more rapidly, relative to the damped circulating ions and exit the trap in a tighter bunch giving narrower larger peaks.

Ion traps can mimic tandem mass spectrometer geometries, in fact they can mimic multiple mass spectrometer geometries allowing complex analyses of trapped ions. A single mass species from a sample can be retained in a trap, i.e. all other species can be ejected and then the retained species can be carefully excited by super-imposing a second oscillating frequency on the first. The excited ions will then collide with the bath gas and will fragment if sufficiently excited. The fragments can then be analysed further. It is possible to retain a fragment ion for further analysis by ejecting other ions and then exciting the fragment ion to fragment. This process can be repeated for as long as sufficient sample exists to permit further analysis. It should be noted that these instruments generally retain a high proportion of fragment ions after induced fragmentation. These instruments and FTICR mass spectrometers (described below) represent a form of temporally resolved tandem mass spectrometry rather than spatially resolved tandem mass spectrometry which is found in linear mass spectrometers.

Traps can also be combined with TOFs in so-called TRAP-TOF geometries. These are quite advantageous for the purposes of this invention as the TRAP collects ions, selects ions and fragments ions that are then injected into the TOF for mass determination. The TOF typically has much higher mass resolution than traps and a wider dynamic range. In addition, the trap can continue isolating ions while the TOF is analysing ions giving good duty cycle.

Fourier Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR MS)

FTICR mass spectrometry has similar features to ion traps in that a sample of ions is retained within a cavity but in FTICR MS the ions are trapped in a high vacuum chamber

by crossed electric and magnetic fields. The electric field is generated by a pair of plate electrodes that form two sides of a box. The box is contained in the field of a superconducting magnet which in conjunction with the two plates, the trapping plates, constrain injected ions to a circular trajectory between the trapping plates, perpendicular to the applied magnetic field. The ions are excited to larger orbits by applying a radio-frequency pulse to two `transmitter plates`, which form two further opposing sides of the box. The cycloidal motion of the ions generates corresponding electric fields in the remaining two opposing sides of the box, which comprise the `receiver plates`. The excitation pulses excite ions to larger orbits which decay as the coherent motions of the ions is lost through collisions. The corresponding signals detected by the receiver plates are converted to a mass spectrum by Fourier Transform (FT) analysis.

For induced fragmentation techniques these instruments operate in a similar manner to an ion trap: all ions except a single species of interest can be ejected from the trap. A collision gas can be introduced into the trap and fragmentation can be induced. The fragment ions can be subsequently analysed. Generally fragmentation products and bath gas combine to give poor resolution if analysed by FT analysis of signals detected by the `receiver plates`, however the fragment ions can be ejected from the cavity and analysed in a tandem configuration with a quadrupole, for example.

Reagents

The methods of this invention require a variety of reagents, which are described in detail below.

Probe synthesis

Isobarically mass tagged probe oligonucleotides and any other oligonucleotides, such as multimeric oligonucleotides (for example branched or comb oligonucleotides; described below) and the components of dendrimers can be synthesised using standard oligonucleotide synthesis methods known in the art. Preferred methods are purely synthetic methods, for example, by the cyanoethyl phosphoramidite method (Beaucage & Caruthers, 1981, Tetrahedron Lett. **22**: 1859-1862; McBride & Caruthers, 1983, Tetrahedron Lett. **24**: 245-248). Synthetic methods useful for making oligonucleotides are also described by Ikuta et al. (1984, Ann. Rev. Biochem. **53**: 323-356), (phosphotriester and phosphite-triester methods), and Narang et al. (1980, Methods Enzymol. **65**: 610-620)

(phosphotriester method). PNA molecules can be made using known methods such as those described by Nielsen et al. (1994, Bioconjug. Chem. **5**: 3-7). PNA is a preferred oligonucleotide analogue for the practice of this invention as PNA is able to hybridise under low salt conditions, even in water if necessary. This means that hybridisation to nucleic acids in tissue sections can take place under whatever conditions suit the experiment.

For many applications of the oligonucleotides of the present invention it is useful to know how stable they are, or more specifically at what temperature they will dissociate. The stability of DNA duplexes can be calculated using known methods for prediction of melting temperatures (Breslauer et al., 1986, PNASUSA **83**(11): 3746-3750; Lesnick & Freier, 1995, Biochemistry **34**:10807-10815; McGraw et al., 1990, Biotechniques **8**: 674-678; and Rychlik et al., 1990, Nucleic Acids Res. **18**: 6409-6412).

Nucleic Acid Analogues

In embodiments of the invention, it may be useful to employ nucleic acid analogues with enhanced binding affinity compared to natural phosphodiester deoxyribosenucleic acids. It is known that RNA analogues with certain modifications at the 2' position of the ribose ring show enhanced binding affinity for RNA targets compared to corresponding DNA/RNA hybrids (see Cummins et al., 1995, Nucleic Acids Res. **23**(11): 2019-24). These RNA analogues also show reduced binding affinity for DNA compared to DNA/DNA hybrids (Tsourkas et al., 2003, Nucleic Acids Res. **31**(6): 5168-74). The ability to bind preferentially to RNA over DNA with enhanced melting temperature makes 2'-modified analogues particularly useful for in situ hybridisation applications for detection of alternatively spliced RNA in a background of genomic DNA.

2'-O-methyl analogues in particular are readily available as phosphoramidite monomers for automated synthesis and are suitable for use with this invention. Additionally or alternatively, 2'-fluoro-modified analogues may be used.

Another nucleic acid analogue for use with this invention are "bridged" analogues such as locked nucleic acids ("LNA"; Thomsen et al., 2005, RNA. **11**(11):1745-8) and 2'-4'-BNA(NC) (Rahman et al., 2008, J Am Chem Soc. **130**(14): 4886-96). Bridged nucleic acid analogues show enhanced binding affinity for RNA compared with their natural nucleic

acid counterparts, and are thus suitable for in situ hybridisation applications. Bridged analogues also show enhanced binding affinity for DNA compared with their natural nucleic acid counterparts, and are therefore useful for detection of chromosomal targets such as chromosomal translocations and for the detection of labelled cDNAs.

It is not normally desirable to synthesise probes that are comprised entirely of bridged analogues. Hence LNA monomers are typically introduced every third base into DNA probes (Válóczi et al., 2004, Nucleic Acids Res. **32**(22): e175; Obernosterer et al., 2007, Nat Protoc. **2**(6): 1508-14). LNA monomers may be introduced into 2'-O-methyl oligonucleotide sequences to enhance binding affinity of the resultant probe (Kierzek et al., 2005, Nucleic Acids Res. **33**(16): 5082-93). When LNA monomers are introduced into 2'-O-methyl probes, the LNA monomers may be positioned every second base but in one embodiment not at the 5' end of a probe.

PNA is another analogue for use with this invention (Nielsen et al., 1994, Bioconjug Chem. **5**(1): 3-7). PNA has enhanced binding affinity for both DNA and RNA targets compared to DNA probes. PNA is less soluble than other DNA analogues and it is currently difficult to produce usable PNA oligonucleotides with a length greater than 20 bases. Hence PNA probes may be shorter than probes made with sugar/phosphate backbones. The invention encompasses chimeric probes comprising lengths of PNA and DNA (see Uhlmann, 1998, Biol Chem. **379**(8-9): 1045-52). These chimeric probes may be longer than PNA-only probes.

Mass Tagged Oligonucleotides

A variety of mass tags can be used with this invention although preferred mass tags (also known as "tags" or "mass markers") are disclosed in WO 97/27327, WO 97/27325, WO 97/27331 and WO 03/025576. Those publications disclose tags that comprise polyamide compounds, essentially peptides or peptide-like tags, which means that these tags can be prepared using a number of peptide synthesis methods that are well known in the art (see for example Jones, 1991, "The chemical synthesis of peptides", Oxford University Press; Fields & Noble, 1990, Int J Pept Protein Res **35**(3): 161-214; Albericio, 2000, Biopolymers **55**(2):123-139). In addition, the use of peptide and peptide-like tags enables coupling of these tags to oligonucleotides using a variety of peptide conjugation techniques that are known in the art. Methods for coupling peptides to oligonucleotides "on column" via 5'

amine functionalities are disclosed in Zaramella et al. (2004, J Am Chem Soc **126**(43): 14029 – 14035). Methods for conjugating peptides to oligonucleotides via thiol groups at the termini of the oligonucleotides are disclosed in Arar et al. (1995, Bioconjug Chem. **6(5)**: 573-577). Oligonucleotides can be coupled to peptides with terminal cysteine residues as disclosed in Wei et al. (1994, Bioconjug Chem. **5(5):** 468-74).

In one approach, peptide tags can be synthesised on Controlled Pore Glass (CPG) beads of the kind used for DNA synthesis. As long as these peptides are suitably protected and as long as the peptide has a free hydroxyl (or trityl protected hydroxyl) at the end of the peptide synthesis, an oligonucleotide can be synthesised directly on the peptide (Haralambidis et al., 1990, Nucleic Acids Res. **18**(3): 493-499). The peptide should be linked to the resin with a suitable linker that will be cleaved under conditions normally used for cleaving and deprotecting oligonucleotides. A 4-Hydroxymethylbenzoic acid (HMBA) linker (Sheppard et al., 1982, Int. J. Peptide Protein Res. **20**: 451) is base-cleavable and suitable for both FMOC peptide synthesis and oligonucleotide synthesis. This can be readily coupled to amino-functionalised CPG beads such as aminopropyl-CPG, which is commercially available.

An example of a synthesis according to this embodiment of the invention is illustrated in Fig. 3. In Fig. 3 it can be seen that tag oligonucleotide conjugates according to this invention may be prepared by initially synthesising the peptide tag component of the conjugate on a controlled pore glass (CPG) resin that is compatible with automated oligonucleotide synthesis. In step (1) of Fig. 3, aminopropyl CPG is initially coupled to the aforementioned HMBA linker. Peptide synthesis is initiated from this linker. In step (2) of the method in Fig. 3, an O-trityl protected serine residue is coupled to the HMBA linker. Spacers, such as beta-alanine, could be introduced between the serine and the HMBA linker if desired to reduce potential steric interference from the resin. In step (3), the desired peptide tag is synthesised using commercially available FMOC-amino acids, typically using multiple cycles of FMOC peptide synthesis. In this example, an FMOC protected photocleavable linker (4-[4-(1-(FMOC-amino)ethyl)-2-methoxy-5-nitrophenoxy)butanoic acid available from IRIS Biotech GmbH, Marktredwitz, Germany) is introduced followed by FMOC-alanine, followed by 1-Fmoc-piperidin-4-ylacetic acid (Sigma Aldrich, UK), followed again by FMOC-alanine, which is finally acetylated to block the terminal amino group so that it does not interfere in subsequent oligonucleotide

synthesis. Between step (3) and step (4), the CPG resin is transferred from the automated peptide synthesiser reaction vessel or column into a column or vessel suitable for use in an automated DNA synthesiser. In step (4) of Fig. 3, the trityl protection group is removed using standard deprotection conditions in the DNA synthesiser, typically trichloroacetic acid or dichloroacetic acid in dichloromethane. In step (5), an optional Dimethoxtrityl (DMTr) protected linker is introduced to reduce steric hindrance from the solid support and the peptide (Spacer Phosphoramidite 9 from Glen Research, Stirling, VA, USA). A wide variety of similar spacers is known in the art and could be substituted for the linker shown, if desired. In step (6), multiple cycles of standard automated (phosphoramidite) DNA synthesis are carried out to generate the desired DNA sequence linked to the peptide. In the final step (7), the side chain deprotection groups are cleaved along with the HMBA linker, typically using NH_3/H_2O steps, releasing the deprotected conjugate into solution. The released conjugate would usually then be purified by high performance liquid chromatography, gel filtration, gel electrophoresis or other standard techniques known in the art.

The method shown in Fig. 3 and described above conjugates the peptide tag at the 3' end of the oligonucleotide. However, "reverse" synthesis (5' to 3') is also possible (see Claeboe et al., 2003, Nucleic Acids Res. **31**(19): 5685-5691). Reverse synthesis phosphoramidites used with the method in Fig. 3 (and also the methods in Figs 4 and 5) would yield oligonucleotides modified at the 5' terminus. Phosphoramidite monomers for 5' to 3' synthesis are commercially available, for example from Glen Research Corporation (Sterling, Virginia, USA).

Many amino acids are known in the art and not all are available as FMOC derivatives. It is, however, comparatively simple for one of ordinary skill in the art to prepare FMOC protected derivatives of unprotected amino acids using standard methods for use with this invention (Fields et al., 1990, Int J Pept Protein Res. **35**(3): 161-214).

PNA peptide conjugates can be readily synthesised as both PNA and peptides can be synthesised by FMOC chemistry. Examples of such probes are disclosed in Thompson et al. (2007, Nucleic Acids Res. **35**(4): e28).

To allow more than one tag to be incorporated per oligonucleotide, mass tags can be incorporated into the oligonucleotide through conjugation to thymidine analogues, for example, as disclosed in Brown et al. (2001, Tetrahedron Lett. **42**: 2587-2592). In that publication, a thymidine analogue is described with a linker coupled to the purine ring of the thymidine. This thymidine analogue has a hydroxyl group protected with an FMOC group on the end of the linker that can be made available after the nucleotide has been coupled into an oligonucleotide during automated oligonucleotide synthesis to allow a phosphoramidite modified tag to be incorporated into an oligonucleotide. Since this analogue can be incorporated within the chain, multiple linkers and hence tags can be couple to the oligonucleotide.

Alternatively, branched peptides can be synthesised incorporating multiple tag peptide sequences. The branched peptides can be synthesised on CPG, with subsequent synthesis of the oligonucleotide probe (see Haralambidis et al., 1990, Nucleic Acids Res. **18**(3): 501-505).

Examples of schemes according to this embodiment of the invention are shown in Figs 4 and 5. In Fig. 4(a), it can be seen that tag oligonucleotide conjugates according to this invention are prepared by initially synthesising the peptide tag component of the conjugate on a controlled pore glass (CPG) resin that is compatible with automated oligonucleotide synthesis. In step (1) of Fig. 4(a), aminopropyl CPG is initially coupled to the aforementioned 4-Hydroxymethylbenzoic acid (HMBA) linker. A spacer could be introduced between the aminopropyl glass and the linker of desired or a CPG resin that has a longer linker could be employed. Peptide synthesis is initiated from the HMBA linker. In step (2) of the method in Fig. 4(a), an N-alpha-Fmoc-N-epsilon-1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl-L-lysine (α-FMOC-ε-Dde-Lys) residue is introduced. Spacers, such as beta-alanine could be introduced between the lysine and the HMBA linker if desired to reduce potential steric interference from the resin. The Dde protection group is removable under 2% hydrazine in Dimethylformamide (DMF) (Bycroft et al., 1993, J. Chem. Soc., Chem. Commun. 778-779) and is used for the purpose of synthesising branched peptides. In step (3) of Fig. 4(a), a further α-FMOC-ε-Dde-Lys residue is introduced, followed by an α-FMOC-O-trityl protected serine residue. The final FMOC group is removed and the serine amino group is acetylated to prevent further reaction. In

step (4), the Dde protection groups are removed. In step (5), the desired peptide tag is synthesised, typically using multiple cycles of peptide synthesis, from both of the free epsilon amino groups exposed by the removal of the Dde groups using commercially available FMOC-amino acids. In this example, an FMOC protected photocleavable linker (4-[4-(1-(FMOCamino)ethyl)-2-methoxy-5-nitrophenoxy)butanoic acid available from IRIS Biotech GmbH, Marktredwitz, Germany) is introduced followed by FMOC-alanine, followed by 1-Fmoc-piperidin-4-ylacetic acid (Sigma Aldrich, UK), followed again by FMOC-alanine, which is finally acetylated to block the terminal amino group so that it does not interfere in subsequent oligonucleotide synthesis. Between step (5) and step (6), the CPG resin is transferred from the automated peptide synthesiser reaction vessel or column into a column or vessel suitable for use in an automated DNA synthesiser. In step (6) of Fig. 4(a), the trityl protection group is removed using standard deprotection conditions in the DNA synthesiser, typically trichloroacetic acid or dichloroacetic acid in dichloromethane.

In step (7) shown in Fig. 4(b), an optional Dimethoxtrityl (DMTr) protected linker/spacer is introduced to reduce steric hindrance from the solid support and the peptide (here, Spacer Phosphoramidite 9 from Glen Research, Stirling, VA, USA). A wide variety of similar spacers is known in the art and could be substituted for the linker shown, if desired. In step (8), multiple cycles of standard automated (phosphoramidite) DNA synthesis are carried out to generate the desired DNA sequence linked to the peptide. In the final step (9), the side chain deprotection groups are cleaved along with the HMBA linker, typically using NH_3/H_2O steps, releasing the deprotected conjugate into solution. The released conjugate would usually then be purified by high performance liquid chromatography, gel filtration, gel electrophoresis or other standard techniques known in the art.

The number of peptides linked to a single oligonucleotide can be varied by varying the number of α-FMOC-ε-Dde-Lysine residues introduced into the peptide in the first steps of the synthesis shown in Fig. 4. If only a single tag is desired then only one α-FMOC-ε-Dde-Lysine residue need be coupled, but if 5 tags were desired this can be achieved by coupling 5 α-FMOC-ε-Dde-Lysine residues.

In principle, the order of certain steps can be varied as well. The introduction of the optional spacer in step 7 of Fig. 4 could take place prior to removal of the Dde groups. This

may be desirable if many peptides or if large peptides are to be coupled to a single oligonucleotide causing steric hindrance that might reduce the efficiency of the introduction the linker.

In further variations on the method in Fig. 4, α-Dde-ε-FMOC-Lysine can be employed in place of α-FMOC-ε-Dde-Lysine. In one variation, the O-trityl serine needed to initiate oligonucleotide synthesis can be coupled at the epsilon amino while the peptide tag is synthesised at the alpha position after removal of the Dde group. Alternatively, the alpha-Dde group can be removed prior to the epsilon-FMOC group, and the O-trityl serine can be coupled at this position. Optional spacers, such as beta-alanine residues can be introduced before coupling of the O-trityl protected serine, if desired.

In the example shown in Fig. 5, it can be seen that tag oligonucleotide conjugates according to a further embodiment of the present invention are prepared by initially synthesising the peptide tag component of the conjugate on a controlled pore glass (CPG) resin that is compatible with automated oligonucleotide synthesis. In step (1) of Fig. 5, aminopropyl CPG is initially coupled to an HMBA linker. Peptide synthesis is initiated from this linker. In step (2) of the method in Fig. 5, an α-FMOC-O-trityl protected serine residue is coupled to the HMBA linker. Spacers, such as beta-alanine, could be introduced between the serine residue and the HMBA linker if desired to reduce potential steric interference from the resin. In step (3), an α-FMOC-ε-FMOC-protected Lysine residue is introduced and coupled to the serine residue. In step (4), the desired peptide tag is synthesised from both of the free amino groups exposed by the removal of the FMOC groups on the lysine amino groups. The peptide tags are synthesised using commercially available FMOC-amino acids. In this example, an FMOC protected photocleavable linker (4-[4-(1-(FMOC-amino)ethyl)-2-methoxy-5-nitrophenoxy)butanoic acid available from IRIS Biotech GmbH, Marktredwitz, Germany) is introduced followed by FMOC-alanine, followed by 1-Fmoc-piperidin-4-ylacetic acid (Sigma Aldrich, UK), followed again by FMOC-alanine, which is finally acetylated to block the terminal amino group so that it does not interfere in subsequent oligonucleotide synthesis. Between step (4) and step (5), the CPG resin is transferred from the automated peptide synthesiser reaction vessel or column into a column or vessel suitable for use in an automated DNA synthesiser. In step (5) of Fig. 5, the trityl protection group is removed using standard deprotection conditions in the DNA

synthesiser, typically trichloroacetic acid or dichloroacetic acid in dichloromethane. In step (6), an optional Dimethoxtrityl protected linker is introduced to reduce steric hindrance from the solid support and the peptide (Spacer Phosphoramidite 9 from Glen Research, Stirling, VA, USA). A wide variety of similar spacers is known in the art and could be substituted for the linker shown, if desired. In step (7), multiple cycles of standard automated (phosphoramidite) DNA synthesis are carried out to generate the desired DNA sequence linked to the peptide. In the final step (8), the side chain deprotection groups are cleaved along with the HMBA linker, releasing the deprotected conjugate into solution. The released conjugate would usually then be purified by high performance liquid chromatography, gel filtration, gel electrophoresis or other standard techniques known in the art.

The number of peptides linked to a single oligonucleotide can be varied by varying the number of α-FMOC-ε-FMOC-Lysine residues introduced into the peptide in the first steps of the synthesis. After the introduction of the first α-FMOC-ε-FMOC-Lysine, removal of the FMOC groups will expose two amino groups. This means in the next cycle of synthesis 2 further α-FMOC-ε-FMOC-Lysine groups could be introduced and four in the next cycle if desired. Three cycles of coupling of α-FMOC-ε-FMOC-Lysine would make 8 amino groups available for the synthesis of peptide tags.

Hybrid methods employing the strategies of Figs 3 to 5 are envisaged. For example, steps (1) to (4) of Fig. 3 could be completed, i.e. up to the removal of the Dde groups, then α-FMOC-ε-FMOC lysine residues could be coupled to the amino groups exposed by removal of the Dde groups allowing further branching to be introduced if desired.

Several photocleavable linkers are commercially available. For example, the FMOC protected linker 4-[4-(1-(FMOCamino)ethyl)-2-methoxy-5-nitrophenoxy)butanoic acid is commercially available from Iris Biotech GmbH (Marktredwitz, Germany) and from Advanced ChemTech, Inc (Kentucky, USA). This can be readily incorporated into peptides during conventional FMOC synthesis.

Branched peptides (or other polyamides), with or without attached oligonucleotides, and comprising a plurality of cleavable branches form another aspect of the invention.

Non-natural amino acids may be used in this invention, including FMOC-Piperazin-1-ylacetate, 1-Fmoc-piperidin-4-ylacetic acid, N,N-Dimethyl glycine, β-Dimethylamino-DL-alanine (all commercially available from Sigma Aldrich, UK).

In addition to the tags shown in Figs 4 and 5, exemplar peptide sequences of the invention include:

Series 1:
Alanine – Piperazin-1-ylacetic acid – Alanine,
Alanine – Piperazin-1-ylacetic acid – Valine,
Valine – Piperazin-1-ylacetic acid – Alanine,
Valine – Piperazin-1-ylacetic acid – Valine;

Series 2:
N,N-Dimethyl glycine - Alanine – Proline – Alanine (SEQ ID NO: 1),
N,N-Dimethyl glycine - Alanine – Proline – Valine (SEQ ID NO: 2),
N,N-Dimethyl glycine - Valine – Proline – Alanine (SEQ ID NO: 3),
N,N-Dimethyl glycine - Valine – Proline – Valine (SEQ ID NO: 4); and

Series 3:
Cysteic Acid - Alanine – Aspartic acid – Alanine (SEQ ID NO: 5),
Cysteic Acid - Alanine – Aspartic acid – Valine (SEQ ID NO: 6),
Cysteic Acid - Valine – Aspartic acid – Alanine (SEQ ID NO: 7),
Cysteic Acid - Valine – Aspartic acid – Valine (SEQ ID NO: 8).

In the series 1 peptides, it can be seen that the collision cleavable linker (Piperazin-1-ylacetic acid) is also a charge-carrying group due to the presence of the tertiary amino group in the piperazine ring. In series 2 and 3 peptides, the charge-carrying groups are separate entities, i.e. N,N-Dimethyl glycine and cysteic acid, respectively. In alternative embodiments, the charge-carrying groups could act as the mass modifier as well, allowing the adjacent mass modifier amino acid group to be removed resulting in a smaller tag if that were desirable.

In the series 2 peptides, the proline residue can be substituted with piperidin-4-ylacetic acid. In series 2, the N,N-Dimethyl glycine can be replaced with any easily protonated positive charge-carrying group including the natural amino acids lysine and histidine, tertiary amino group containing molecules such as β-Dimethylamino-alanine. Secondary amino containing groups may also be introduced such as nipecotic acid. Pyridine containing compounds such as nicotinic acid are also appropriate.

Fig. 6a illustrates a mass tagged oligonucleotide probe based on the peptide, (N) - Acetate - Valine – Piperazin-1-ylacetic acid – Alanine – Beta-alanine – (C) [SEQ ID NO: 9]. This structure is another variant of the series-1 peptides shown above. The mass tag fragment released from SEQ ID NO: 9 by photocleavage ("P") is also shown.

Fig. 6b illustrates a set of 9 isobaric tags (with $[M+H]^+$ = 437.3) constructed by combinatorial rearrangement of the mass modifier amino acids, beta-alanine, alanine and valine, and isotopes of acetic acid in the tags. All of the isotope-substituted reagents are commercially available (for example from Cambridge Isotope Laboratories, Inc; Andover, MA, USA).

Large numbers of different sets of tags can be constructed by changing the amino acids and other entities used as mass modifiers. The acetate group blocking the N-terminal amino group of the peptide tags could be replaced with any carboxylic acid, e.g. propionic or butyric acids, which can be obtained with isotope substitutions (for example from Cambridge Isotope Laboratories, Inc; Andover, MA, USA). Similarly, further isobaric compounds can be obtained by swapping the positions of the alanine and valine residues, for example. Synthesis of every combination of isotope substituted amino acids shown in this sequence would give rise to hundreds of different tags, although not all would be isobaric with each other. If every possible, isotopic variant of the sequence shown in Fig. 6b was synthesised, the tags would fall into a series of isobaric sets. Although these variants are too numerous to enumerate here, they are within the scope of the present invention.

In all these series the mass modifiers comprise individual amino acids of valine or alanine or their isotopes. However, other amino acids can be used for this purpose. In addition, more than one amino acid can be inserted into the sequence to act as mass modifiers. To

ensure cleavage is most favoured at a particular site in the sequence, beta-amino acids can also be used to act as mass modifiers as these do not readily support formation of oxazolone structures and thus do not cleave as easily as alpha amino acids at any given collision energy. In Fig. 6b the combination of alanine and beta-alanine is used to act as a mass modifier. Similarly, the combination of valine and acetate are also used to act as a single mass modifier in Fig. 6b.

In an alternative embodiment, an acid cleavable linker can be used. Since most MALDI matrix materials are acidic, addition of the matrix will effect cleavage of the mass tags. A simple method for introducing an acid labile group is to include a P3'-N5' phosphoramidate at the 5' terminus of the oligonucleotide adjacent to the mass tag (Shchepinov et al., 2001, Nucleic Acids Res. **29**(18): 3864-72).

In a further embodiment, the entire probe label complex can be desorbed, and cleavage of the tags can take place by collision using Post Source Decay in a Time-Of-Flight mass spectrometer or in the mass analyser of an ion trap instrument or in a collision cell in alternative geometries that are used with MALDI, such as the Q-TOF geometry. PNA probes with a collision cleavable linker between the tag peptide and a PNA probe have been described previously (Thompson et al., 2007, Nucleic Acids Res. **35**(4): e28).

In an alternative embodiment, branched peptides may be coupled to one or more oligonucleotides using "click chemistry". A popular and versatile "click reaction" is 1,3-cyclo-addition of azides with terminal acetylenes using a Copper catalyst at room temperature. 5'-Hexynyl Phosphoramidite is commercially available from Glen Research Corporation (Sterling, Virginia, USA), which permits the introduction of an alkyne into the 5'-terminus of an oligonucleotide. Similarly, Azidobutyrate N-HydroxySuccinimide Ester is also available from Glen Research. This can be used to introduce an azide group into a peptide with a free amino group. Alternatively, Fmoc-propargyl-Glycine is available from Sigma Aldrich allowing an alkyne function to be introduced into a peptide by FMOC synthesis. In this situation, Azidobutyrate N-HydroxySuccinimide Ester can be used to introduce and azide group into an amine-modified oligonucleotide. Multiple alkynyl groups can also be incorporated into oligonucleotides at internal positions allowing multiple conjugation of azide functionalised labels (Gierlich et al., 2006, Org Lett. **8**(17): 3639-42).

For the present invention, a branched peptide synthesised according to the methods shown in Figs 4 or 5 could be produced but the trityl-serine can be omitted and the terminal amino group of the peptide can be coupled with Azidobutyrate N-HydroxySuccinimide Ester. Azide derivatised branched peptides could then be coupled to oligonucleotides with multiple alkyne groups to give a highly labeled conjugate.

A branched peptide comprising or consisting of multiple tags according to the present invention can also be conjugated to an oligonucleotide in a non-covalent manner, for example by employing biotin-avidin binding. Biotin can be introduced into a branched peptide by standard methods known in the art. In particular, a biotinylated lysine residue may be incorporated at any position in a peptide during standard FMOC synthesis using N-α-Fmoc-N-ε-biotinyl-L-lysine, which is commercially available (for example from Merck Chemicals Ltd, Nottingham, UK). With reference to the methods shown in Figs 3 to 5, the serine residue used for the coupling of an oligonucleotide could be readily replaced with biotinylated lysine. A variety of avidin counter-ligands for biotin are available, including monomeric and tetrameric avidin and streptavidin.

Methods for conjugating oligonucleotides to antibodies using avidin are known in the art and these can be adapted to the conjugation of branched peptides to oligonucleotides. DNA-peptide conjugates can be prepared by conjugating oligonucleotides to streptavidin. These streptavidin-oligonucleotide conjugates can then be linked to biotinylated branched peptides in the same manner as previously described for antibodies (Niemeyer et al., 2003, Nucleic Acids Res. **31**(16): e90). Alternatively, antibodies can be conjugated to avidin and biotinylated oligonucleotides can be linked to the avidin complex (Sano et al., 1992, Science. **258**(5079): 120-2).

Similarly, a branched peptide according to the invention can be conjugated to a peptide or protein (for example, an affinity ligand such as an antibody, or a functional fragment thereof) by chemically reacting the branched peptide with avidin (Iwai et al., 1988, Anal. Biochem. **171**: 277-282). Branched peptides comprising mass tags of the present invention with a free cysteine thiol group can be conjugated to amino groups on avidin or an antibody using the reagent SMPB (Succinimidyl 4-[p-maleimidophenyl] butyrate) as described by Iwai et al. (1998, above). A cysteine with a free thiol can be introduced into

the peptide shown in Figs 3 to 5 to replace the trityl-serine residue using standard FMOC synthesis methods.

Multimeric Oligonucleotides

In some embodiments, signals from a bound probe molecule may be enhanced by incorporating multiple tags into the probe molecule. This can be done by coupling more than one tag to each probe molecule, for example multiple peptide tags as described above.

Additionally or alternatively, the probe molecule may comprise a multimeric oligonucleotide to which one or more tags (for example, peptide tags) are attached. Suitable examples of multimeric oligonucleotides include a branched oligonucleotide, a comb oligonucleotide and a dendrimeric oligonucleotide.

Oligonucleotide units of the multimeric oligonucleotide may be covalently linked directly to each other through phosphodiester bonds or through interposed linking agents such as nucleic acid, amino acid, carbohydrate or polyol bridges, or through other cross-linking agents that are capable of cross-linking nucleic acid or modified nucleic acid strands. The site(s) of linkage may be at the ends of the unit (in either normal 3'-5' orientation or randomly oriented) and/or at one or more internal nucleotides in the strand.

In linear multimeric oligonucleotides, individual units are linked end-to-end to form a linear polymer. A branched oligonucleotide (for example, bDNA) may comprise two, three or more oligonucleotide units emanating from a point of origin to form a branched structure. The point of origin may be another oligonucleotide unit or a multifunctional molecule to which at least two oligonucleotide units can be covalently bound (see for example US5,124,246 and Horn et al., 1997b, Nucleic Acids Res. **25**: 4842-4849). A comb (or "fork") oligonucleotide may comprise an oligonucleotide unit backbone with one or more pendant oligonucleotide units. The pendant units usually depend from a modified nucleotide or other organic moiety having appropriate functional groups to which oligonucleotides may be conjugated or otherwise attached (see for example Horn et al., 1997a, Nucleic Acids Res. **25**: 4835-4841). A dendrimeric oligonucleotide comprises non-oligonucleotide components to which oligonucleotides of the same or different sequences

are attached (see for example Shchepinov et al., 1997, Nucleic Acids Res. **25**: 4447-4454 and Shchepinov et al., 1999, Nucleic Acids Res. **27**: 3035-3041).

The multimeric oligonucleotide for use in the invention may be linear, branched, or comprise a combination of linear and branched portions. There may be at least two branch points in the multimer, for example at least 3, preferably 5 to 10. The multimer may include one or more segments of double-stranded sequences.

A probe molecule comprising a multimeric oligonucleotide may include a sequence complementary to a target linked to two or more "address" sequences to which a secondary "address complement" probe may be designed to bind. This address complement probe would be coupled to a mass tag according to this invention. The probe molecule and address sequences may be in the form of a Y-shaped oligonucleotide of a structure described by Suzuki et al. (2000, Nucleic Acids Symp Ser. **44**: 125-126). Alternatively, a single long probe sequence comprising a target recognition sequence and multiple address sequences may be used. Methods for the use of branched and comb oligonucleotide probes for in situ hybridisation (see Player et al., 2001, J Histochem Cytochem. **49**(5): 603-12; Kenny et al., 2002, J Histochem Cytochem. **50**(9): 1219-27) may be adapted for use in the present invention.

Oligonucleotide-tagged Probes

In certain embodiments of the invention, a polypeptide may be detected by binding of an oligonucleotide-labelled ligand. Examples of such ligands include a nucleic acid aptamer, a peptide, an antibody and a carboydrate. The binding of the oligonucleotide-labelled ligand may be detected by hybridisation of an oligonucleotide labelled with one or more mass tags of this invention. An oligonucleotide-labelled ligand in the form of an oligonucleotide-peptide conjugate used for target binding can be produced in the same way as mass tagged oligonucleotides described above.

Aptamers are folded nucleic acid ligands with high binding affinities for non-nucleic acid targets. Aptamers have a variety of advantages for use as ligands for the detection of a variety of different molecular targets: they can be synthesised in conventional oligonucloetide synthesisers or they can be generated by PCR, they can be evolved to higher levels of specificity and they can easily by linked to additional sequences for

detection by secondary, labelled oligonucleotides or dendrimers (Stoltenburg et al., 2005, Anal Bioanal Chem. **383**(1): 83-91; Tombelli et al., 2004, Biosens Bioelectron. **20**(12): 2424-34).

An oligonucleotide-labelled ligand in the form of an antibody-oligonucleotide conjugate may be produced by methods known in the art. Intein Thioesters are useful means for coupling cysteine-derivatised oligonucleotides to polypeptides. Protein sequences of interest, such as antibody clone libraries, can be cloned into plasmids containing intein sequences. The resulting intein containing hybrid structures reacts with free thiols on an oligonucleotide in a mild reaction that is highly site-specific (Lovrinovic et al., 2005, Mol Biosyst **1**(1): 64-69). A variety of bifunctional reactive linkers are also available, for example from Pierce (Pittsburgh, PA), and can be used to cross-link various functional groups. Pierce provides linkers for amino group to thiol coupling, amino to amino coupling and thiol to thiol coupling.

Conventional chemical coupling of oligonucleotides to antibodies is also known in the art. An antibody may be activated with maleimide and coupled to thiol derivatised oligonucleotides (Hendrickson et al., 1995, Nucleic Acids Res. **23**(3): 522-9). In another approach, an aldehyde-modified antibody may be coupled to a hydrazide modified oligonucleotide (Kozlov et al., 2004, Biopolymers. **73**(5): 621-30).

An oligonucleotide-antibody conjugate may alternatively be prepared by initially conjugating an oligonucleotide to streptavidin. This streptavidin-oligonucleotide conjugate can then be linked to a biotinylated antibody (Niemeyer et al, 2003, Nucleic Acids Res. **31**(16): e90). Alternatively, an antibody can be conjugated to avidin, and a biotinylated oligonucleotide can be linked to the avidin complex (Sano et al., 1992, Science. **258**(5079): 120-2).

It has been shown that oligonucleotide-antibody conjugates retain both the ability of the antibody to bind to its target and the ability of the oligonucleotide to base-pair normally with a complementary oligonucleotide (Kuijpers et al., 1993, Bioconjug Chem. **4**(1): 94-102) and that these conjugates will bind to their targets and tissue allowing for detection following hybridisation of a labelled complementary oligonucleotide (Bos et al., 1994, Cancer Res. **54**(13): 3479-86). This work employed radiolabelled oligonucleotides for the

detection step but oligonucleotide-label ligands of this invention will be equally effective for detection of bound antibody conjugates.

MS/MS and MS^n analysis of Peptide Tandem Mass Spectrometric Tags

Tandem mass spectrometers allow ions with a predetermined mass-to-charge ratio to be selected and fragmented by collision induced dissociation (CID). The fragments can then be detected providing structural information about the selected ion. When peptides are analysed by CID in a tandem mass spectrometer, characteristic cleavage patterns are observed, which allow the sequence of the peptide to be determined. Natural peptides typically fragment randomly at the amide bonds of the peptide backbone to give series of ions that are characteristic of the peptide. CID fragment series are denoted a_n, b_n, c_n, etc. for cleavage at the nth peptide bond where the charge of the ion is retained on the N-terminal fragment of the ion. Similarly, fragment series are denoted x_n, y_n, z_n, etc. where the charge is retained on the C-terminal fragment of the ion.

Trypsin and thrombin are favoured cleavage agents for tandem mass spectrometry as they produce peptides with basic groups at both ends of the molecule, i.e. the alpha-amino group at the N-terminus and lysine or arginine side-chains at the C-terminus. This favours the formation of doubly charged ions, in which the charged centres are at opposite termini of the molecule. These doubly charged ions produce both C-terminal and N-terminal ion series after CID. This assists in determining the sequence of the peptide. Generally speaking only one or two of the possible ion series are observed in the CID spectra of a given peptide. In low-energy collisions typical of quadrupole based instruments the b-series of N-terminal fragments or the y-series of C-terminal fragments predominate. If doubly charged ions are analysed then both series are often detected. In general, the y-series ions predominate over the b-series.

In general, peptides fragment via a mechanism that involves protonation of the amide backbone follow by intramolecular nucleophilic attack leading to the formation of a 5-membered oxazolone structure and cleavage of the amide linkage that was protonated (Schlosser & Lehmann, 2000, Mass Spectrom. **35**: 1382-1390). Fig. 9a shows one proposed mechanism by which this sort of fragmentation takes place. This mechanism requires a carbonyl group from an amide bond adjacent to a protonated amide on the N-terminal side of the protonated amide to carry out the nucleophilic attack. A charged

oxazolonium ion gives rise to b-series ions, while proton transfer from the N-terminal fragment to the C-terminal fragment gives rise to y-series ions as shown in Fig. 9a. This requirement for an appropriately located carbonyl group does not account for cleavage at amide bonds adjacent to the N-terminal amino acid, when the N-terminus is not protected and, in general, b-series ions are not seen for the amide between the N-terminal and second amino acid in a peptide. However, peptides with acetylated N-termini do meet the structural requirements of this mechanism and fragmentation can take place at the amide bond immediately after the first amino acid by this mechanism. Peptides with thioacetylated N-termini will cleave particularly easily by the oxazolone mechanism as the sulphur atom is more nucleophilic than an oxygen atom in the same position.

Fragmentation of the amide backbone of a peptide can also be modulated by methylation of the backbone. Methylation of an amide nitrogen in a peptide can promote fragmentation of the next amide bond C-terminal to the methylated amide and also favours the formation of b-ions. The enhanced fragmentation may be partly due to the electron donating effect of the methyl group increasing the nucleophilicity of the carbonyl group of the methylated amide, while the enhanced formation of b-ions may be a result of the inability of the oxazolonium ion that forms to transfer protons to the C-terminal fragment as shown in Fig. 9b. In the context of this invention thioacetylation of the N-terminus of a tag peptide can be used to enhance cleavage of the tag peptide at the next amide bond. Similarly, methylation of the nitrogen atom of an N-terminal acetyl or thioacetyl group will also enhance cleavage of the adjacent amide bond.

The ease of fragmentation of the amide backbone of a polypeptide or peptide is also significantly modulated by the side chain functionalities of the peptide. Thus the sequence of a peptide determines where it will fragment most easily. In general it is difficult to predict which amide bonds will fragment easily in a peptide sequence. This has important consequences for the design of the peptide mass tags of this invention. However, certain observations have been made that allow peptide mass tags that fragment at the desired amide bond to be designed. Proline, for example, is known to promote fragmentation at its N-terminal amide bond (Schwartz et al., 1997, Biol. Mass Spectrom. **21**: 92) as fragmentation at the C-terminal amide gives rise to an energetically unfavourable strained bicyclic oxazolone structure. Aspartic acid also promotes fragmentation at its N-terminal amide bond. Asp-Pro linkages, however, are particularly labile in low energy CID analysis

(Wysocki et al., 2000, J Mass Spectrom. **35**(12): 1399-1406) and in this situation aspartic acid seems to promote the cleavage of the amide bond on its C-terminal side. Thus proline, and asp-pro linkages can also be used in the tag peptides of this invention to promote fragmentation at specified locations within a peptide.

A typical tandem mass spectrometer geometry is a triple quadrupole, which comprises two quadrupole mass analysers separated by a collision chamber, also a quadrupole. This collision quadrupole acts as an ion guide between the two mass analyser quadrupoles. A gas can be introduced into the collision quadrupole to allow collision with the ion stream from the first mass analyser. The first mass analyser selects ions on the basis of their mass/charge ratio which pass through the collision cell where they fragment. The fragment ions are separated and detected in the third quadrupole. Induced cleavage can be performed in geometries other than tandem analysers. Ion trap mass spectrometers can promote fragmentation through introduction of a gas into the trap itself with which trapped ions will collide. Ion traps generally contain a bath gas, such as helium but addition of neon for example, promotes fragmentation. Similarly photon induced fragmentation could be applied to trapped ions. Another favourable geometry is a Quadrupole/Orthogonal Time of Flight tandem instrument where the high scanning rate of a quadrupole is coupled to the greater sensitivity of a reflectron TOF mass analyser to identify the products of fragmentation.

Conventional "sector" instruments are another common geometry used in tandem mass spectrometry. A sector mass analyser comprises two separate "sectors", an electric sector which focuses an ion beam leaving a source into a stream of ions with the same kinetic energy using electric fields. The magnetic sector separates the ions on the basis of their mass to generate a spectrum at a detector. For tandem mass spectrometry, a two sector mass analyser of this kind can be used where the electric sector provide the first mass analyser stage, the magnetic sector provides the second mass analyser, with a collision cell placed between the two sectors. Two complete sector mass analysers separated by a collision cell can also be used for analysis of mass tagged peptides.

MALDI matrix dyes

Various compounds have been found to be useful as matrices (also referred to herein as matrix materials) for MALDI analysis of large biomolecules. Such compounds are generally

characterised by a number of properties. The compounds generally have a strong extinction coefficient at the frequency of the laser used for desorption. The compounds are also able to isolate analyte molecules in a solid solution and the compounds are sufficiently volatile to rapidly sublime when exposed to laser shots in the MALDI mass spectrometer. The subliming dye should vaporise rapidly in a jet that entrains the embedded analyte molecules and for most purposes this should take place without fragmentation of the analyte (although fragmentation may sometimes be desirable if structural information about the analyte is sought). However, a matrix should not be too volatile as experiments can sometimes take several hours and the analyte/matrix co-crystal must remain stable under vacuum in the ion source for this period of time. The property of volatility to laser irradiation can be measured approximately by determining the initial velocity of analyte ions generated by the matrix. It has been observed that higher initial velocities correspond to "softer" ionisation, i.e. reduced fragmentation, (Karas & Glückmann, 1999, J Mass Spectrom. **34**: 467–477) but high initial ion velocities of some matrices also correlates to rapid sublimation under vacuum.

Different matrices have different properties in terms of their ability to assist the desorption of embedded analytes and in the subsequent sensitivity with which the analytes are detected. It has been found empirically that certain matrices are more appropriate for the analysis of particular analytes than others. For example, 3-hydroxypicolinic acid has been found to be most effective for analysing oligonucleotides (Wu et al., 1993, Rapid Commun. Mass Spectrom. **7**:142-146), while 2,5-dihydroxybenzoic acid and 4-hydroxy -alpha-cyano-cinnamic acid (HCCA) are both most effective for the analysis of peptides and proteins (Strupat et al., 1991, Int. J. Mass Spectrom. Ion Proc. **111**: 89 – 102; Beavis et al., 1992, Org. Mass Spectrom. **27**: 156–158). Various cinnamic acid derivatives have been found to be effective for the analysis of proteins (Beavis & Chait, 1989, Rapid Commun Mass Spectrom. **3**(12): 432-435) and the choice of matrix is dependant on the nature of the analyte, for example sinnapinic acid may be preferred over HCCA for large peptides and polypeptides, while HCCA is generally preferred for smaller peptides. 2,5-dihydroxybenzoic acid may produce less fragmentation than the cinnamic acid derivatives in some cases.

Most of the matrices described above are acidic matrices. Basic matrices have also been developed and may be more appropriate for the analysis of acid-sensitive compounds (Fitzgerald et al., 1993, Anal Chem. **65**(22): 3204-3211).

Infrared MALDI (IR-MALDI) is similar in principal to ultraviolet MALDI (UV-MALDI) in that analytes must be embedded in a matrix that preferably has a strong extinction coefficient at the frequency of the laser in the desorption instrument. Appropriate matrices tend to be different compounds from those used in UV-MALDI and liquid matrices are often used. Glycerol, urea, ice and succinic acid have all been shown to be effective matrices for IR-MALDI (Talrose et al., 1993, Rapid Commun Mass Spectrom. **13**(21): 2191-2198). However, some UV-MALDI matrices, such as cinnamic acid derivatives, also work as IR matrices (Niu et al., 1998, J Am. Soc. Mass Spectrom. **9**: 1-7).

Liquid matrices for UV-MALDI have also been explored (Ring & Rudich, 2000, Rapid Commun Mass Spectrom. **14**(6): 515-519; Sze et al., 1998, J Am Soc Mass Spectrom. **9**(2): 166-174; Karas et al., 1991, Mass Spectrom Rev **10**: 335). The simplest examples of liquid matrices comprise solutions of the matrices used as solids. True liquid matrices are also known such as nitrobenzoyl alcohol. Both types of matrix have some advantages in terms of sample consistency, stability under vacuum and ease of handling, although solid matrices tend to be more sensitive. In the context of the present invention, the improvements in sensitivity may justify the use of liquid matrices. This may have particular advantages in the automation of sample preparation, as liquid handling robotics are widely available and the use of solutions of matrices, for solid matrix co-crystallisation, which readily clog dispensing devices can be avoided.

Matrix Application

Various sample preparation methods can be used to obtain signals from biological samples: rinsing the sample in saturated dihydroxybenzoic acid dissolved in highly purified water, or coating samples by spraying a solution of hydroxy-alpha-cyano-cinnamic acid or laying a section of tissue on a matrix that has been pre-coated with a matrix. For a review see Caldwell & Caprioli (2005, Mol Cell Proteomics. **4**(4): 394-401). Acoustic spraying has been reported for coating tissue sections (Aerni et al., 2006, Anal Chem. **78**: 827-834). Robotic spotting of matrix has also been reported (Groseclose et al., 2007, J Mass Spectrom. **42**: 254-262).

Matrix can also be pre-coated onto targets (Miliotis et al., 2002, Rapid Commun Mass Spectrom. **16**(2): 117-26). Tissue sections can then be laid directly onto the pre-coated targets for analysis.

Calibrants

MALDI spectra can vary quite considerably in terms of intensity, making comparison of quantities between spectra challenging. To provide accurate maps of the intensity of a particular tag ion and thus a map of the binding of the corresponding probe, it is important to be able to make quantitative comparisons between tag ion spectra from different locations from a tissue section. To make these comparisons, the intensities of the tag ions in spectra under comparison may be normalised. Normalisation may be carried out by incorporating a known quantity of a tag or tagged probe into the sample.

In one embodiment, a tag or tagged probe for normalisation is included at a known concentration in the matrix applied to the tissue section. Alternatively or in addition, a known quantity of a spike target can be added to the tissue sample.

In some embodiments, it may be desirable to use more than one type of calibrant with the same specimen or sample.

When spectra are obtained according to the method of this invention, the intensity of the peaks of the calibrants can be used to normalise the spectra. In the simplest case, where a single calibrant ion is dissolved in the matrix, normalisation can be achieved by dividing all of the intensities of the other tag ions by the intensity of the calibrant tag ion.

With more than one calibrant ion, calibration curves can be plotted to adjust ion intensities over a wide range of intensities.

Signal Processing

Although the methods of this invention can be practiced on any MS/MS capable instrument, the fastest MS/MS instruments are likely to be based on Time-of-Flight (TOF) technology, such as TOF/TOF, Trap/TOF and Quadrupole/TOF geometries. Speed is important to allow decent resolution images to be obtained in a reasonable amount of time.

We now describe some of the issues in analysis of TOF data. Processing of Time-Of-Flight data is usually performed by software provided by the manufacturer of the instrument, e.g. the MassLynx software provided by Micromass (Manchester, UK) to operate their ESI-TOF and Q-TOF instrumentation. It is, however, sometimes preferable to be able to process the data directly and the general steps necessary to process TOF data to render it compatible with the methods of this invention are described below. For a review of some of the standard digital signal processing techniques described below, see for example Smith, S. W. (1997, "The Scientist and Engineer's Guide to Digital Signal Processing", California Technical Publishing).

Raw data from the TOF is obtained in the form of digital signals from an analogue to digital converter that receives ion arrival times from the detector. The first step is to convert the ion arrival time data into ion mass-to-charge ratio data using a calibration file. Typically the digital spectra from the TOF mass analyser are contaminated by low levels of random noise. Preferably, this noise is removed prior to further analysis. Various methods of removing noise are applicable. In general the noise levels are very low compared to the ion signals. The simplest noise elimination method, therefore, is to set a threshold intensity below which the signal will ignored (or removed). However, the noise level for a Time-Of-Flight mass analyser is found to vary as the mass-to-charge ratio increases so it is better to apply a varying threshold for different mass-to-charge ratios. A standard threshold function could be determined for a given instrument relating noise to the mass-to-charge ratio and this could be used to eliminate signals below the threshold level of intensity. A preferred method, however, would be to make a data-dependant noise-estimation for different mass-to-charge ratios for each spectrum, as this allows random variations between analyses on a particular instrument to be accounted for and it makes the method independent of the instrument used. This can be done by splitting the raw spectrum into bins and estimating the noise in each bin. An interpolation or spline function describing an appropriate curve can then be fitted to the noise estimates for each bin to provide an adaptive threshold that varies over the full mass-to-charge ratio range of the spectrum. Signals below the calculated threshold are then removed from the spectrum.

After the random background noise has been removed the digital signal are smoothed prior to attempting to find ion peaks in the data. Smoothing can be achieved by various methods. Typically the digital mass spectrum data would be convoluted with a low

bandpass filter. A low bandpass filter generally smoothes a digital signal by effectively determining a moving average of the signal. This removes very high frequency signals from the data that correspond to small random variations in the digitised signal intensities for each ion. The digital signal can be convoluted with a number of different filter kernels that have a smoothing effect, such as a simple square function, which produces a modified spectrum in which a moving average has been applied where there is equal weighting to every point in the moving average. A preferred filter kernel applies a higher weighting to the central point in the moving average. Appropriate filter kernels include filters derived from a windowed sinc function, Blackman windows and Hamming windows. In a more preferred embodiment, the TOF spectrum is smoothed by convolution with a filter kernel derived from a Gaussian function.

Identification of peaks in a digital signal is essentially the same as for a continuous signal. With a continuous signal the first and second differentials of the signal are calculated; maxima and minima of the signal, i.e. peaks and troughs, are identified where the first differential is zero, while maxima are identified where the second differential is negative. For a discrete signal a Laplacian filter determines appropriate corresponding difference equations that facilitate detection of peaks in the digital signal.

In the context of this invention, the methods will be applied to the MS/MS spectra obtained from tag peptides. Once a list of peaks has been identified from the TOF data with their corresponding mass-to-charge ratios, the rapid analysis techniques described below can be applied to this list of peaks.

Rapid Analysis

MALDI TOF spectra are known for poor reproducibility in terms of peak intensities (Albrethsen, 2007, Clin Chem. **53**(5): 852-8). Individual spectra generated from single laser shots are usually insufficient for meaningful interpretation. Typically, a spectrum is produced by summing up between 10 and 100 individual laser shots. This means that to generate a one megapixel image from conventional spectra may require as many as 100 million laser shots, which would be extremely time-consuming.

In a further embodiment of this invention, a method of rapidly scanning a sample is provided. In the field of image encoding and compression for electronic transmission, one

method for compressing an image is to only record the points in the image where the image information changes. In simple terms, an image is defined as a series of areas of the same colour and a set of instructions, which define the relative positions and sizes of these different areas. A similar approach can be taken with imaging mass spectrometry for actually generating an image. The control system for the mass spectrometer can generate a full mass spectrum for the first spot on the tissue by acquiring and summing up 10 or 20 or other suitable number of laser shots. The laser can then be focused on the next spot and a laser shot can be acquired. In principle, the shot can be compared with the first spectrum to see if there is any change from the first spectrum. If no change in peak intensities is observed the laser can be focused onto the next location on the tissue section. This can be repeated until a change in intensities is observed, i.e. the point where the image changes, so that another full spectrum can be obtained. In practice, however, making this comparison is not trivial since MALDI TOF spectra are not particularly reproducible.

As described earlier, the methods and reagents of this invention provide a means to determine spectrum quality and to compare spectra and individual shots quantitatively. The quality of each shot (or spectrum produced by summing a number of shots) can be determined by using the intensities of calibrant ions. One method to compare the quality of shots or spectra is to determine whether the intensity of the calibrant ion is above a predetermined threshold. If more than two or more calibrant ions are added to the matrix or sample in a predetermined ratio, then a second method can be used to assess spectrum or shot quality: the ratio of the intensity of the first calibrant to the second can be used to determine whether the spectrum or shot is of a good quality. If there is a small or no deviation from the expected ratio then it is reasonable to assume that the spectrum is of a better quality than a spectrum where the determined ratio deviates from the expected ratio by a substantial amount. Thus the number of shots required from each location can be varied in accordance with a measurement of the quality of the summed spectrum.

An alternative approach is to perform a survey scan, i.e. survey spectra generated from 1 shot or summation of a small number of shots (such as 2 or 3) from each location, to rapidly generate a crude image. The data system of the mass spectrometer can then use the data from adjacent spectra to determine which locations on the sample need further shots to improve the image: adjacent spectra or shots can be compared to determine

whether any change has occurred in the quantities of targets that have been probed using the tagged probes of this invention. If no change (or only a small percentage change – pre-defined by the user) is detected then it can be assumed that the region of the tissue from which the image was obtained is homogeneous.

Alternatively, a survey scan can be performed and each shot can be assessed for quality using spiked calibrants as described above. Areas of the sample where the spectra give calibrant intensities below the predetermined threshold or, if more than one calibrant ion is used, areas of the sample where the measured ratios of calibrant ions deviate significantly from the expected ratios, can be sampled again until the spectra obtained from those areas are of a good quality. This method can be improved by acquiring high quality spectra, i.e. spectra with multiple shots (e.g. by summation of 10 to 100 shots), relatively frequently, for example, after every 10 or 20 locations where survey spectra are obtained.

The basis steps of a rapid method of imaging a planar sample in accordance with the invention comprises the following steps:

The planar sample is contacted with labelled probe molecules.

Unbound probes are washed from the planar sample.

The planar sample is coated with a matrix or laid on a surface, pre-coated with matrix where the matrix contains a known concentration of at least one tag spike or at least one tagged probe molecule spike.

A first location on the planar sample is sampled a number of times by a laser in a MALDI MS/MS mass spectrometer to generate a spectrum by repeating the following steps:

1. Desorb ions from the surface of the sample

2. Gate or trap desorbed ions to isolate ions with mass-to-charge ratios corresponding to tags of this invention.

3. Fragment the isolated tag ions by collision induced dissociation to release characteristic daughter ions that confirm the identity of the tags and allow relative quantification of the tags.

4. Record the intensity of the daughter ions or add the intensities of the daughter ions to the previously recorded intensities to build a composite spectrum

5. Compare the intensity of the spike ion with the known lowest concentration with background noise and determine whether the ratio of the spike ion signal to the background noise is above a predetermined threshold.

6. If the ratio of the spike ion signal to the background noise is above a predetermined threshold move to record the composite spectrum as final for the current location and select a new location to sample and repeat steps 1 to 7.

7.If the ratio of the spike ion signal to the background noise is not above a predetermined threshold repeat steps 1 to 5.

In one embodiment of this aspect of the invention, the background noise is taken as the mean of the intensity of all of the ions that are detected whose mass-to-charge ratios do not correspond to fragments from the tags of this invention.

Another rapid method of imaging a planar sample in accordance with the invention comprises the following steps:

The planar sample is contacted with labelled probe molecules.

Unbound probes are washed from the planar sample

The planar samples is coated with a matrix or laid on a surface, pre-coated with matrix where the matrix contains a known concentration of at least two tag spikes or at least two tagged probe molecule spikes.

A first location on the planar sample is sampled a number of times by a laser in a MALDI MS/MS mass spectrometer to generate a spectrum by repeating the following steps:

1. Desorb ions from the surface of the sample

2. Gate or trap desorbed ions to isolate ions with mass-to-charge ratios corresponding to tags of this invention.

3. Fragment the isolated tag ions by collision induced dissociation to release characteristic daughter ions that confirm the identity of the tags and allow relative quantification of the tags.

4. Record the intensity of the daughter ions or add the intensities of the daughter ions to the previously recorded intensities to build a composite spectrum.

5. Compare the intensity of the two spike ions with the known lowest concentrations with each other and determine the difference between the measured ratio of the spike ion signal and the expected ratio.

6. If the absolute value of the difference between the measured ratio of the spike ion signals and the expected ratio is less than a predetermined threshold move to record the composite spectrum as final for the current location and select a new location to sample and repeat steps 1 to 7.

7. If the absolute value of the difference between the measured ratio of the spike ion signals and the expected ratio is greater than the predetermined threshold repeat steps 1 to 5.

A further rapid method of imaging a planar sample according to the present invention comprises the following steps:

The planar sample is contacted with labelled probe molecules.

Unbound probes are washed from the planar sample

The planar samples is coated with a matrix or laid on a surface, pre-coated with matrix where the matrix contains a known concentration of at least two tag spikes according to this invention or at least two tagged probe molecule spikes.

A first location on the planar sample is sampled a number of times by a laser in a MALDI MS/MS mass spectrometer to generate a spectrum by repeating the following steps:

1. Desorb ions from the surface of the sample

2. Gate or trap desorbed ions to isolate ions with mass-to-charge ratios corresponding to tags of this invention.

3. Fragment the isolated tag ions by collision induced dissociation to release characteristic daughter ions that confirm the identity of the tags and allow relative quantification of the tags.

4. Record the intensity of the daughter ions or add the intensities of the daughter ions to the previously recorded intensities to build a composite spectrum

5. Compare the intensity of the two spike ions with the known lowest concentrations with each other and determine the difference between the measured ratio of the spike ion signal and the expected ratio.

6. If the absolute value of the difference between the measured ratio of the spike ion signals and the expected ratio is above a first predetermined threshold move to record the composite spectrum as final for the current location and select a new location to sample and repeat steps 1 to 7.

7. If the absolute value of the difference between the measured ratio of the spike ion signals and the expected ratio is greater than the first predetermined threshold repeat steps 1 to 5.

Further locations on the planar sample are sampled a number of times by a laser in a MALDI MS/MS mass spectrometer to generate a spectrum by repeating the following steps:

8. Desorb ions from the surface of the sample

9. Gate or trap desorbed ions to isolate ions with mass-to-charge ratios corresponding to tags of this invention.

10. Fragment the isolated tag ions by collision induced dissociation to release characteristic daughter ions that confirm the identity of the tags and allow relative quantification of the tags.

11. Record the intensity of the daughter ions or add the intensities of the daughter ions to the previously recorded intensities to build a composite spectrum.

12. Compare the intensity of the two spike ions with the known lowest concentrations with each other and determine the difference between the measured ratio of the spike ion signal and the expected ratio.

13. If the absolute value of the difference between the measured ratio of the spike ion signals and the expected ratio is less than a second predetermined threshold move to record the composite spectrum as final for the current location and select a new location to sample and repeat steps 8 to 14.

14. If the absolute value of the difference between the measured ratio of the spike ion signals and the expected ratio is greater than the predetermined second threshold repeat steps 8 to 12.

Fig. 7 is a flow diagram illustrating, in a general form, the steps described above. "Y" in Fig. 7 denotes "yes" and "N" denotes "no". In step 700 the specimen is contacted with probe molecules in accordance with the invention. In step 705 unbound probe molecules are removed by washing or rinsing with a buffer. In step 710, the specimen is contacted with a MALDI matrix material, as previously described. At this stage calibrants may also be introduced into the specimen. The calibrants may be part of the matrix material or may be applied in a separate step.

Step 715 is the option of performing a survey scan is available. If a survey scan is to be performed the process continues in step 720. If a survey scan is not performed the process continues at step 725. As previously described, in a survey scan, a crude image is generated by analysing each spot on the specimen based on one or a few shots (i.e. irradiations and detection of ions). Based on the crude image, a determination of where additional shots are required can be made. The additional shots are started in step 725.

In step 725 a portion of the specimen is irradiated. This results in the generation of ions that are selected using an apparatus as described with reference to Fig. 2. The ions of interest e.g. those corresponding to the mass tags (or the probes plus mass tags) and the calibrants, are selected, for example using an ion trap, quadropole, in step 730. The selected ions are fragmented by collision with a bath gas in step 735, and the fragments detected and recorded in step 740. The amounts recorded are also normalised in step 740 to allow them to be compared and combined with other sets of results. Example techniques for normalisation have already been described.

In step 745 a determination of whether that portion of the specimen needs to be reanalysed is made. This determination may be made on the basis of a survey scan, if one has been performed, or on the basis of signal to noise ratio or a ratio of calibrant ions, as already described.

If the portion is to be reanalysed, the same portion is irradiated again in step 750 and steps 730 to 745 are repeated.

If the portion of the specimen does not need to be reanalysed, the detected and normalised amounts are recorded as final for that portion in step 755. A determination is then made of whether further portions of the specimen need to be analysed or reanalysed in step 760, i.e. whether the analysis is complete. Again, this determination may be made on the basis of the results of a survey scan, or may simply be on the basis of whether a scanning process has covered all the desired portions of the specimen.

If the analysis of the specimen is complete, the process ends at step 770. If further analysis is required, a different (new) portion of the specimen is irradiated in step 765 and steps 730 to 745 subsequently repeated.

Tissue Preparation

In general, tissue for analysis should be treated to minimise degradation or diffusion from their starting positions of molecules of interest. Rnase inhibitors, DNAse inhibitors and Protease inhibitors can be added to washing and fixing buffers to minimise or eliminate endogenous nucleases, proteases and peptidases activity from degrading target molecules.

In order to minimise sample diffusion of molecules of interest away from their starting locations, several types of sample preparation procedure can be used.
In one preferred general procedure, tissue sections are dried slowly for several hours. MALDI matrix material (also referred to herein as "matrix") may then be applied, e.g. by electrospray addition of matrix. The chosen matrix should be electrosprayed directly onto the dried tissue slice followed by another drying step of 15 minutes in a vacuum desiccator. Alternatively, the dried tissue section can be applied to a target that has been pre-coated with matrix.

Alternatively, tissues can be fixed using one of the many techniques known in the art (Tbakhi et al., 1998, Am J Pathol. **152**(1): 35-41). Exemplar methods are fixation with neutral buffered formaldehyde or Bouin's solution as these are generally considered the best fixatives for in situ hybridisation (Weiss & Chen, 1991, J Histochem Cytochem. **39**(9): 1237-1242).

Blotting

Direct analysis of tissue has several drawbacks, including interference from signals from low molecular weight ions (MW < 500 Da) of low interest such as lipids or carbohydrates, which may obscure signals from low molecular weight molecules of interest peptides in the tissue and the potential of leakage of tissue fluid to other areas of the tissue sample preparation. In an effort to maintain the precise spatial position of a molecule in a sample with that of the actual target specimen imaged, samples may be blotted onto a stable substrate for subsequent analysis.

The target may be washed with water to remove salts and other water-soluble contaminants that do not adhere to the target surface. Other wash buffers include ammonium citrate, ammonium acetate.

Various types of target surfaces may be used such as stainless steel metal targets, C-18 micro bead covered target, nitrocellulose membranes, PVDF (Polyvinylidene difluoride) membranes.

In the simplest approach, blotting may be achieved by simply placing freshly cut tissue (with or without permeabilisation) onto the target surface for a short period, 10-30 seconds and then carefully lifting off the tissue.

Tissues may be rendered permeable prior to blotting by application of an organic solvent, such as ethanol or methanol, or by application of detergents. Detergents are, however, generally not preferred for use with mass spectrometry as they tend to suppress spectra.

Nucleic acids may also be blotted onto nitrocellulose or nylon membranes. These techniques are typically applied to gel separated nucleic acids to facilitate transfer of nucleic acids to a medium on which they can be readily probed (Brown, 2001, Curr Protoc Protein Sci., Appendix 4 & 4G; Krumlauf, 1996, Methods Mol Biol. **58**: 113-28; Pellé & -Murphy, 1993, Nucleic Acids Res. **21**(11): 2783-2784). The same process should be equally applicable to a tissue section. Typically, a gel (or in the context of this invention, a tissue section), is laid on the membrane. Layers of porous material, such as paper toweling are placed on top of the gel or tissue section and pressure is applied evenly to the gel/tissue section (either using suction, or by placing a stack of paper towels and a weight on top of the membrane and gel), to ensure good and even contact between gel and membrane. Buffer transfer by capillary action from a region of high water potential to a region of low water potential (usually filter paper and paper tissues) is then used to move the DNA from the gel on to the membrane; ion exchange interactions bind the DNA to the membrane due to the negative charge of the DNA and positive charge of the membrane. The membrane is then baked, i.e., exposed to high temperature (60 to 100 °C) (in the case of nitrocellulose) or exposed to ultraviolet radiation (nylon) to permanently and covalently crosslink the DNA to the membrane. The membrane can then be contacted with a probe labeled with the tags of this invention.

Proteins can also be extracted by electroblotting onto a polyvinylidene difluoride membrane after which enzymatic digestion of the proteins can take place on the membrane (Vestling & Fenselau, 1994, Biochem Soc Trans. **22**(2): 547-551).

Other applications:

The description above has focussed on the imaging analysis of tissue sections using the methods of this invention. Other applications of the imaging techniques described include imaging of microarrays and detection of samples that have been spotted out on a planar array.

A microarray is a 2-dimensional array of nucleic acid probes designed to be complementary to target sequences in a sample. Practically speaking a microarray could comprise an array of wells on microtitre plates, for example, such that each well contains a single immobilised oligonucleotide that is a member of the array.

Alternatively, the array may be synthesised combinatorially on a glass "chip" according to the methodology of Southern or that of Affymetrix, Santa Clara, California (see for example: Pease et al., 1994, PNASUSA **91**: 5022-5026; Maskos & Southern, 1993, Nucleic Acids Res. **21**: 2269-2270; Southern et al, 1994, Nucleic Acids Res. **22**: 1368-1373) or using related ink-jet technologies such that discrete locations on the glass chip are derivatised with one member of the hybridisation array (Barbulovic-Nad et al., 2006, Crit Rev Biotechnol. **26**(4): 237-59).

In a typical microarray analysis, a label is introduced into a target sample and the target sample is hybridised to the microarray. The identity and amount of each target sequence in the sample is determined by the intensity of the signal from the label that is detected at the location corresponding to the probe that is complementary to the target sequence For the purposes of this invention it is anticipated that the tags of this invention can be readily substituted for the tags used in the many methods for labelling cDNA or cRNA with fluorescent tags (Schindler et al., 2005, Anal Biochem. **344**(1): 92-101; Singh et al., 2005, Am J Physiol Cell Physiol. **288**(5): C1179-89; Kurn et al., 2005, Clin Chem. **51**(10): 1973-81).

In particular the method disclosed by Ericsson et al. (2008, Nucleic Acids Res. **36**(8): e45) is well suited to the practice of this invention. This publication discloses the use of circularising oligonucleotide probes that are amplified by linear amplification in a two stage process. The second stage of the process produces a linear amplification product covalent linked to the surface of a microarray. Captured rolling circle products at each location on the array are probed with fluorescently labelled oligonucleotides prior to detection. The fluorescently labeled oligonucleotides described in this method can be directly substituted with the probes of this invention. In such a modified microarray experiment, scanning of the array would obviously be performed in a MALDI MS/MS instrument in accordance with the methods of this invention.

Example 1: Preparation of a peptide-tagged oligonucleotide

An oligonucleotide probe conjugate with the structure shown in Fig. 1 was synthesised. The peptide tag sequence was:

(N-terminus) – Acetate - Alanine - Piperazin-1-ylacetic acid – Alanine – Photocleavable Linker – Glutamic Acid – Amide – (C-terminus) (SEQ ID NO: 10).

The synthesis was performed using standard FMOC amino acid chemistry on a PAL resin, which releases the C-terminal amino acid as an amide in the final deprotection and cleavage step. The synthesis was performed on a custom-made synthesiser constructed from a Gilson 215 liquid handling robot configured for peptide synthesis. The peptide was purified by HPLC.

The free glutamic carboxylic acid was used to couple the peptide to an oligonucleotide with the following sequence:

5'-GTATGCACCTGGGCTCTTTGCAGGTCTCT-3' (SEQ ID NO: 11).

The oligonucleotide was synthesised using standard phosphoramidite procedures on an Applied Biosystems Expedite synthesiser. After completion of the oligonucleotide synthesis, an aminohexyl linker (6-(4-Monomethoxytritylamino)hexyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite) was introduced at the 5' terminus. The amino group was deprotected and then the peptide was coupled using a published method (Zaramella et al., 2004, J Am Chem Soc. **126**(43): 14029-35). The peptide was coupled using the following relative concentrations of peptide, condensing reagent and base in dimethylformamide: 20:19:43 of peptide: HBTU (O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate) : NMM (1-methylmorpholine) respectively with the final concentration of peptide being 0.04 M.

The conjugate was purified by HPLC. The conjugate was then analysed by MALDI MS/MS on a Shimadzu Axima Performance TOF-TOF instrument (with Mode Reflectron, Power: 66, Gate: 324.05-333.08, P.Ext. @ 500 (bin 52), 190 mV[sim=21441 mV] Profiles 1-113 Smooth Av 1). The conjugate was analysed by co-crystallisation with both 2,5-dihydroxybenzoic acid (DHB) and 4-hydroxy -alpha-cyano-cinnamic acid (HCCA). Fig.

8 shows a mass spectrum of the photocleaved and desorbed tag peptide released from the conjugate undergoing consecutive cleavage to give the expected daughter ion (see also Fig. 1).

Example 2: Selection of MALDI matrix material for tag cleavage and detection

A further peptide, hereinafter referred to as Peptide Tag 1, was synthesised. Peptide Tag 1 has the structure:

(N-terminus) - Acetate – Leucine – Piperazin-1-ylacetic acid – Leucine – Photocleavable Linker – Glutamic Acid – Amide – (C-terminus) [SEQ ID NO: 12],

and is shown in Fig. 10a (top) together with its photocleavage ("P") fragment (bottom).

Equal concentrations of Peptide Tag 1 were analysed in three different matrices: 4-hydroxy-alpha-cyano-cinnamic acid (HCCA), 2,5-dihydroxybenzoic acid (DHB) and sinapinic acid (SA). Samples of 600 pmol of Peptide Tag 1 were co-crystallised in saturated (SA – 18 mg/ml; HCCA - 20 mg/ml; DHB - 25 mg/ml) solutions of each matrix where each matrix was dissolved in 1:1 acetonitrile/water with 0.1% Trifluoroacetic acid. 1 µl of each sample was spotted down in micro-wells on a stainless steel MALDI target and allowed to dry. The target was then analysed by MALDI MS/MS on a Shimadzu Axima Performance TOF-TOF instrument operated in MS-mode. For each matrix, a series of spectra were collected as laser power was varied. 10 shots were acquired at each of 121 locations within a square raster of 400 x 400 µm and summed for each matrix sample at each laser power. The intensity of the native "peptide" ion, i.e. the Peptide Tag 1 ions that have not undergone photocleavage (m/z 821.4 for $[M+H]^+$ ion), and the intensity of the ion corresponding to the expected photocleavage fragment (m/z 412.3) were recorded for each matrix (see Fig. 10a for structures of precursor tag peptide ion and the photocleaved tag peptide "fragment" ion). These values are plotted in Figs 10b (for HCCA), 10c (for DHB) and 10d (for SA). By comparison of Figs 10b, 10c and 10d, DHB gives the highest overall intensity signal for the photocleavage "fragment" ions at a laser power of 90. Note that the laser power is measured in arbitrary units defined by the instrument. SA gives similar but slightly lower intensities compared to DHB, while the signal intensity from HCCA is quite low.

In the same experiment, further spectra were obtained while operating the instrument in two different MS/MS modes. The Shimadzu instrument can perform Post Source Decay (PSD) analysis and full Collision Induced Dissociation (CID) of a selected precursor ion and both methods were applied to determine the conditions that give rise to the most intense signal from the desired daughter (m/z 257.2) ion generated by fragmentation of the photocleaved fragment ion (m/z 412.3). Again laser power was varied and the intensities of parent tag ion (precursor ion selected by the instrument) and the daughter ion were recorded for each laser power. The results of this experiment are shown in Figs 11a, 11b and 11c for PSD and Figs 11d, 11e and 11f for CID. Both DHB and SA give good results in either PSD or CID modes. Again DHB appears to be a bit more sensitive than SA but both give much higher signals than HCCA. Both give the highest signals for daughter ion detection at a laser power of 90 in the PSD mode. DHB gives the highest signal in CID mode at a laser power of 90 while SA gives the highest signal in CID mode at a laser power of 100. SA was selected for further use as it gives more homogenous deposition than DHB, which is important for consistent results in imaging.

In a further experiment, the optimum concentration of SA was determined. Equimolar concentrations of Peptide Tag 1 550 pmol/µl were dissolved in SA at various concentrations in 1:1 acetonitrile:water 0.1 % TFA and spotted onto a target. The samples were analysed by PSD to determine the highest daughter ion signal at a laser power of 90. It can be seen from Fig. 12 that the highest signal was obtained with a concentration of 4 mg/ml of SA. For practical purposes, however, it was found that a concentration of 6mg/ml sinapinic acid produces more homogenous and reproducible deposition of matrix when sprayed.

Example 3: Synthesis of a set of three isobaric tags and conjugation to oligonucleotides

Three further peptides with the structures shown in Fig. 13(a) were synthesised. These peptides give rise to three isobaric fragments after photocleavage (note Peptide Tag 4 doesn't photocleave but is synthesised to have the structure of a photocleaved peptide tag). The peptide tag sequences were:

Peptide Tag 2: (N-terminus) - Acetate-$^{13}C_6$ Leucine - Piperazin-1-ylacetic acid – Leucine – Photocleavable Linker – Glutamic Acid – Amide – (C-terminus) [SEQ ID NO: 13];

Peptide Tag 3: (N-terminus) - Acetate- Leucine - Piperazin-1-ylacetic acid – $^{13}C_6$ Leucine – Photocleavable Linker – Glutamic Acid – Amide – (C-terminus) [SEQ ID NO: 14]; and

Peptide Tag 4: (N-terminus) - Acetate- D_3 Leucine - Piperazin-1-ylacetic acid – D_3 Leucine – Amide – (C-terminus) [SEQ ID NO: 15].

Synthesis was performed using standard FMOC amino acid chemistry on a PAL resin, which releases the C-terminal amino acid as an amide in the final deprotection and cleavage step. The synthesis was performed on a custom-made synthesiser constructed from a Gilson 215 liquid handling robot configured for peptide synthesis. The peptide was purified by HPLC.

Peptide Tags 2 and 3 were conjugated to oligonucleotides. Peptide Tag 4 was synthesised with the structure of the expected photocleavage fragment and was designed to be used as a spike that will be added to the matrix to assist with normalisation of spectra.

The free glutamic carboxylic acid on Peptide Tag 2 was used to couple to the peptide to an oligonucleotide with the following sequence:

5'-GTATGCACCTGGGCTCTTTGCAGGTCTCT-3' (SEQ ID NO: 16) to give the conjugate:

CONJUGATE2: PEPTIDETAG2-5'-GTATGCACCTGGGCTCTTTGCAGGTCTCT-3' (i.e. SEQ ID NO: 13 conjugated to SEQ ID NO: 16).

Similarly, Peptide Tag 3 was coupled to an oligonucleotide

5'- TCTCTGGACGTTTCTCGGGTCCACGTATG-3' (SEQ ID NO: 17), having a reverse sequence of SEQ ID NO: 16,

to give the conjugate:

CONJUGATE3: PEPTIDETAG3-5'- TCTCTGGACGTTTCTCGGGTCCACGTATG -3' (i.e. SEQ ID NO: 14 conjugated to SEQ ID NO: 16).

The oligonucleotides were synthesised using standard phosphoramidite procedures on an Applied Biosystems Expedite synthesiser. After completion of the oligonucleotide synthesis, an aminohexyl linker (6-(4-Monomethoxytritylamino)hexyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite) was introduced at the 5' terminus. The amino group was deprotected and then the peptide was coupled using a published method (Zaramella et al., 2004, J Am Chem Soc. **126**(43): 14029-35). The peptide was coupled using the following relative concentrations of peptide, condensing reagent and base in dimethylformamide: 20:19:43 of peptide: HBTU (O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate) : NMM (1-methylmorpholine) respectively with the final concentration of peptide being 0.04 M.

The conjugates were purified by HPLC. An experiment was carried out to assess the relative sensitivity of Peptide Tag 4 (the matrix spike) compared to the two conjugates. To this end, both of the oligonucleotide/peptide (CONJUGATE2 and CONJUGATE3) conjugates were mixed in equimolar quantities with Peptide Tag 4. The mixture was then analysed by MALDI MS/MS on a Shimadzu Axima Performance TOF-TOF instrument using post source decay on ions with m/z 418. The conjugate was analysed by co-crystallisation with sinapinic acid (4 mg/ml as determined from example 2). Fig. 13(b) shows a mass spectrum (869 mV 297 mV 244 mV) of the desorbed and photocleaved tag peptides at m/z 418 selected for PSD undergoing consecutive cleavage to give the expected daughter ions giving 3 distinct peaks at m/z 257, 260 and 263. After laser desorption, both conjugates and the free peptide tag 4 give rise to a single ion at 418 with successive fragmentation to give the three expected daughter ions demonstrating the multiplexing capability of the Tag Peptides of this invention. The intensity of the 260 ion is slightly greater than the other two. This ion corresponds to the Peptide Tag 4, which was not conjugated and the higher intensity for this ion reflects the fact that it does not need to undergo photocleavage. Significantly, the difference in intensity between the conjugated and unconjugated peptides is not very great implying that the efficiency of the photocleavage process is high.

Fig. 13(c) shows intensity-adjusted spectra (310 mV 56 mV 52 mV) for the three tags showing that there is very little "cross-talk" or overlap between the tags from higher mass isotopes of each tag, i.e. the higher mass isotopes of the lightest tag only contribute a neglible amount to the signal for the higher mass tags and so on.

Example 4: Normalisation of signal intensities

Quantification in MALDI mass spectrometry is complicated by matrix effects as well as the competitive nature of the ionisation process (Karas et al., 2000, J Mass Spectrom. **35**(1):1-12). Some MALDI effects are illustrated in Fig. 14(a). In this figure, an image of a MALDI target is shown. The image is a map of the ionisation intensities of a series of four samples of the calcium channel blocker drug verapamil. This molecule was selected because it illustrates the issue of quantification by MALDI. The four samples shown contain the same concentration of drug but each sample differs due to the presence of different excess "background ions". In the first sample, the drug was mixed with the phospholipid dipalmitoyl phosphatidylethanolamine (DPPE), in the second sample the drug was mixed with Dipalmitoyl Phosphatidylcholine (DPPC), in the third with cholesterol and finally a control with no background ion. The two phospholipids and cholesterol were each made up at 10 mg/ml. The Verapamil was made up at 1 mg/ml. 1 µl of drug solution was spotted down with 1µl of each of the lipids and left to dry. The targets were then spray coated with SA matrix (6 mg/ml) and imaged by MALDI MS (m/z at 455). In the presence of different backgrounds, the intensity of the drug ions varies considerably despite the drug being present at the same concentration in all samples.

Fig. 14(b-i) shows a further image of a MALDI target. This image is a map of the ion intensities of the daughter ion from four samples with the same concentration of CONJUGATE2 (daughter ion at m/z 257 produced by fragmentation of the m/z 418 parent ion) in the presence of various background ions. Again, the intensity of the tag ion varies with the background ion. Fig. 14(b-ii) shows the same MALDI target but the image is now a map of the intensities of the daughter ion from CONJUGATE3 (daughter ion at m/z 263 produced by fragmentation of the m/z 418 parent ion), which was added to the matrix as normalisation spike. Again, the intensity varies with background. Fig. 14(b-iii) shows the effect of normalising the intensity of CONJUGATE2 by dividing this with the intensity of CONJUGATE3 from the same location. The normalised intensities of CONJUGATE2 are homogenised effectively using a normalisation tag.

Example 5: Quantification of signal intensities

Fig. 15(b) left graph shows the results of a dilution series of CONJUGATE2 as depicted in Fig. 15(a) left graphic, while Fig. 15(b) right graph shows a similar dilution series of

CONJUGATE3 as depicted in Fig. 15(a) right graphic. 4 concentrations of each conjugate were made up; 200 fmol, 1 pmol, 5 pmol and 25 pmol were diluted in 10 mg/ml BSA. 1 µl of this sample was deposited on a metal target. The target was then spray coated with SA matrix was (6 mg/ml sinapinic acid; 1:1 acetonitrile:water with 0.1 % TFA). The samples were then imaged using PSD on the photocleavage product at m/z 418. The intensities of the daughter ions at m/z 257 and 263 (see Fig. 15(a), middle graphic) for were recorded from each concentration for CONJUGATE2 and CONJUGATE3 respectively. From both graphs in Fig. 15(b), it can be seen that the intensities of each tag daughter ion increase in a direct linear relationship with the concentration of the conjugates.

Example 6: Hybridisation of probe to a target fixed on a tissue section

Fig. 16 left image shows a pair of kidney tissue sections from a mouse. The kidney was sectioned and then fixed by immersion of the tissue in 10% neutral buffered formalin for 2 hours. Kidney tissue was selected as it does not express a sequence similar to HER2_TARGET below.

50 pmol of synthetic Her2 target (see below) was sprayed onto half of one of the kidney sections (shown in Fig. 16 left image, lower section, left side) using a CAMAG ATS4 automated TLC sampler (CAMAG, Switzerland).

HER2_TARGET:

5'_AGGGGTTTGCTCCGGAGAGACCTGCAAAGAGCCCAGGTGCATACCTTGGCAATCT GCATA_3' (SEQ ID NO: 18).

After application of the HER2_TARGET sequence (SEQ ID NO: 18), the tissue was post-fixed with cold (4°C) 0.4% paraformaldehyde/Phosphate Buffered Saline for 20 minutes. In situ hybridisation (ISH) was carried out using CONJUGATE2. The following protocol was used:

1. The tissue sections were rinsed in Diethyl Pyrocarbonate (DEPC) treated H_20.
2. The tissue sections were then immersed in Prehybridisation buffer: 50µl of Prehybridisation buffer was pipetted onto each section (2 ml Pre-hybridisation buffer: 240ul 5M NaCl; 200ul 10xPE; 20ul 10mg/ml salmon sperm DNA; 400ul 50% polyethylene glycol 6000; 1140ul DEPC H_20). A coverslip was then laid on each

section to prevent evaporation. The sections were incubated at 37°C for 1hr. Prehybridisation buffer was removed. The prehybridisation step blocks non-specific binding sites, decreasing background.

3. Each tissue section was then exposed to Hybridisation buffer. 50µl hybridisation buffer was pipetted onto each section (1 ml Hybridisation buffer: 120 ul 5M NaCl; 100µl 10xPE (500 mM PIPES; 50 mM EDTA; pH adjusted to 7.2 with 1 M KOH); 10µl 10mg/ml salmon sperm DNA; 200ul 50% polyethylene glycol 6000; 100 µl of CONJUGATE2 (550 pmol/µl); made up to 1ml with DEPC H_20). The sections were covered with coverslips and incubated at 37°C overnight.
4. Coverslips were removed from sections by rinsing in 4xSSC.
5. The tissue sections were then washed 3 times for 10 minutes in 1 x SSC (Saline Sodium Citrate; 150 mM NaCl; 15 mM Sodium Citrate) at 37°C.
6. The tissue sections were then washed twice for 1 minute in DEPC H_2O at 37°C to remove salt prior to MALDI mass spectrometric analysis.
7. The two tissues sections were laid onto a metal target.
8. The tissue sections were then desiccated slowly.
9. Both desiccated tissue sections were then sprayed with 6mg/ml sinapinic acid matrix (again using the CAMAG ATS4 instrument). The sections were visualised together by MALDI TOF MS/MS using PSD selecting the m/z 418 peptide tag photocleavage fragment from CONJUGATE2 as the precursor.

Fig. 16 right image shows the distribution of the daughter ion at m/z 257 produced by PSD of the 418 ion over the two kidney sections. The daughter is located to the spiked area of the tissue. Some migration of the target appears to have occurred during the post-fixing process. Endogenous RNA is unlikely to migrate in this way, as it is typically complexed with proteins and other intracellular molecules that will ensure it fixes well during the initial fixation step.

Example 7: Synthesis of a set of two isobaric tags and conjugation to oligonucleotides

Two peptides with the structures shown in Fig. 17 were synthesised. These peptides give rise to two isobaric fragments after photocleavage. The peptide tag sequences were:

Peptide Tag 5: (N-terminus) - Acetate- Leucine - Piperazin-1-ylacetic acid – Valine – Photocleavable Linker – Glutamic Acid – Amide – (C-terminus) (SEQ ID NO: 19; top structure in Fig. 17)

and

Peptide Tag 6: (N-terminus) - Acetate- Valine - Piperazin-1-ylacetic acid – Leucine – Photocleavable Linker – Propargyl Glycine – Amide – (C-terminus) (SEQ ID NO: 20; bottom structure in Fig. 17).

The synthesis was performed using standard FMOC amino acid chemistry on a PAL resin, which releases the C-terminal amino acid as an amide in the final deprotection and cleavage step. The synthesis was performed on a custom-made synthesiser constructed from a Gilson 215 liquid handling robot configured for peptide synthesis. The peptide was purified by HPLC.

Peptide Tags 5 and 6 were conjugated to oligonucleotides with the sequence below:

5'-GTATGCACCTGGGCTCTTTGCAGGTCTCT-3' (SEQ ID NO: 21).

The free glutamic carboxylic acid on Peptide Tag 5 was used to couple to the peptide to an oligonucleotide with the above sequence but modified with a 5' amino group using the same protocol as Example 3.

CONJUGATE4:

PEPTIDE TAG5 - 5'-GTATGCACCTGGGCTCTTTGCAGGTCTCT-3' (i.e. SEQ ID NO: 19 conjugated to SEQ ID NO: 21).

Similarly, Peptide Tag 6 (SEQ ID NO: 20) was coupled to three oligonucleotides with the same sequence as SEQ ID NO: 21 using the Copper Catalysed Azide Alkyne Cycloaddition (CuAAC) reaction (Gogoi K et al., 2007, Nucleic Acids Res. **35**(21):e139). Each oligonucleotide had 1, 2 or 4 internal amino groups on the second thymidine residue

of SEQ ID NO: 21, to produce CONJUGATE_1TAG, CONJUGATE_2TAG and CONJUGATE_4TAG, respectively.

The oligonucleotides were synthesised using standard phosphoramidite procedures on an Applied Biosystems Expedite synthesiser. To incorporate internal amino groups at the second thymidines of SEQ ID NO: 21, Amino-Modifier C6 dT (5'-Dimethoxytrityl-5-[N-(trifluoroacetylaminohexyl)-3-acrylimido]-2'-deoxyUridine,3'-[(2-cyanoethyl)-(N,N-diisopropyl)]-phosphoramidite; Glen Research, Sterling, VA, USA) was incorporated. The Trifluoroacetyl group protecting the amino groups was removed during the standard final deprotection step.

The oligos were then modified to incorporate an azide function at each of the free amino groups. Azidobutyrate NHS Ester (Glen Research, Sterling, VA, USA) was used to introduce the azide to the amino groups. A solution of the azide reagent (100 equivalents) in DMSO (100 µl) was added to a solution of a purified oligo (0.2 micromole, 1 equivalent) in 0.5M sodium carbonate buffer pH 8.75 (100 µl). The reaction was left at Room Temperature for 5 hours, NAP-G25 gel filtered and then purified by reverse phase HPLC.

The reaction conditions for the coupling of the azide oligonucleotides to the propargyl peptide (Peptide Tag 6 [SEQ ID NO: 20]) depended on the number of azide sites in the oligonucleotides. For coupling of the oligonucleotides with one and two azide sites the aqueous reaction comprised: 1 equivalent of oligonucleotide, 20 equivalents of CuSO4, 200 equivalents of sodium ascorbate, 140 equivalents of tri-hydroxypropyltriazole and 50 equivalents of peptide. For the oligonucleotide with 4 azide sites using 1 equivalent of oligonucleotide, 50 equivalents of CuSO4, 500 equivalents of sodium ascorbate, 350 equivalents of tri-hydroxypropyltriazole and 80 equivalents of peptide.

To carry out the coupling, a solution of sodium ascorbate was added to a degassed tri-hydroxypropyltriazole solution, followed by CuSO4 solution, peptide solution and oligo solution, respectively. With each addition step, the reaction was degassed with argon for 1 minute. Then the reaction was left in the dark at RT for 2 hours. After the reaction was completed, the conjugates were desalted by NAP-G25 (X2) gel filtration and then purified by reverse phase HPLC.

Example 8: Analysis of multiply labeled probes by MALDI-TOF/TOF mass spectrometry

The conjugates from Example 7 were made up to a concentration of 50 pmol/µl in water. The oligonucleotide/peptide conjugates were then analysed by mixing equimolar quantities of CONJUGATE4 with the other conjugates: CONJUGATE_1TAG, CONJUGATE_2TAG and CONJUGATE_4TAG respectively. The mixture was then analysed by MALDI MS/MS on a Shimadzu Axima Performance TOF-TOF instrument using post source decay on ions with m/z 398. All four conjugates yield [M+H]+ ions at m/z 398 after Photocleavage. CONJUGATE_1TAG, CONJUGATE_2TAG and CONJUGATE_4TAG produce a daughter ion at m/z 257 after Post Source Decay. For comparison, CONJUGATE_4 was used as an internal standard, which produces a daughter ion at m/z 243. The conjugates were analysed by co-crystallisation with sinapinic acid (20 mg/ml).

Samples were prepared in duplicate with three replicates taken from the each pair. The samples were as follows:

1. 0.5 µL CONJUGATE4, 0.5 µL water and 0.5 µL sinapinic acid (20mg/mL)
2. 0.5 µL CONJUGATE4, 0.5 µL CONJUGATE_1TAG and 0.5 µL sinapinic acid (20mg/mL)
3. 0.5 µL CONJUGATE4, 0.5 µL CONJUGATE_2TAG and 0.5 µL sinapinic acid (20mg/mL)
4. 0.5 µL CONJUGATE4, 0.5 µL CONJUGATE_4TAG and 0.5 µL sinapinic acid (20mg/mL)

For the PSD analysis the ion gate was set such that the m/z 398 ion was selected. Typical spectra are shown in Fig. 18. It can be seen in Fig. 18 that the three ions of interest are visible. The m/z 398 parent ion and the two daughter ions at m/z 243 and m/z 257 can be detected when present.

The ion intensity counts for the three ions of interest from the replicated analyses are plotted in Table 1 below. Typical MALDI spot to spot variation can be seen from analysis to analysis, but in general the ion ratios are reproducible. Replicates of sample 4 where the oligonucleotide was conjugated to more tags show a much higher ratio of the 257 ion to the control ion at 243 indicating that conjugation of more tag to the oligonucleotide increases signal intensity as expected. Similarly, sample 3 replicates (2 tag conjugate) has a signal intensity between that of the singly labelled and quadruply labelled probes.

Table 1

PSD	m/z 398	m/z 243	m/z 257	Ratio of 257/243
1	410	67		
1	682	191		
1	914	195		
2	922	123	109	89%
2	653	105	69	66%
2	640	107	71	66%
3	*567*	*71*	*39*	*55%*
3	331	29	32	110%
3	263	29	35	121%
4	261	26	60	231%
4	646	55	114	207%
4	429	36	76	211%

Although the present invention has been described with reference to preferred or exemplary embodiments, those skilled in the art will recognise that various modifications and variations to the same can be accomplished without departing from the spirit and scope of the present invention and that such modifications are clearly contemplated herein. No limitation with respect to the specific embodiments disclosed herein and set forth in the appended claims is intended nor should any be inferred.

All documents cited herein are incorporated by reference in their entirety.

Claims

1. A method for analysing a specimen, comprising the steps of:

a) contacting the specimen with probe molecules, each of the probe molecules including one or more mass tags coupled to each probe molecule via a first cleavable linker, and allowing probe molecules to become bound to the specimen;

b) removing unbound probe molecules from the specimen;

c) contacting the specimen with a Matrix-assisted Laser Desorption Ionisation (MALDI) matrix material, or laying the specimen on a surface coated with a MALDI matrix material;

d) irradiating a portion of the specimen with a laser beam to release ions from the specimen;

e) selecting, from the ions released in step (d), ions with mass-to-charge ratios corresponding to the mass tags or derivatives of the mass tags;

f) recording the amount and type of ions selected in step (e), together with the location of the portion of the specimen, as a result;

g) repeating steps (d) to (f) on a different portion of the specimen.

2. The method according to claim 1, wherein steps (d) to (f) are repeated a plurality of times for at least one portion of the specimen.

3. The method according to claim 1 or 2, wherein the step of recording comprises combining the amount of ions selected in step (e) with a previously recorded amount of selected ions for that portion of the specimen to produce a composite result.

4. The method according to any preceding claim, wherein the specimen is a biological specimen.

5. The method according to any preceding claim, wherein the specimen is a tissue sample, an array of tissue samples or a microarray.

6. The method according to any preceding claim, wherein the specimen is a plurality of samples blotted onto a substrate from a tissue sample.

7. The method according to any preceding claim, wherein the specimen comprises a sample or samples labelled with oligonucleotide-labelled ligands, wherein the probe molecules bind to the oligonucleotide-labelled ligands.

8. The method according to any preceding claim, wherein the specimen is contacted with a plurality of different probe molecules, each different probe molecule labelled with a different mass tag.

9. The method according to claim 8, wherein at least two of the different mass tags are isotopes of one another.

10. The method according to claim 8 or 9, wherein at least two of the different mass tags are of equal mass.

11. The method according to any preceding claim, wherein at least one of the probe molecules comprises a plurality of mass tags.

12. The method according to any preceding claim, wherein the released ions comprise probe molecule-mass tag complexes and the step of selecting includes splitting the mass tags from the probe molecules.

13. The method according to claim 12, wherein the mass tags are split from the probe molecules by photoillumination, collision or by a chemical reagent.

14. The method according to any preceding claim wherein at least some of the mass tags comprise at least two mass modifiers coupled together via a second cleavable linker.

15. The method according to claim 14, further comprising of step of:
fragmenting the ions selected in step (e) to produce daughter ions prior to step (f);
wherein step (f) comprises detecting and recording the amount and type of each daughter ion together with the location of the portion of the specimen.

16. The method according to claim 15 or 16, wherein the second cleavable linker is a collision cleavable linker.

17. The method according to any preceding claim, wherein the first cleavable linker is a photocleavable linker.

18. The method according to any preceding claim, wherein the specimen is a planar tissue sample.

19. The method according to any preceding claim, wherein the method comprises raster scanning the specimen with the laser beam.

20. The method according to any preceding claim, further comprising the step of optimising the power of the laser beam prior to step (d) to provide the largest amount of a predetermined ion or ions selected in step (e).

21. The method according to any preceding claim, further comprising the step of optimising the concentration of the MALDI matrix material to provide the largest amount of ions selected in step (e) for a predetermined MALDI matrix material and a predetermined laser power.

22. The method according to any preceding claim further comprising the step of selecting a concentration of the MALDI matrix material that provides the most homogenous deposition of the MALDI matrix material on the specimen using a spraying operation.

23. The method according to any preceding claim, further comprising the step of normalising the amount of selected ions recorded in step (f).

24. The method according to claim 23, wherein the step of normalising comprises the steps of:

including in the MALDI matrix material or specimen a predetermined concentration of calibrant that produces calibrant ions when irradiated by the laser beam;

detecting and recording an amount of calibrant ions produced following the step of irradiating; and

dividing the amount of ions selected in step (e) by the amount of calibrant ions to provide a normalised amount of selected ions.

25. The method according to claim 23, wherein the step of normalising comprises the steps of:

including in the MALDI matrix material or specimen a predetermined concentration of a plurality of different calibrants that produce calibrant ions when irradiated by the laser beam;

detecting and recording an amount of each type of calibrant ion produced following the step of irradiating and producing a calibration function from the amounts; and

calibrating the amount of each type of ion selected in step (e) with the calibration function to provide a normalised amount of each type of selected ion.

26. The method according to claim 24 or 25, wherein at least one calibrant is a mass tag calibrant or a probe molecule calibrant.

27. The method according to claim 26, wherein the at least one calibrant is an isotope of one of the mass tags of the probe molecules.

28. The method according to any preceding claim, wherein the step of recording comprises producing digital signals indicative of the amount and type of selected ions, further comprising the step of removing noise from the digital signals by removing all signals below a threshold level of recorded intensity, wherein the threshold level is dependent on the mass to charge ratio of the recorded ions to which the signals correspond.

29. The method according to claim 28, further comprising the step of smoothing the digital signal.

30. The method according to any preceding claim, further comprising the steps of:

analysing the results and determining, based on the results, whether to repeat steps (d) to (f) for a portion of the specimen.

31. The method according to claim 30, wherein the step of determining comprises calculating a difference value based on the difference between the recorded result for that portion and the recorded result for at least one adjacent portion, and comparing the difference value with a predetermined threshold.

32. The method according to claim 30, wherein the step of determining comprises determining a measure of the quality of the result based on an amount of detected calibrant ions and comparing the measure with a predetermined threshold.

33. The method according to claim 32, wherein at least one calibrant is a mass tag calibrant or a probe molecule calibrant.

34. The method according to claim 33, wherein the at least one calibrant is an isotope of one of the mass tags of the probe molecules.

35. The method according to any preceding claim, comprising performing steps (a) to (g) to produce a survey scan, and selecting portions of the specimen on which to repeat steps (d) to (f) based on an analysis of the survey scan.

36. The method according to claim 35, wherein the step of irradiating is performed using a first laser spot size to produce a survey scan, and wherein the step of irradiating is performed using a second laser spot size on selected portions of the specimen, wherein the second laser spot size is smaller than the first laser spot size.

37. The method according to any preceding claim, further comprising the steps of:

including in the MALDI matrix material or specimen a predetermined concentration of at least one calibrant that produces calibrant ions when irradiated by the laser beam;

detecting and recording an amount of the calibrant ions produced following the step of irradiating,

comparing the amount of calibrant ions with a detected background noise to provide a ratio;

determining, prior to step (g), whether the ratio is above a predetermined threshold; and

if the ratio is above a predetermined threshold, recording in step (f) the result as final for that portion of the specimen, and

if the ratio is not above the predetermined threshold, returning to step (d) for the same portion of the specimen.

38. The method according to claim 37, wherein a plurality of different calibrant ions are added to the MALDI matrix material or specimen, and the step of comparing comprises comparing the amount of calibrant ions of the calibrant with the lowest detected concentration with the background noise to provide the ratio.

39. The method according to any one of claim 1 to 36, further comprising the steps of:

including in the MALDI matrix material or specimen a predetermined concentration of a first calibrant that produces first calibrant ions when irradiated by the laser beam and a known concentration of a second calibrant that produces second calibrant ions when irradiated by the laser beam;

detecting and recording an amount of the first calibrant ions and an amount of the second calibrant ions, produced following the step of irradiating;

comparing the amounts of the first and second calibrant ions with one another to provide a ratio;

determining the difference between the ratio and an expected ratio; and

if the value of the difference is less than a predetermined threshold, recording in step (f) the result as final for that portion of the specimen, and

if the value of the difference is not less than the predetermined threshold, returning to step (d) for the same portion of the specimen.

40. The method according to any one of claims 37 to 39, wherein at least one calibrant is a mass tag calibrant or a probe molecule calibrant.

41. The method according to claim 40, wherein the at least one calibrant is an isotope of one of the mass tags of the probe molecules.

42. The method according to any one of the preceding claims, further comprising the step of producing an image of the distribution of each type of selected ion across the specimen.

43. A probe molecule comprising a multimeric polyamide, the multimeric polyamide having two or more polyamide branches, wherein at least two of the polyamide branches comprise or consist of a mass tag.

44. The probe molecule according to claim 43, wherein one or more or all of the polyamide branches comprising or consisting of a mass tag have a cleavable linker via which the polyamide branches are cleavable from the branched polyamide.

45. The probe molecule according to claim 44, wherein the cleavable linker is photocleavable, collision cleavable and/or acid cleavable.

46. The probe molecule according to claims 44 or 45, wherein one, two, more or all of the mass tags comprise a first mass modifier sequence coupled to a second mass modifier sequence via the cleavable linker (for example, a collision cleavable linker).

47. The probe molecule according to claim 46, wherein the first mass modifier sequence, the second mass modifier sequence and/or the cleavable linker comprise a charge-carrying entity.

48. The probe molecule according to any of claims 43 to 47, wherein one or more or all of polyamide branches comprise a charge-carrying entity.

49. The probe molecule according to any of claims 43 to 48, wherein the polyamide branches are substantially identical to each other.

50. The probe molecule according to any of claims 43 to 49, wherein the polyamide is a peptide.

51. The probe molecule according to any of claims 43 to 50, wherein the polyamide is coupled to a protein (for example, an affinity ligand such as an antibody) and/or an oligonucleotide.

52. The probe molecule according to claim 51, wherein the oligonucleotide is coupled to the multimeric polyamide via a covalent bond (for example, via a reaction with a free or trityl-protected hydroxyl group on the polyamide).

53. The probe molecule according to claim 51 or 52, wherein the oligonucleotide is coupled to the multimeric polyamide via a non-covalent bond (for example, via biotin-avidin binding).

54. The probe molecule according to any of claims 51 to 53, wherein the oligonucleotide is up to about 100 nucleotides in length, for example about 30-60 or about 35-50 or about 40 nucleotides in length.

55. A set of probe molecules comprising at least two probe molecules as defined in any of claims 43 to 54, wherein the mass tags are isobaric.

56. A multimeric polyamide comprising a plurality of cleavable polyamide branches.

57. The multimeric polyamide according to claim 56, wherein one or more or all of the cleavable polyamide branches comprise or consist of a mass tag.

58. The multimeric polyamide according to claim 56 or 57, wherein one or more or all of the cleavable polyamide branches comprise a charge-carrying entity.

59. The multimeric polyamide according to any of claims 56 to 58, wherein the cleavable peptide branches are photocleavable, collision cleavable or acid cleavable.

60. The multimeric polyamide according to any of claims 56 to 59, wherein the polyamide is a peptide.

61. A probe molecule comprising a multimeric oligonucleotide, wherein a first oligonucleotide sequence of the multimeric oligonucleotide binds to a predetermined target and a second oligonucleotide sequence of the multimeric oligonucleotide is linked to a mass tag.

62. The probe molecule according to claim 61, further comprising an intermediate oligonucleotide covalently or non-covalently coupled to the mass tag, the intermediate oligonucleotide being hybridisable to the second oligonucleotide sequence (thereby linking the mass tag to the second oligonucleotide sequence).

63. The probe molecule according to claim 61, wherein the second oligonucleotide sequence is covalently or non-covalently coupled to the mass tag.

64. The probe molecule according to any of claims 61 to 63, wherein the mass tag is cleavable from the probe molecule.

65. The probe molecule according to claim 64, wherein the mass tag is photocleavable, collision cleavable or acid cleavable.

66. The probe molecule according to any of claims 61 to 65, wherein the multimeric oligonucleotide is a branched oligonucleotide, a comb oligonucleotide or dendrimeric oligonucleotide.

67. The probe molecule according to any of claims 61 to 66, wherein the predetermined target and the mass tag are located on different branches of the multimeric oligonucleotide.

68. The probe molecule according to any of claims 61 to 67, wherein the mass tag is a polyamide, for example a peptide.

69. The probe molecule or multimeric polyamide according to any of claims 43 to 68, wherein the polyamide is a peptide comprising or consisting of any one or more of the sequences: Alanine – Piperazin-1-ylacetic acid – Alanine, Alanine – Piperazin-1-ylacetic acid – Valine, Valine – Piperazin-1-ylacetic acid – Alanine, Valine – Piperazin-1-ylacetic acid – Valine, SEQ ID NOs 1-10, 12-15 and 19-20, the peptides shown in Figs 4 and 5, and modifications of any of these peptides (for example, as described herein).

70. The probe molecule or multimeric polyamide according to any of claims 43 to 69, for use in mass spectrometry (MS), for example MALDI MS/MS.

71. A probe molecule or multimeric polyamide according to any of claims 43 to 70, in which the probe molecule or multimeric polyamide comprises an oligonucleotide sequence which binds to a target oligonucleotide sequence of an oligonucleotide-labelled ligand.

72. The probe molecule or multimeric polyamide according to claim 71, in which the oligonucleotide-labelled ligand is a nucleic acid aptamer, a peptide, an antibody or a carbohydrate capable of binding to a specimen target.

73. Use of a probe molecule or multimeric polyamide as defined in any of claims 43 to 72 in a method as defined in any one of claims 1 to 42.

74. A kit comprising a plurality of probe molecules or multimeric polyamides according to any one of claims 43 to 72 together with instructions for their use.

75. A method, probe molecule, multimeric polyamide or kit substantially as described herein with reference to the accompanying figures.

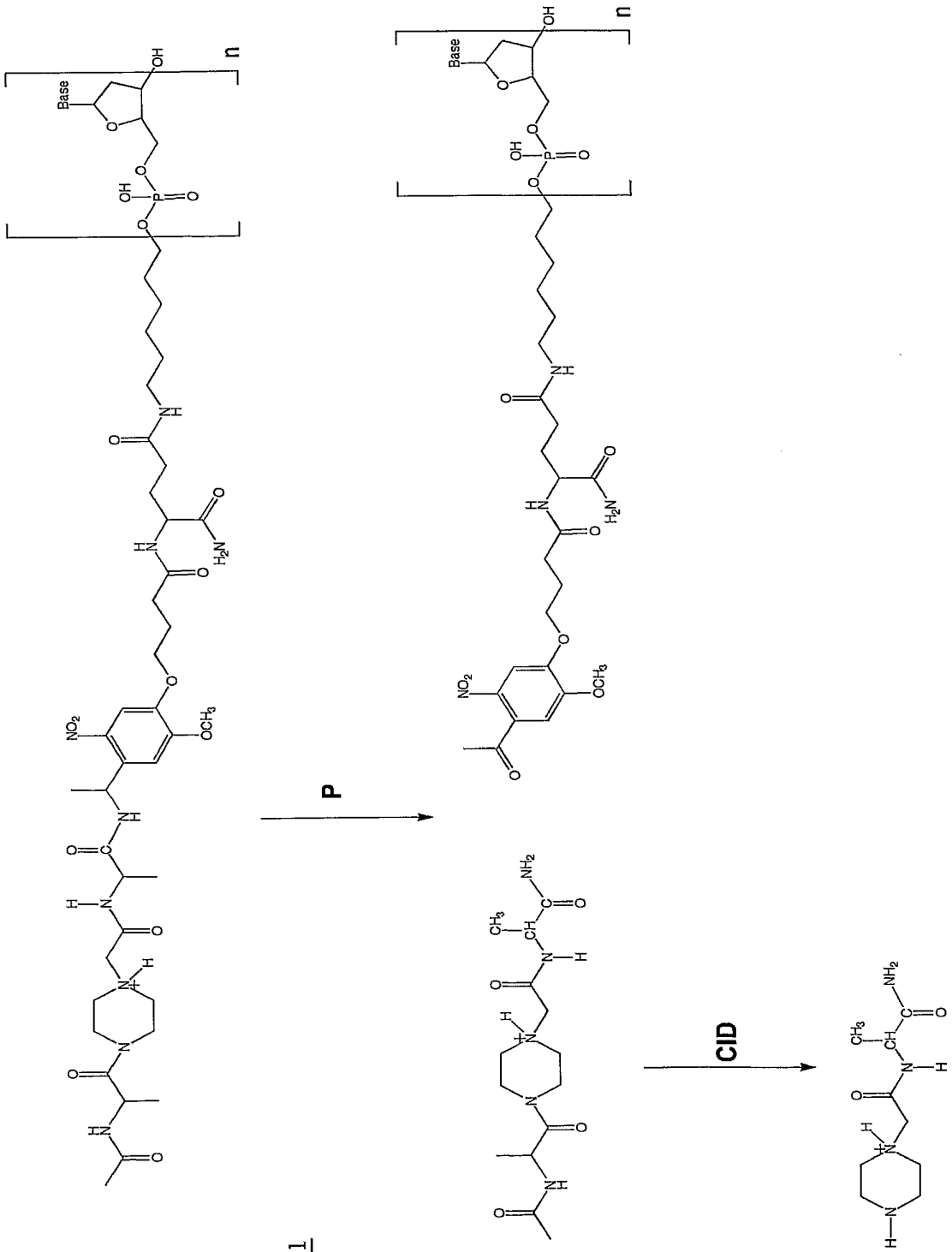


Fig. 1

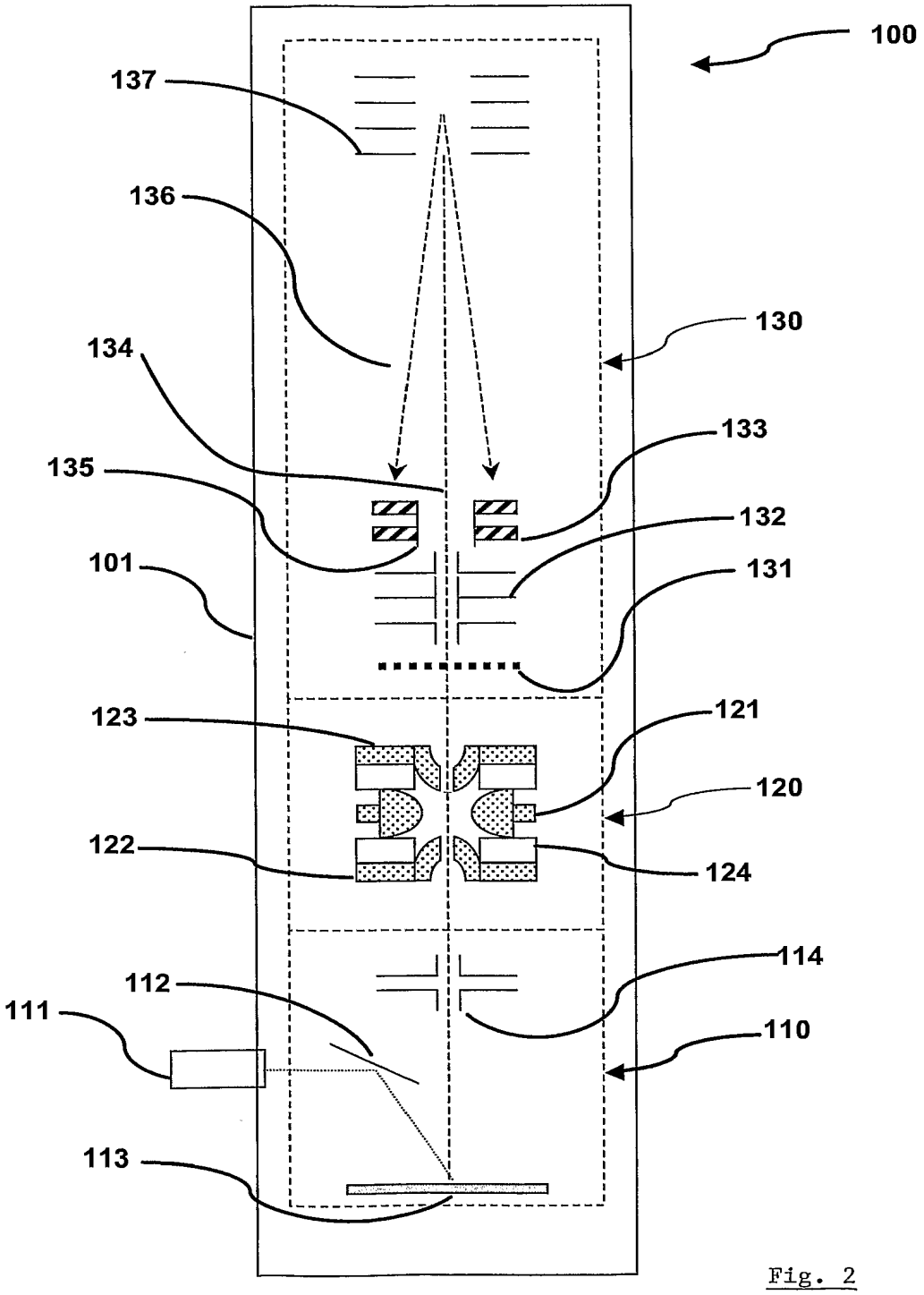

100
137
136
130
134
133
135
132
101
131
123
121
120
122
124
114
112
111
110
113

Fig. 2

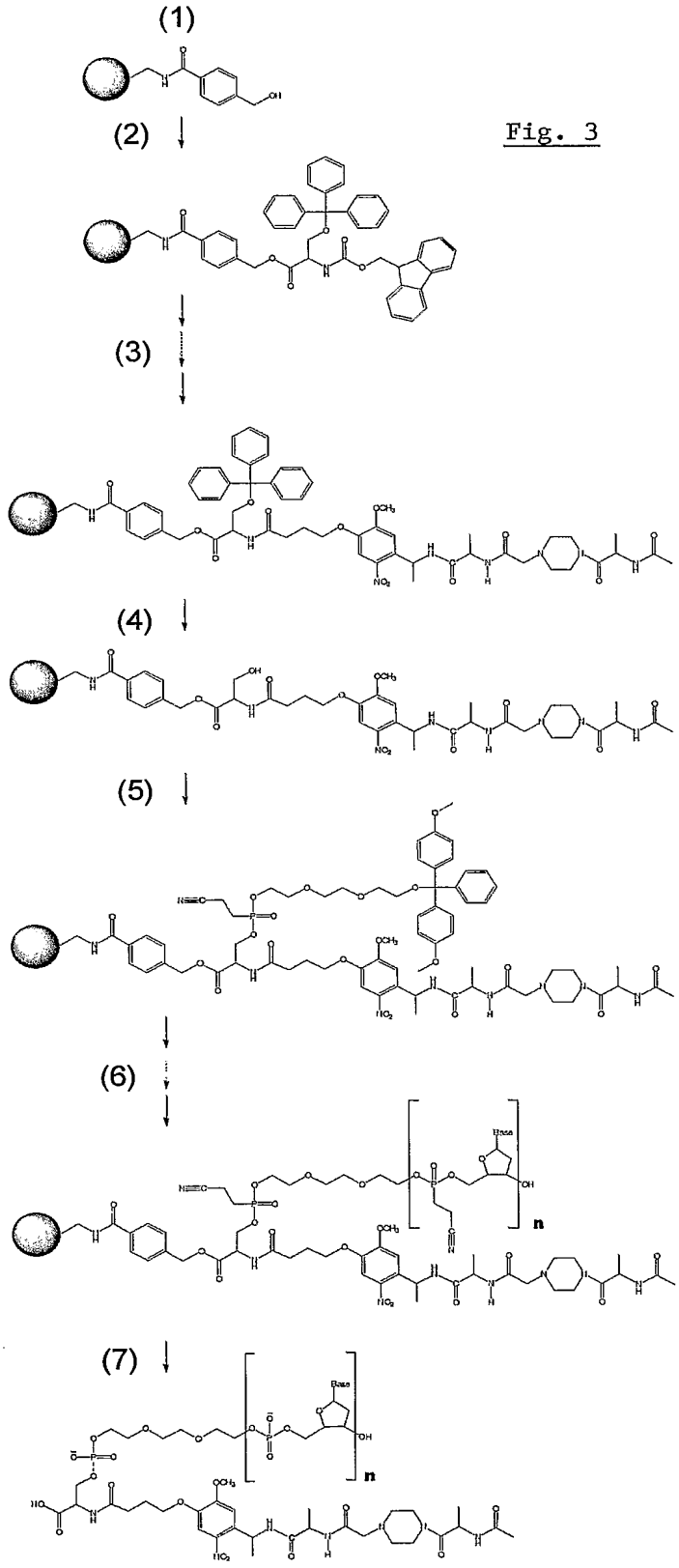


Fig. 3

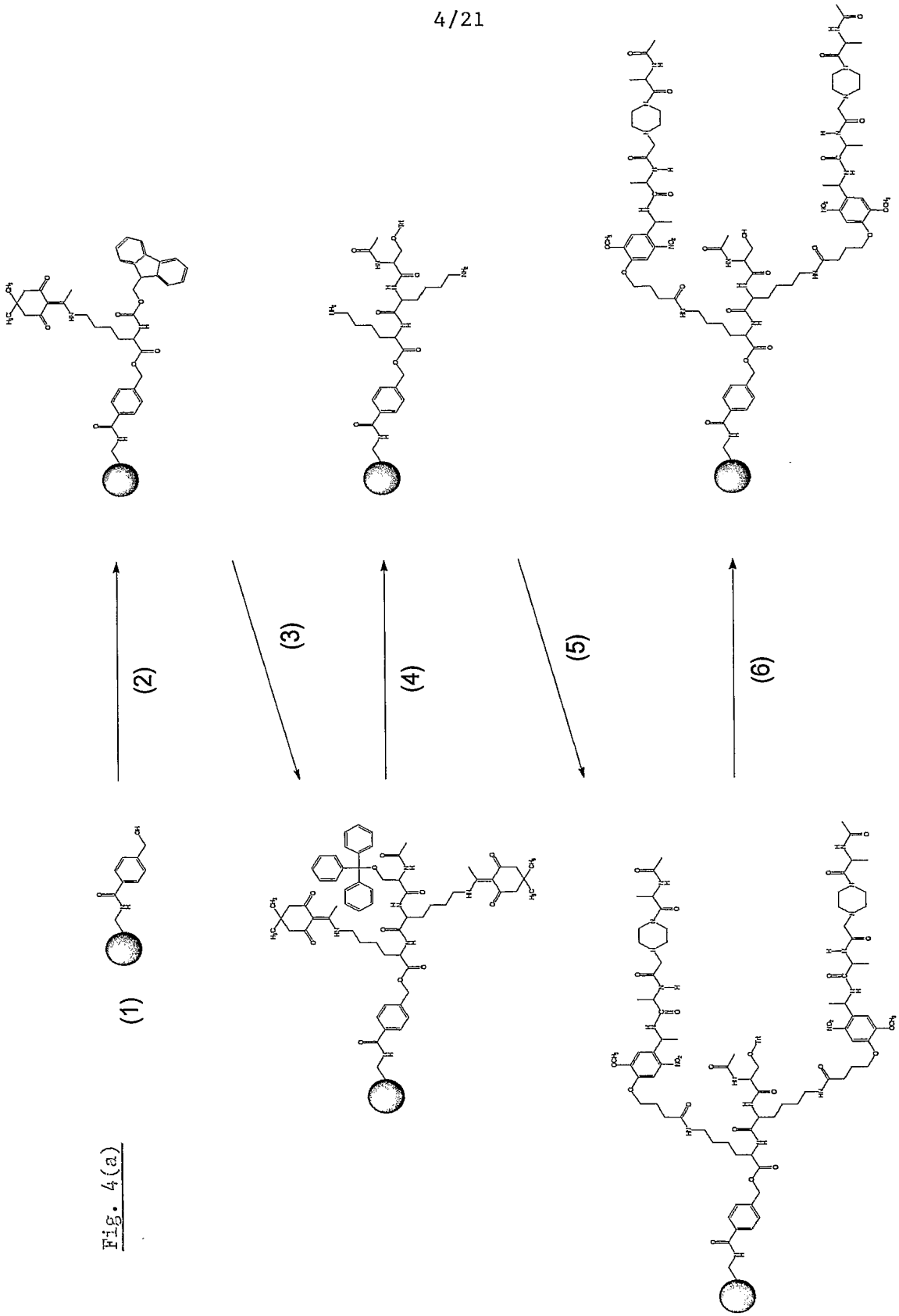


Fig. 4(a)

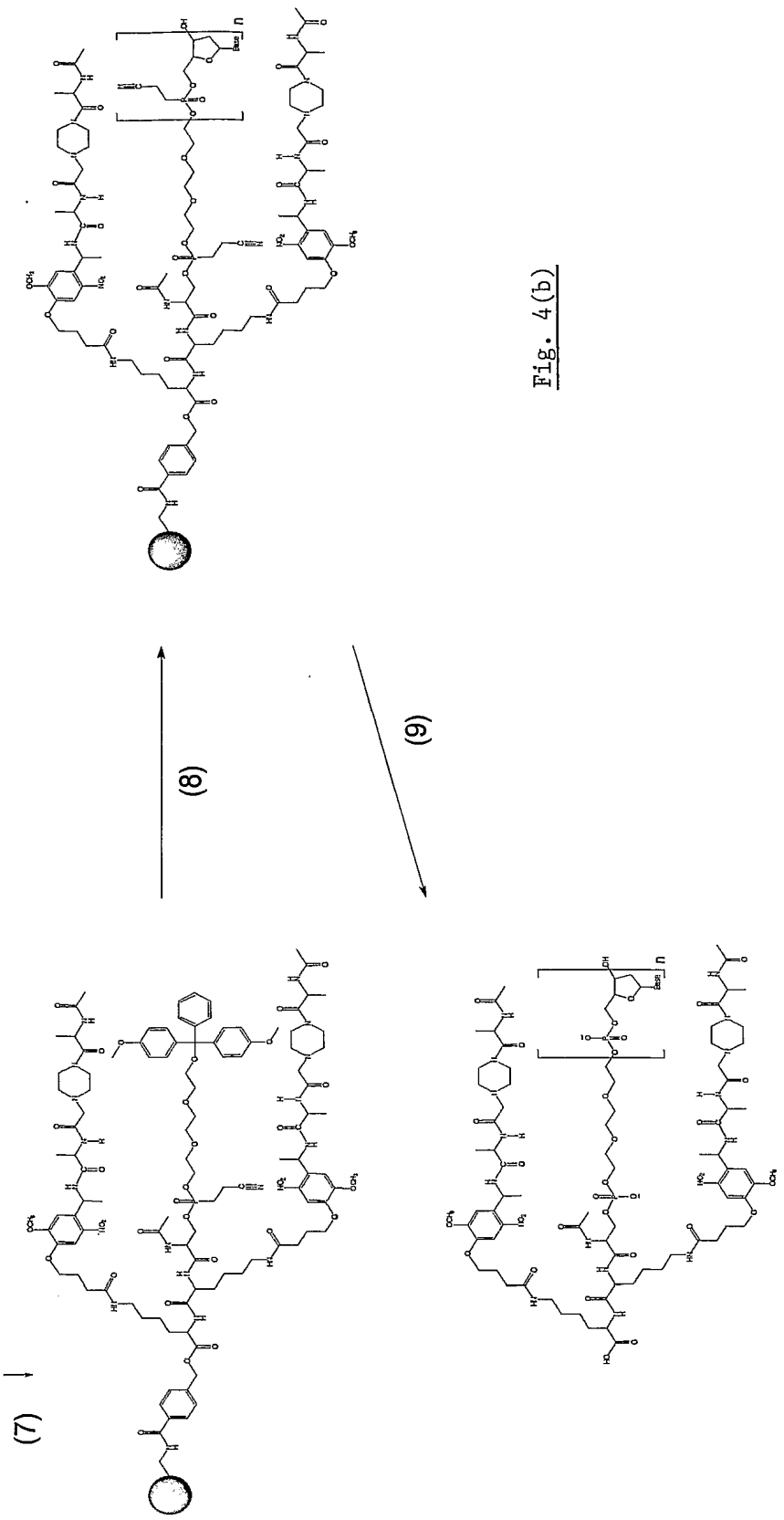


Fig. 4(b)

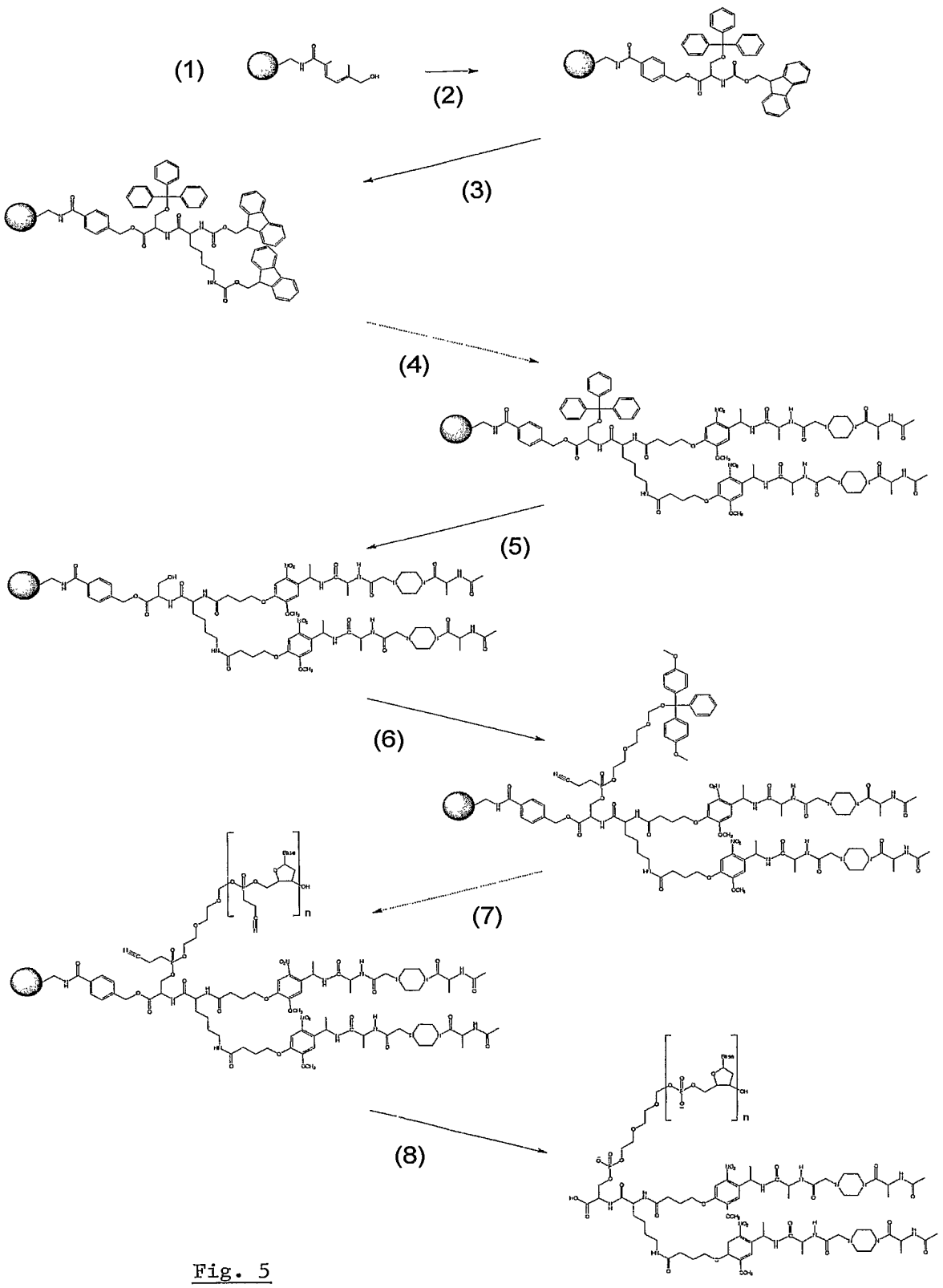


Fig. 5

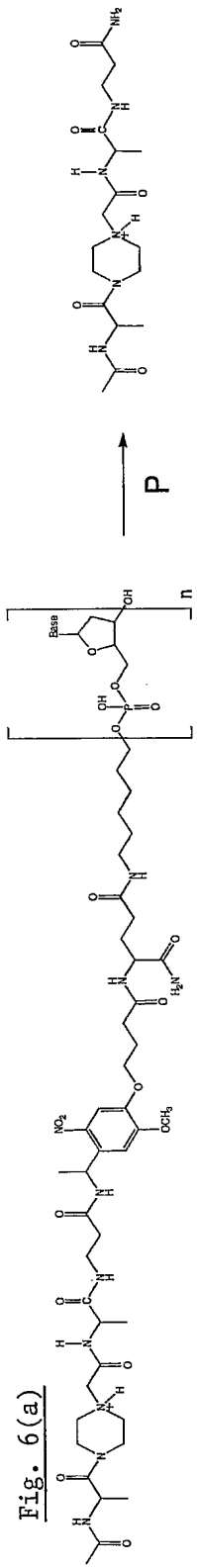


Fig. 6(a)

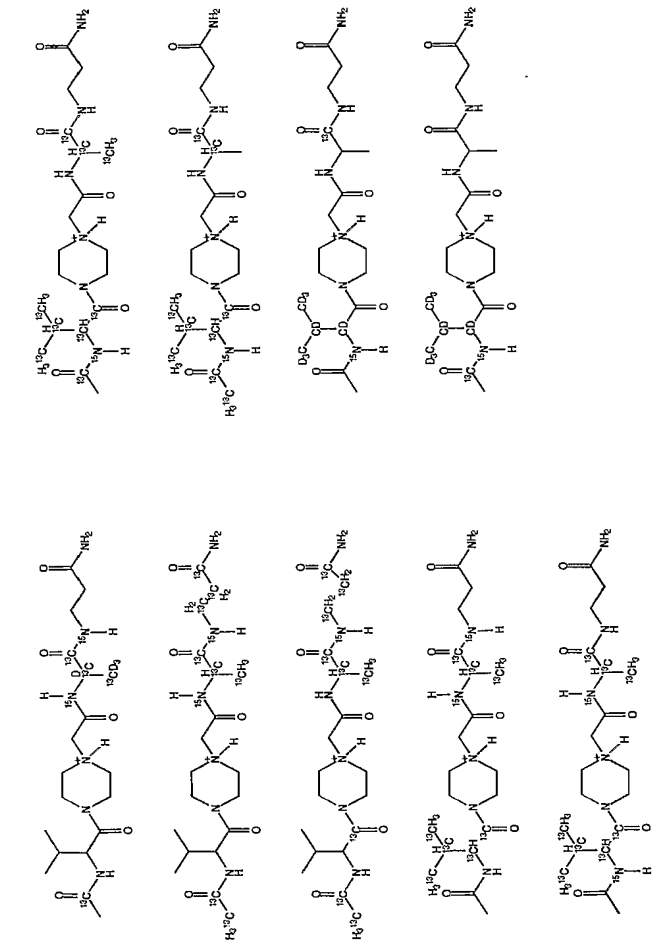


Fig. 6(b)

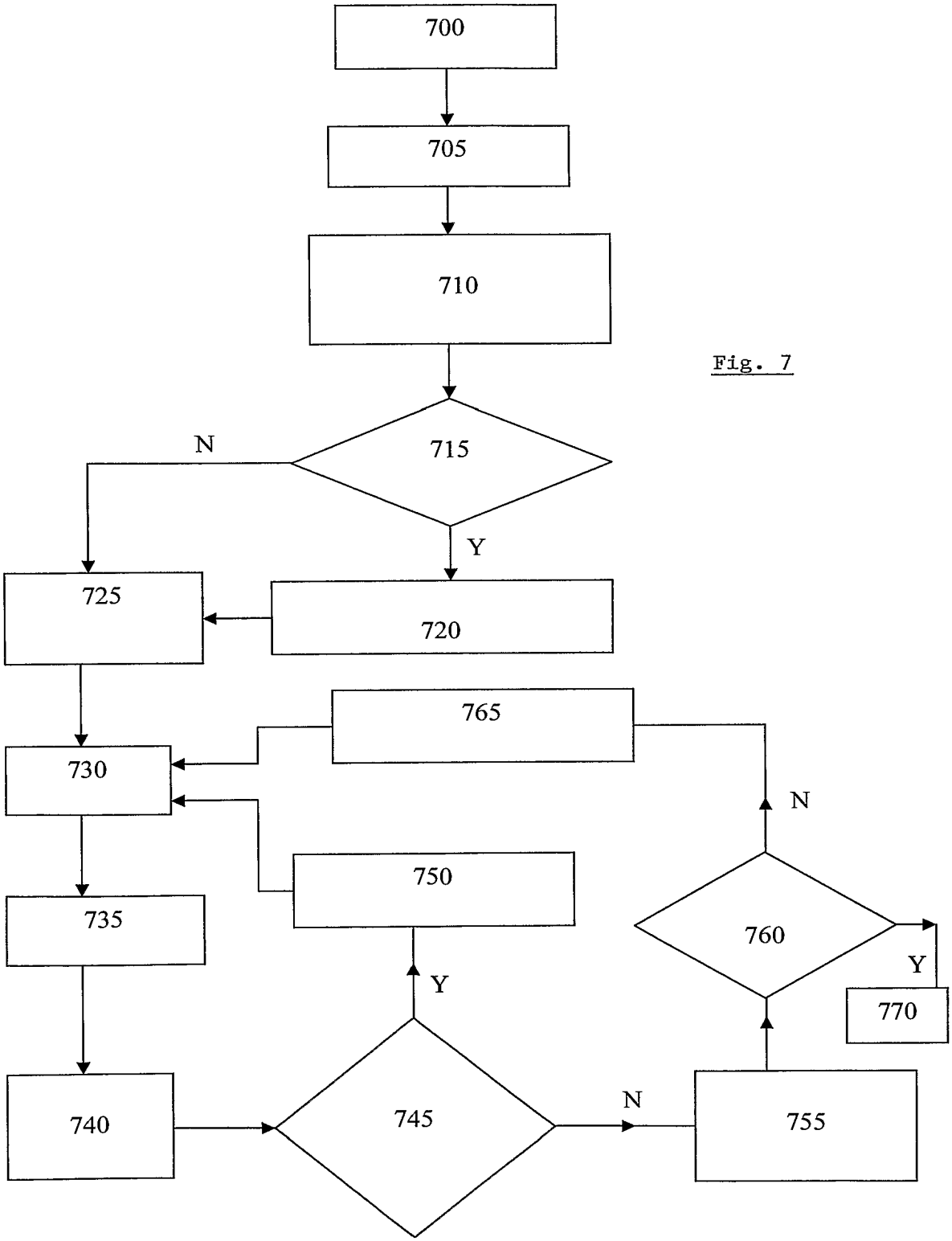

700
705
710
715
N
Y
725
720
765
730
N
750
735
760
Y
770
Y
740
745
N
755

Fig. 7

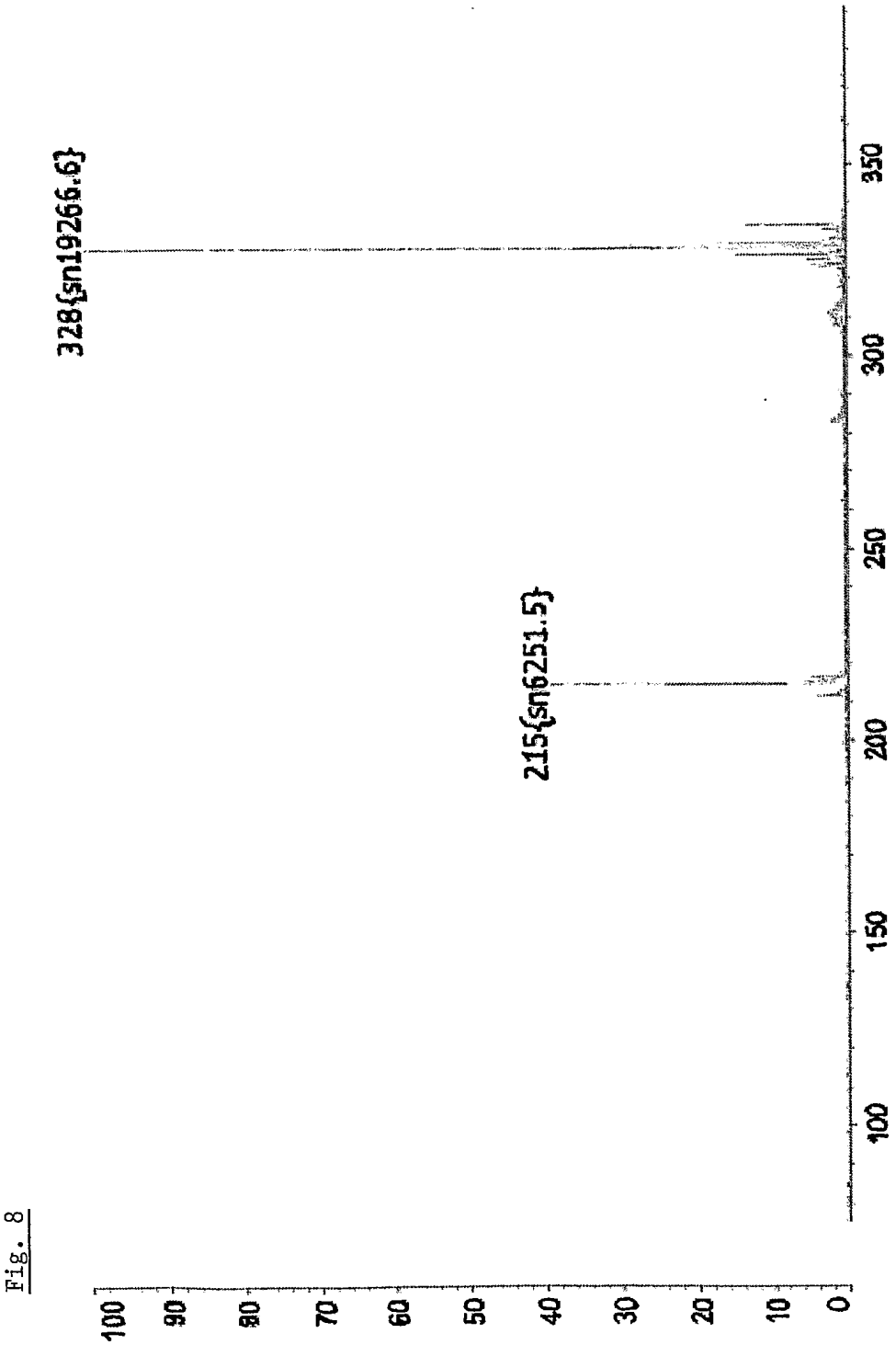

328{sn19266.6}
215{sn6251.5}
100
150
200
250
300
350
0
10
20
30
40
50
60
70
80
90
100

Fig. 8

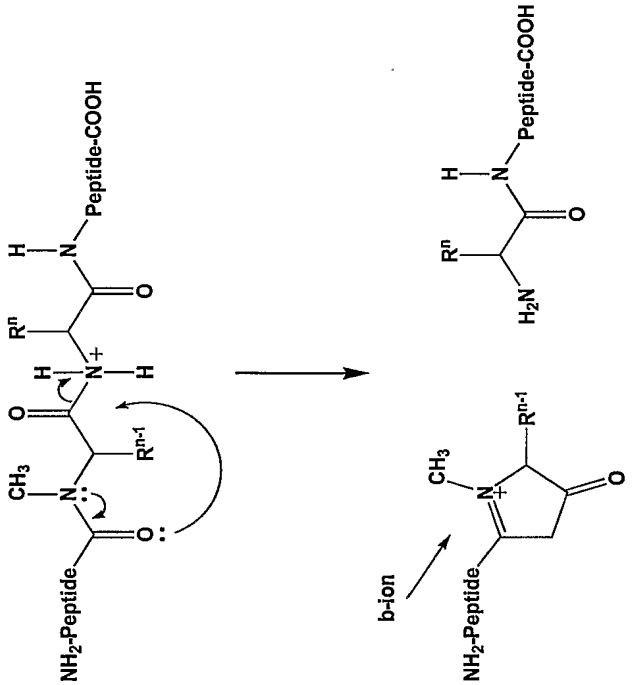

Peptide-COOH
R^n
R^{n-1}
CH_3
NH_2-Peptide
b-ion
H_2N

Fig. 9(b)

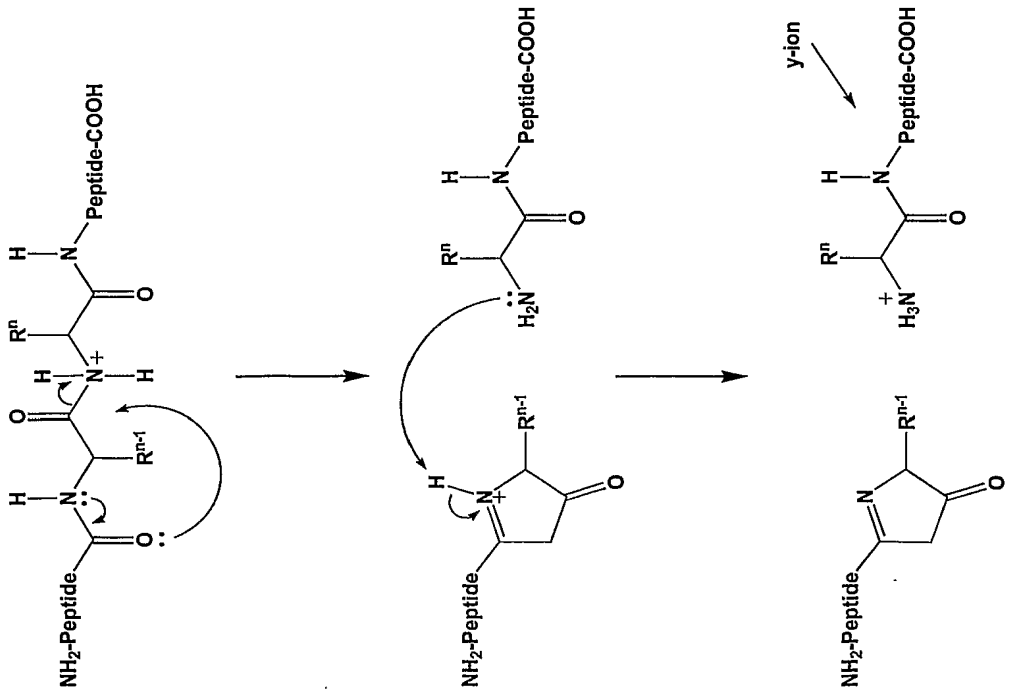

Peptide-COOH
R^n
R^{n-1}
NH_2-Peptide
H_2N
y-ion
H_3N

Fig. 9(a)

Fig. 10(a)

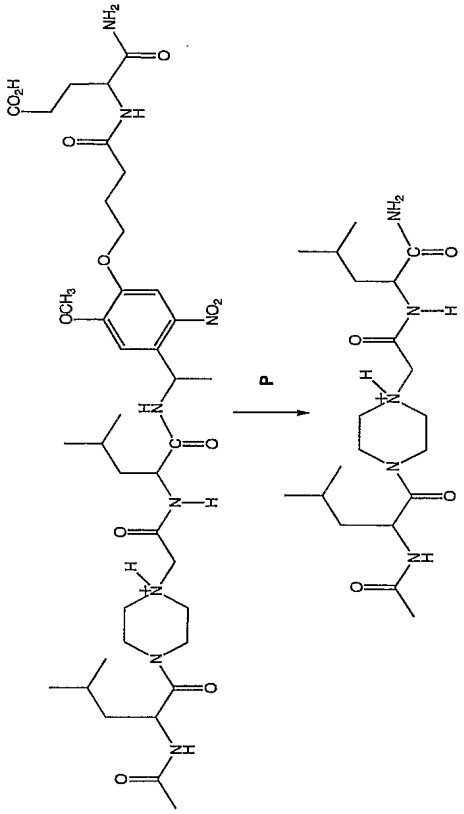


Fig. 10(b)

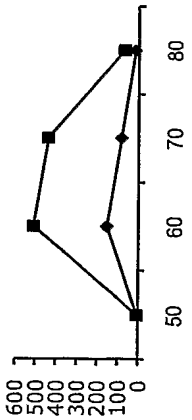


Fig. 10(c)

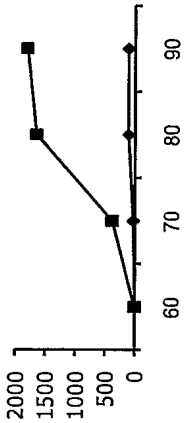


Fig. 10(d)

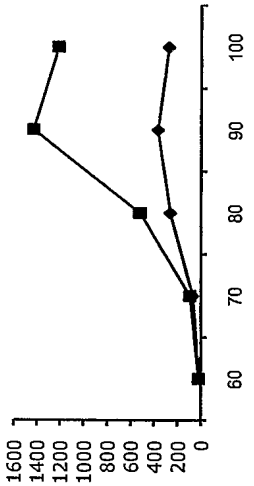


Fig. 11(a)

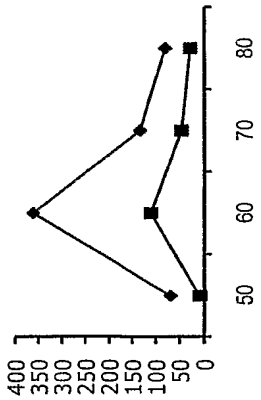


Fig. 11(b)

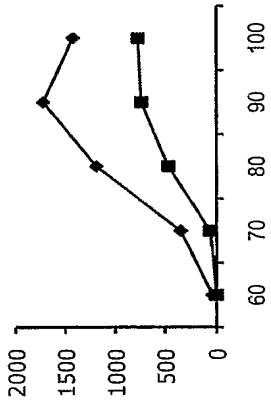


Fig. 11(c)

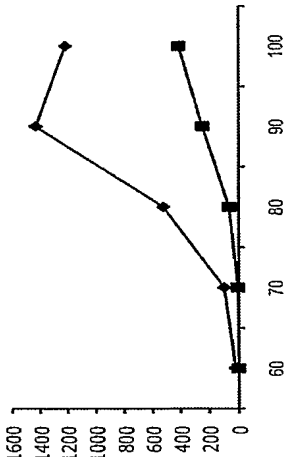


Fig. 11(d)

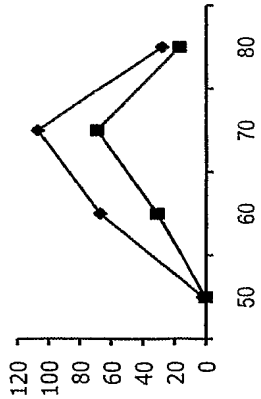


Fig. 11(e)

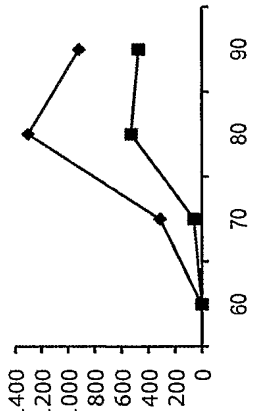
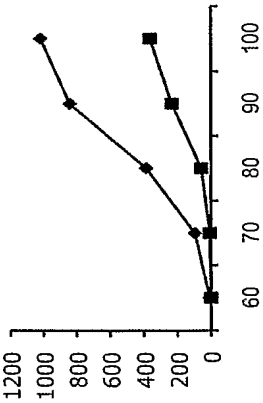


Fig. 11(f)



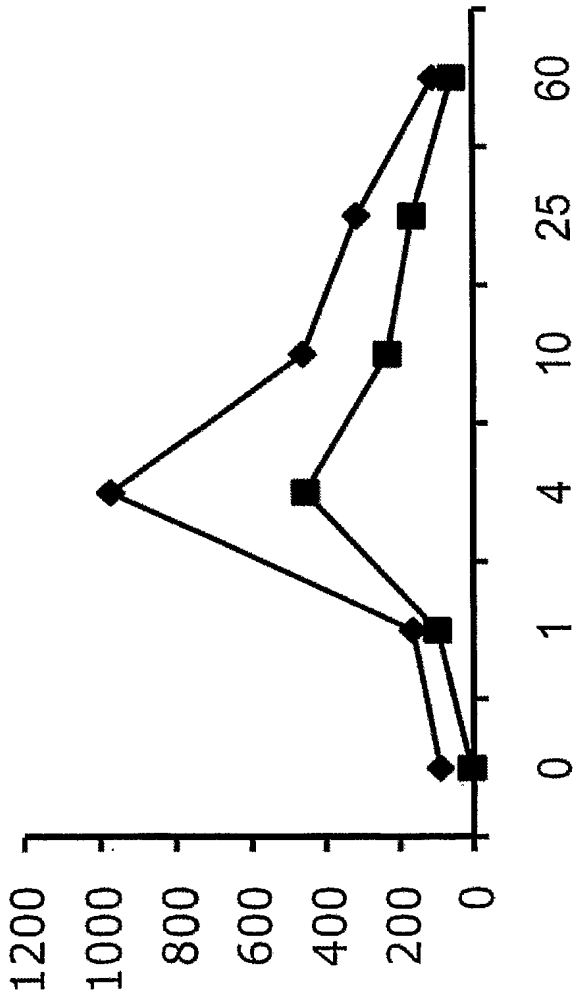


Fig. 12

Fig. 13(a)

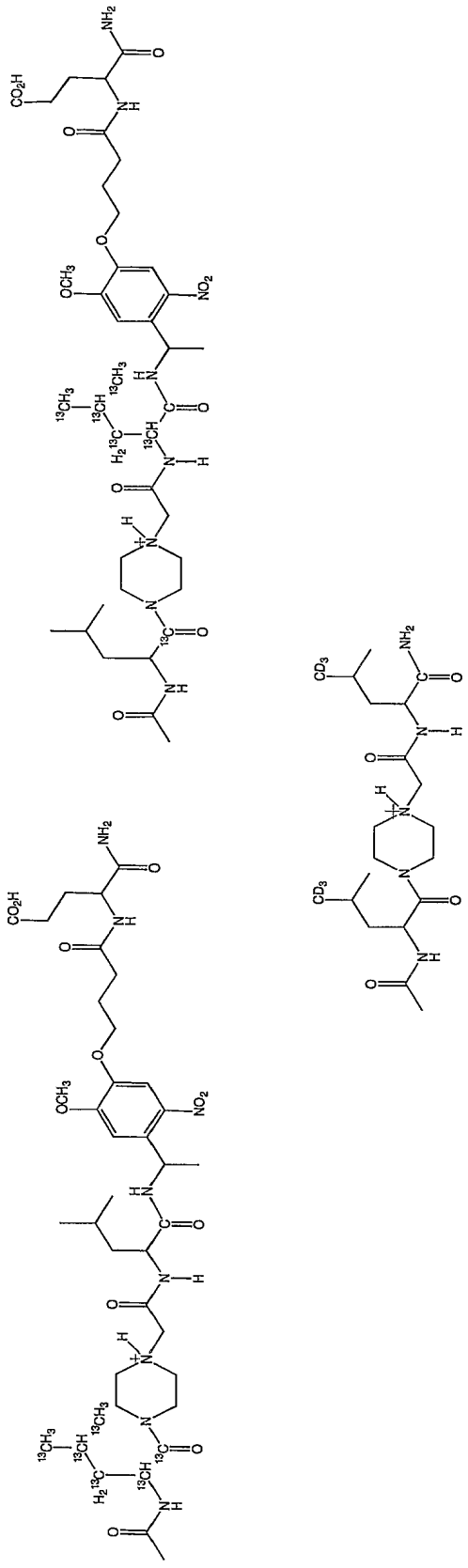


Fig. 13(b)

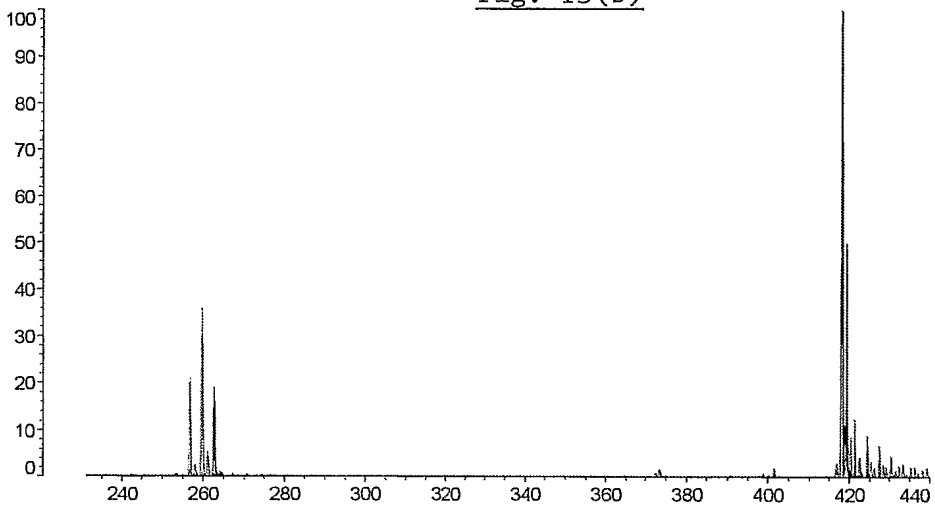


Fig. 13(c)

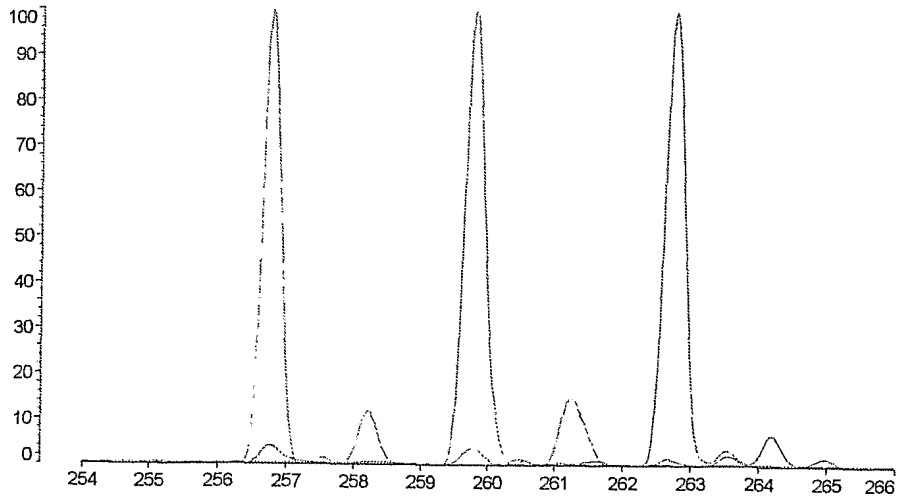


Fig. 14(a)

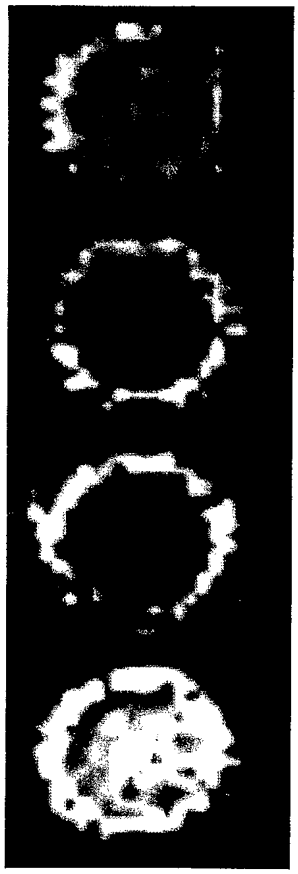


Fig. 14(b-i)

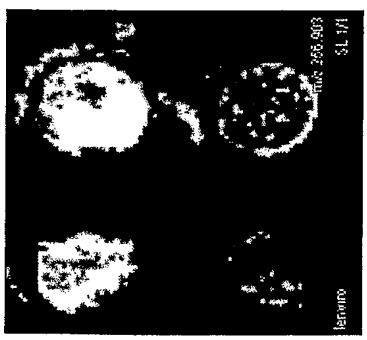


Fig. 14(b-ii)

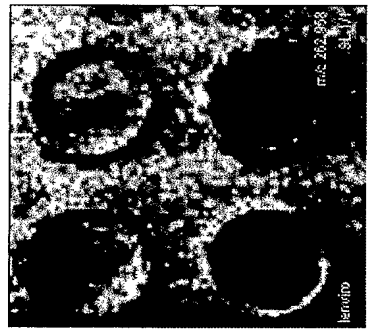


Fig. 14(b-iii)

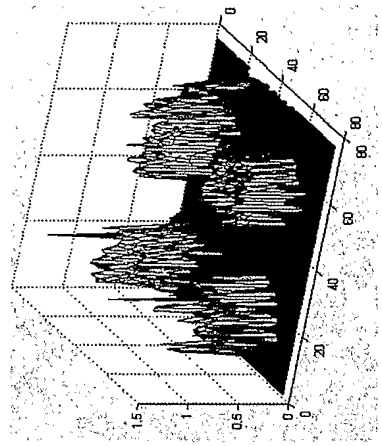
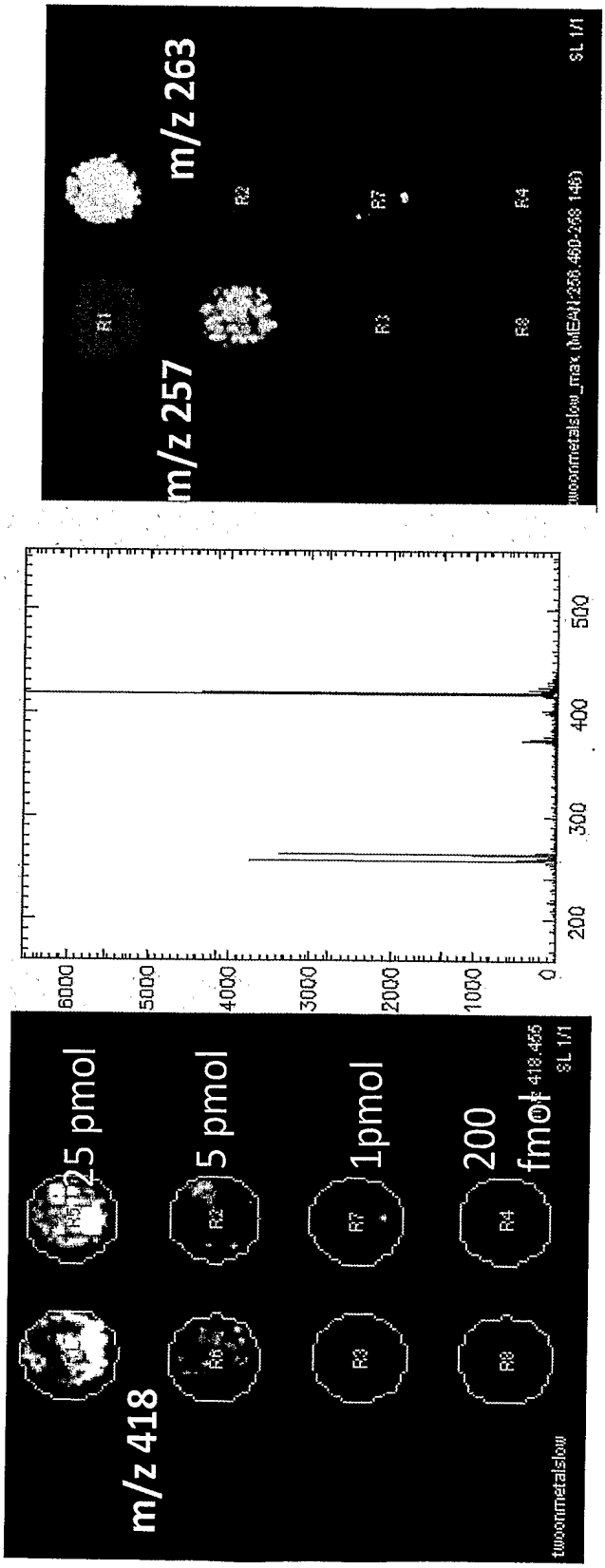


Fig. 15(a)


m/z 418
25 pmol
5 pmol
1pmol
200 fmol
m/z 257
m/z 263
6000
5000
4000
3000
2000
1000
0
200
300
400
500

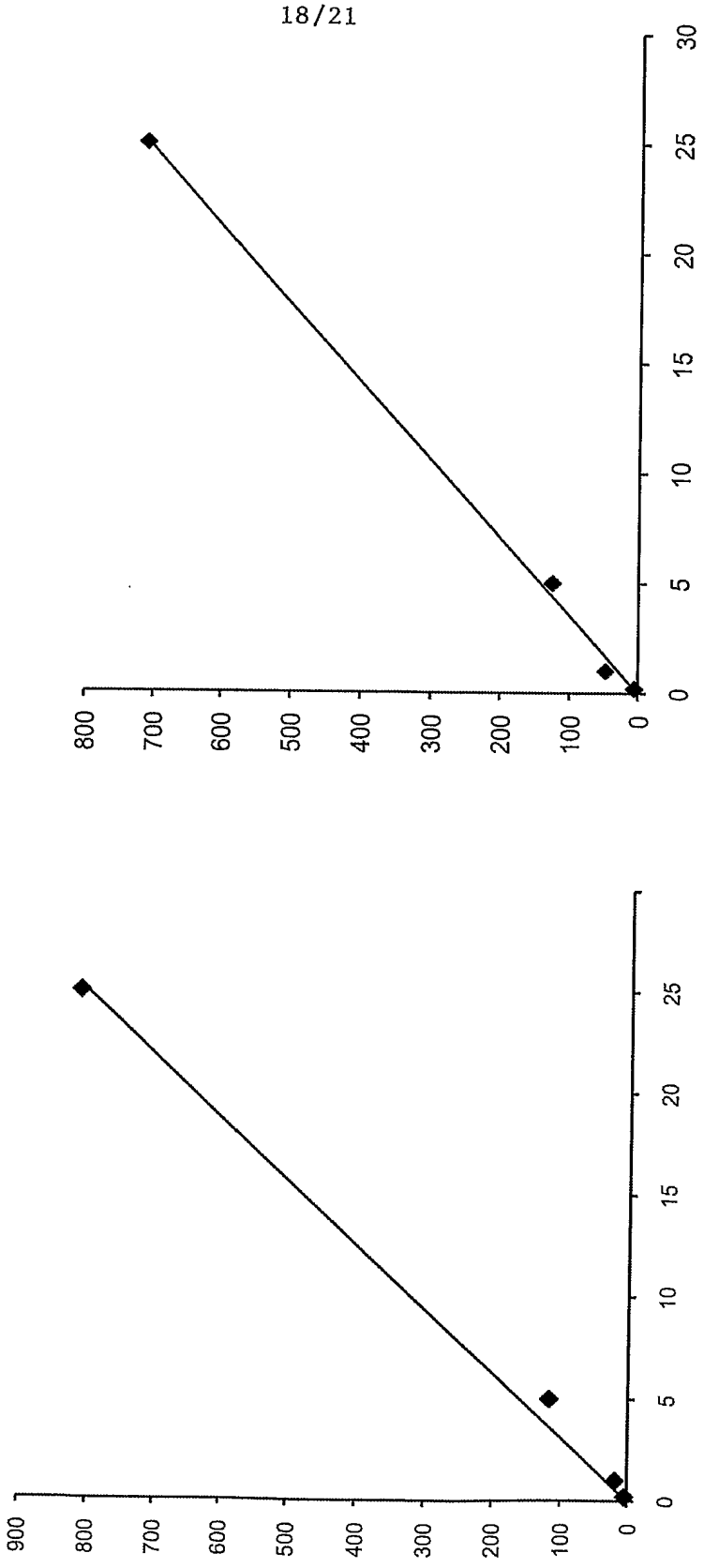

0
100
200
300
400
500
600
700
800
900
0
5
10
15
20
25
0
100
200
300
400
500
600
700
800
0
5
10
15
20
25
30

Fig. 15(b)

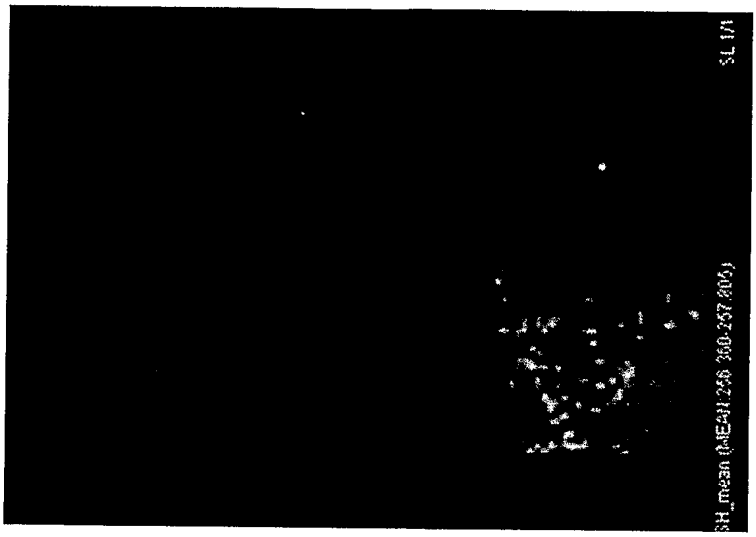


Fig. 16

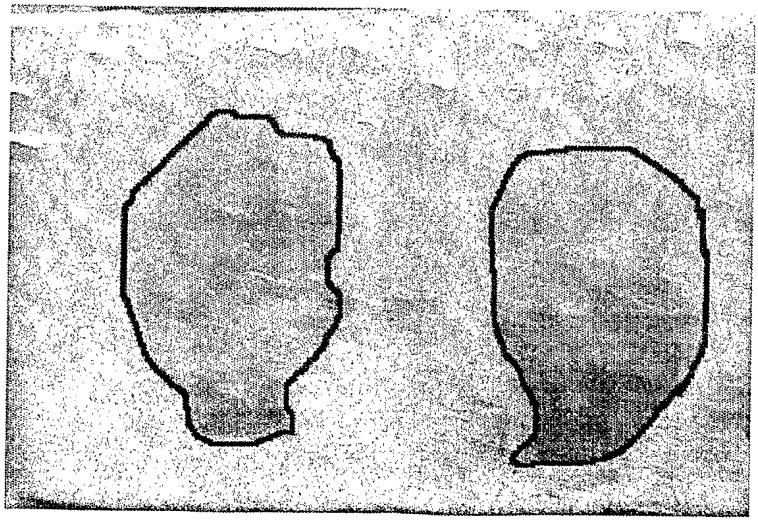


Fig. 17

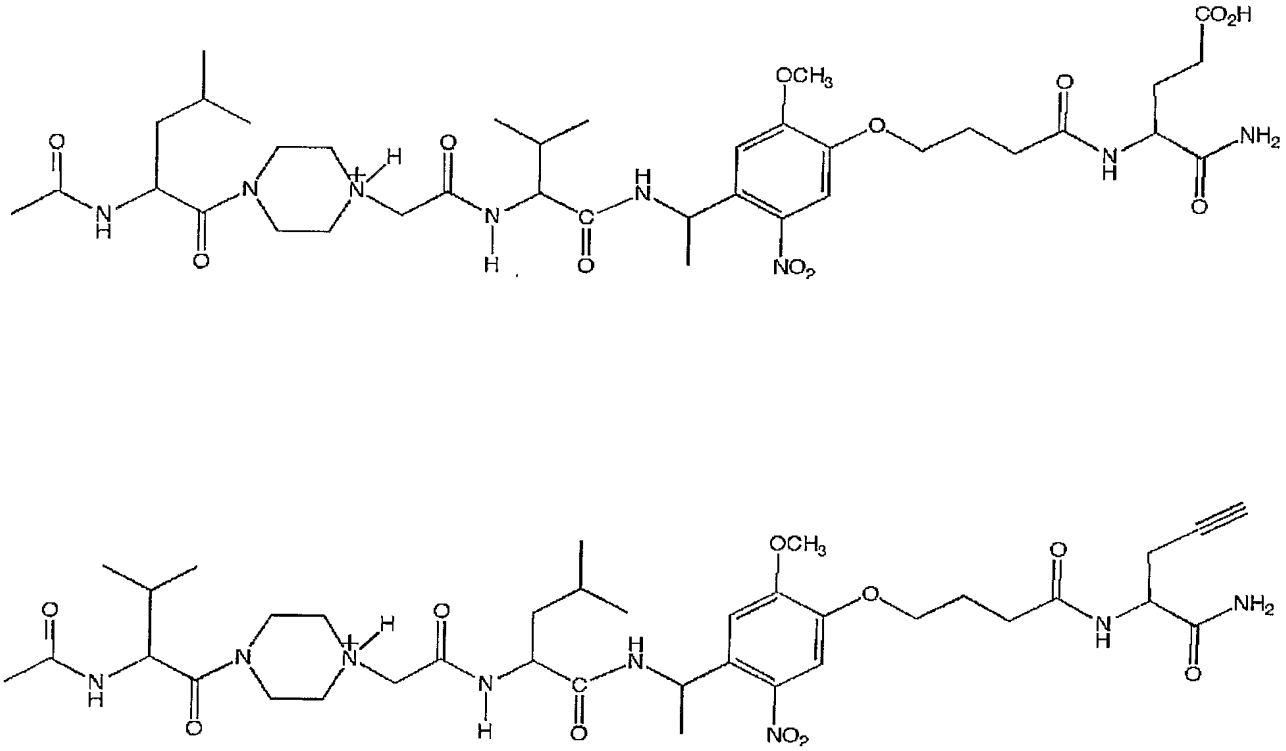
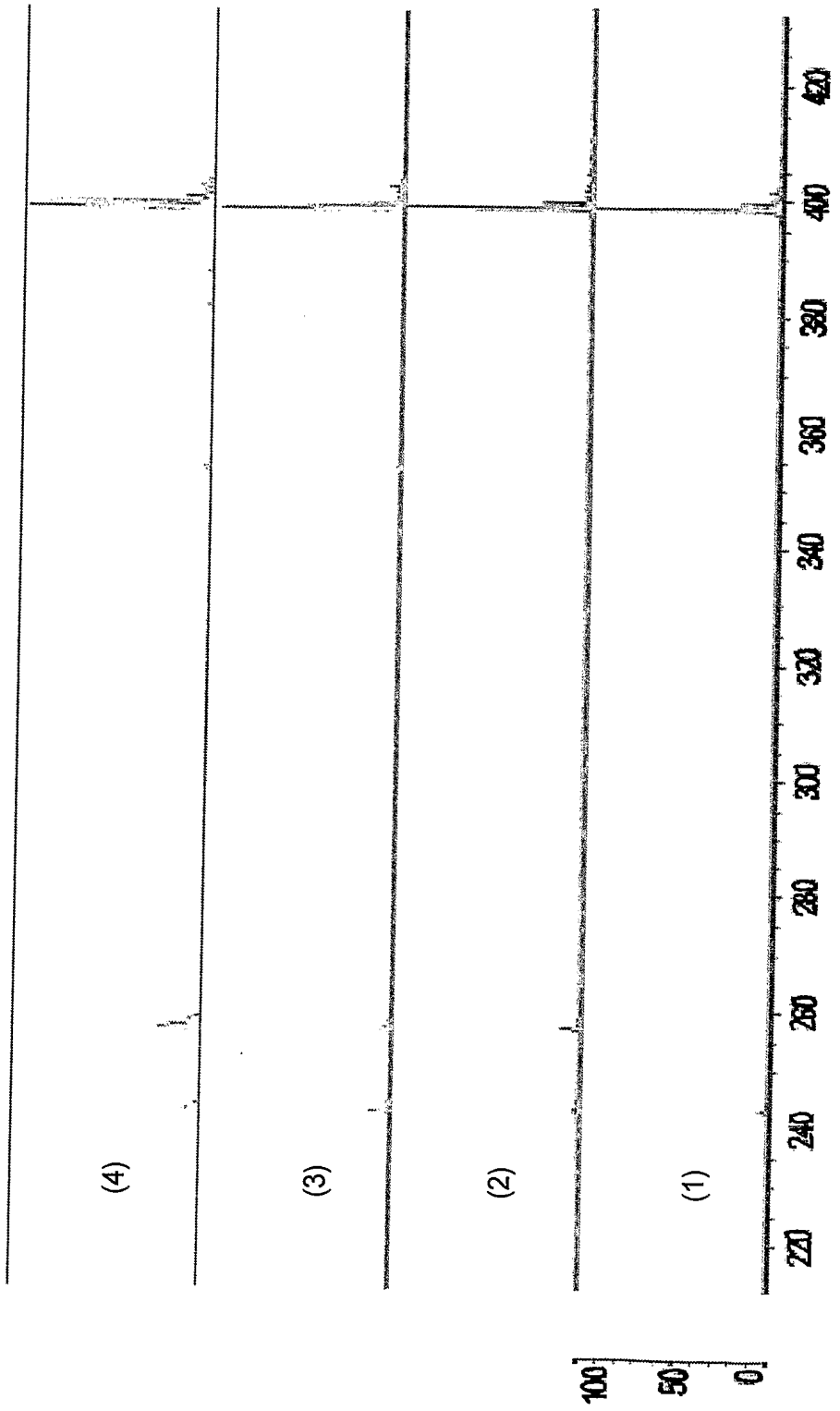


Fig. 18



INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2009/001583

A. CLASSIFICATION OF SUBJECT MATTER
INV. G01N33/68 C12Q1/68

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G01N C12Q

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2007/031717 A (ELECTROPHORETICS LTD [GB]; SCHAEFER JUERGEN [DE]) 22 March 2007 (2007-03-22) claims	1-75
X	WO 2006/017208 A (APPLERA CORP [US]) 16 February 2006 (2006-02-16) claims	1-75
X	US 2006/166201 A1 (SCHATZ PHILIPP [DE] ET AL) 27 July 2006 (2006-07-27) claims paragraph [0016]	1-75
X	US 2005/196786 A1 (LEVY SHAWN [US] LEVY SHAWN [US] ET AL) 8 September 2005 (2005-09-08) the whole document	1-75

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 September 2009	02/10/2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Routledge, Brian

4

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2009/001583

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2005/067648 A (UNIV VANDERBILT [US]; LEVY SHAWN UNIV VANDERBILT [US]; LEVY SHAWN [US]) 28 July 2005 (2005-07-28) the whole document	1-75
X	WO 2004/086050 A (XZILLION GMBH & CO KG [DE]; PROTEOME SCIENCES PLC [GB]; HAMON CHRISTIA) 7 October 2004 (2004-10-07) the whole document	1-75
Y	WO 03/025576 A (XZILLION GMBH & CO KG [DE]; THOMPSON ANDREW HUGIN [GB]; HAMON CHRISTIA) 27 March 2003 (2003-03-27) cited in the application the whole document	1-75
X	WO 01/68664 A (BRAX GROUP LTD [GB]; SCHMIDT GUNTER [GB]; THOMPSON ANDREW HUGIN [GB];) 20 September 2001 (2001-09-20) claims	1-75
Y	WO 97/27327 A (DARWIN MOLECULAR CORP [US]; NESS JEFFREY VAN [US]; TABONE JOHN C [US];) 31 July 1997 (1997-07-31) cited in the application the whole document	1-75
Y	WO 97/27325 A (DARWIN MOLECULAR CORP [US]; NESS JEFFREY VAN [US]; TABONE JOHN C [US];) 31 July 1997 (1997-07-31) cited in the application the whole document	1-75
Y	US 5 124 246 A (URDEA MICHAEL S [US] ET AL) 23 June 1992 (1992-06-23) cited in the application the whole document	1-75

4

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2009/001583

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 2007031717	A	22-03-2007	AU	2006290550 A1	22-03-2007
			CA	2622220 A1	22-03-2007
			EP	1926997 A1	04-06-2008
			JP	2009508135 T	26-02-2009
WO 2006017208	A	16-02-2006	CA	2572754 A1	16-02-2006
			EP	1776588 A1	25-04-2007
US 2006166201	A1	27-07-2006	AU	2003266197 A1	29-03-2004
			WO	2004022779 A2	18-03-2004
			DE	10240746 A1	18-03-2004
			EP	1534864 A2	01-06-2005
US 2005196786	A1	08-09-2005	NONE		
WO 2005067648	A	28-07-2005	NONE		
WO 2004086050	A	07-10-2004	AU	2004223631 A1	07-10-2004
			CA	2520297 A1	07-10-2004
			EP	1606623 A2	21-12-2005
			JP	2006523305 T	12-10-2006
			US	2007023628 A1	01-02-2007
WO 03025576	A	27-03-2003	CA	2460131 A1	27-03-2003
			DE	60223696 T2	29-01-2009
			DK	1425586 T3	11-02-2008
			ES	2296996 T3	01-05-2008
			JP	4290003 B2	01-07-2009
			JP	2005503557 T	03-02-2005
			PT	1425586 E	31-12-2007
			US	2005048489 A1	03-03-2005
WO 0168664	A	20-09-2001	AT	364841 T	15-07-2007
			AU	4083401 A	24-09-2001
			AU	2001240834 B2	17-11-2005
			CA	2403114 A1	20-09-2001
			CN	1427953 A	02-07-2003
			DE	60128900 T2	07-02-2008
			DK	1275004 T3	15-10-2007
			EP	1275004 A2	15-01-2003
			ES	2287104 T3	16-12-2007
			IL	151732 A	03-11-2008
			JP	2003529059 T	30-09-2003
			NO	20024344 A	14-11-2002
			PT	1275004 E	14-08-2007
			US	2003194717 A1	16-10-2003
			US	2008167197 A1	10-07-2008
			US	2008145850 A1	19-06-2008
WO 9727327	A	31-07-1997	AT	187501 T	15-12-1999
			AU	717330 B2	23-03-2000
			AU	1707997 A	20-08-1997
			BR	9707060 A	28-12-1999
			CA	2243989 A1	31-07-1997
			CN	1212019 A	24-03-1999
			CN	1515541 A	28-07-2004
			CZ	9802183 A3	16-12-1998
			DE	69700902 D1	13-01-2000

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2009/001583

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9727327 A		DE 69700902 T2	20-07-2000
		EP 0850320 A2	01-07-1998
		ES 2143848 T3	16-05-2000
		GR 3032843 T3	31-07-2000
		HU 9900459 A2	28-06-1999
		JP 2000507091 T	13-06-2000
		JP 2008220370 A	25-09-2008
		NZ 331042 A	28-10-1999
		PL 328272 A1	18-01-1999
		PT 850320 E	31-05-2000
WO 9727325 A	31-07-1997	AT 191515 T	15-04-2000
		AT 434054 T	15-07-2009
		AU 717945 B2	06-04-2000
		BR 9707016 A	04-01-2000
		CA 2243546 A1	31-07-1997
		CN 1212020 A	24-03-1999
		CZ 9802285 A3	16-12-1998
		DE 69701612 D1	11-05-2000
		DE 69701612 T2	31-08-2000
		DK 0840804 T3	17-07-2000
		EP 0840804 A1	13-05-1998
		EP 0962537 A2	08-12-1999
		ES 2145580 T3	01-07-2000
		GR 3033887 T3	30-11-2000
		HU 9901856 A2	28-09-1999
		JP 2000503845 T	04-04-2000
		JP 2008200040 A	04-09-2008
		NZ 331043 A	28-01-1999
		PL 328239 A1	18-01-1999
		PT 840804 E	31-10-2000
US 5124246 A	23-06-1992	NONE	

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 July 2010 (15.07.2010)

PCT

(10) International Publication Number
WO 2010/081114 A2

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

(21) International Application Number:
PCT/US2010/020680

(22) International Filing Date:
11 January 2010 (11.01.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/193,946 12 January 2009 (12.01.2009) US

(71) Applicant *(for all designated States except US)*: **20/20 GENESYSTEMS, INC.** [US/US]; 20/20 GeneSystems, Inc., 9430 Key West Avenue, Rockville, Maryland 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **JAMES, William, M.** [US/US]; 11703 Judy Place, Potomac, Maryland 20854 (US). **RAIT, Vladimir** [US/US]; 4838 Cloister Drive, Rockville, Maryland 20852 (US).

(74) Agents: **MORALES, Joseph, L.** et al.; Whiteford, Taylor & Preston LLP, Seven Saint Paul Street, Baltimore, Maryland 21202-1636 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, KM, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (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, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

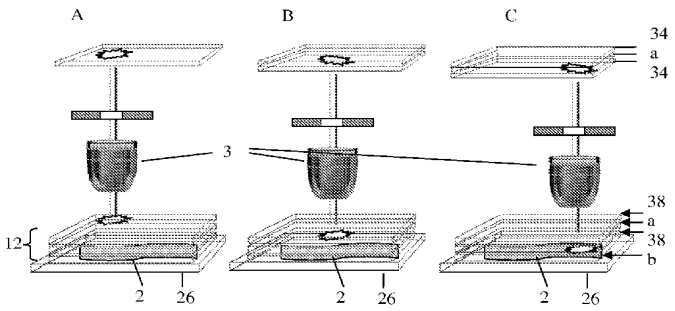
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

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

(54) Title: OLIGONUCLEOTIDE-COATED AFFINITY MEMBRANES AND USES THEREOF

Figure 15



(57) **Abstract**: A method of analyzing tissue sections in a manner that provides information about the presence and expression levels of multiple biomarkers at each location within the tissue section. The method comprises the preparation of membranes having covalently bound oligonucleotides and the use of those membranes for evaluation of various markers in the sample. The membranes may be arranged in stacks, wherein each layer has a different oligonucleotide capture strand. Transfer oligonucleotides complementary to the capture strands are attached through a cleavable bond to antibodies that recognize and bind to specific biomarkers present in the tissue sample. The tissue sample is exposed to the antibody-transfer strand conjugate and then treated with a cleaving reagent. Upon cleavage, the transfer strand migrates through the stack and binds to the capture strand. The level of expression of the biomarker may be determined by measuring expression of a reporter on the transfer strand.

WO 2010/081114 A2

Oligonucleotide-Coated Affinity Membranes and Uses Thereof

CROSS-REFERENCE TO RELATED APPLICATION

This application is a continuation in part, and claims the benefit of, U.S. Patent Application No. 61/193,946 filed January 12, 2009. That application is incorporated herein in its entirety.

GOVERNMENT SUPPORT

This invention was made with government support under Cooperative Agreement #70NANB7H7028 awarded by the U.S. Department of Commerce, National Institute of Standards and Technology. The U.S. Government has certain rights in the invention.

BACKGROUND

Membranes are employed in a wide variety of biological assays and *in-vitro* diagnostics products in laboratory, field, and point-of-care settings. These include, for example, various gel blotting procedures as well as "dip-stick" lateral flow tests for pathogens and other analytes. These devices and approaches typically employ fibrous membranes made from nitrocellulose or PVDF.

More recently, an alternative to fibrous membranes has been employed for a variety of diagnostics and biodetection applications. These membranes, known as "track-etched" membranes ("TEMs") comprise thin films with discrete pores that are formed through a combination of charged-particle bombardment (or irradiation) followed by chemical etching (see photograph FIG. 17). The particle bombardment results in the formation of damaged areas in the film (tracks) which are subsequently etched to form pores with a defined size.

Recent uses of TEMs in various biological assays are described for example by Hanot et al. in "Industrial applications of ion track technology," *Nucl. Instrum. Methods Phys. Res.* Sect. B, 267: 1019-1022 (2009) and by Jones et al. in "Expanding the use of track-etched membranes," *IVD Technology* Nov./Dec/ 2002. The authors describe the employment of TEMs in biosensors, cytology, bacterial detection, and a variety of other biological fields.

The aforementioned references describe use of "off the shelf" TEMs in their native form. A few other applications use TEMs as substrates that are coated with various probes and capture moieties. For example U.S. Patent No. 5,968,745 to Thorp et al. describes a polymer electrode for detecting nucleic acid hybridization that that utilizes oligonucleotide probes bound to a single track-etched membrane via carbodiimide condensation.

One highly innovative use of TEMs for the analysis of tissue sections is the "Layered Peptide Array" ("LPA") as described by Emmert-Buck, et al. in U.S. Patent Application Pub. No.

2009/0215073 A1 (Appl. No. 12/289,736) and in *Clinica Chimica Acta* 376 (2007) 9-16 and the *Journal of Molecular Diagnostics*, Vol. 9, No. 3, July 2007 pgs. 297-304 (each of these references are incorporated herein in their entirely). An LPA is a stack of TEMs, each of which is coated with a different peptide or protein antigen; each layer in the stack has a specific affinity to a different antibody. Antibodies to target proteins are applied to the tissue section in much the same manner as in immunohistochemistry. After washing, the antibodies are released from the tissue section and passed vertically through the peptide-coated TEMs while maintaining their two-dimensional position. The antibodies are specifically captured by the target layer to which a mimic of the natural target antigen has been coated. Alternatively, Emmert-Buck et al. disclose use of a cocktail of conjugated antibodies (a primary antibody attached to a transfer or "shuttle" antibody) that can be applied to the tissue; the shuttle antibody is then cleaved and captured on a complementary affinity ligand coated upon a layer of the stack. The layers of the stack are then separated, and the transfers are read.

The Layered Peptide Array approach, which is a subset of related techniques known in the literature as "Layered Expression Scanning (LES)", significantly increases the number of markers quantifiable per tissue section. (See U.S. Patent Nos. 7,214,477, 6,969,615, 6,602,661, U.S. Patent Application No. 11/189,038; Englert CR et al., *Layered expression scanning: rapid molecular profiling of tumor samples*. Cancer Res. 2000; 1526-30; Tangrea MA et al., *Layered expression scanning: multiplex analysis of RNA and protein gels.* Biotechniques 2003; 1280-5; Gillespie JW et al. Molecular profiling of cancer. Toxicol. Pathol. 2004; 67-71; Galperin MM, et al., Multimembrane dot-blotting: a cost-effective tool for proteome analysis. Biotechniques 2004; 1046-51; Gannot G et al. Layered peptide arrays: high-throughput antibody screening of clinical samples. J. Mol. Diagn. 2005; 427-36; Patel V et al. *Profiling EGFR activity in head and neck squamous cell carcinoma by using a novel layered membrane Western blot technology.* Oral Oncol. 2005; 503-8; all incorporated herein by reference).

Importantly, LPAs and LES permit the analysis of multiple biomarkers in various 2-D samples such as tissue sections while preserving the localization of these biomarkers. In other words, when used with tissue sections this approach combines classical pathology with multiplex array based technologies. These techniques address important unmet needs in the emerging field of "Personalized Medicine" – the development and use of therapies specifically targeted to the disease characteristics of individual patients is widely predicted to be the key driver of 21st century medicine. Better diagnostics linked to drugs are anticipated

to significantly improve patient outcomes while reducing healthcare costs by avoiding prescriptions of expensive drugs that prove ineffective for many patients. Many of the new targeted therapies that have come to market in recent years cost over $75,000 per patient per year but are effective in only a narrow subset of patients (~15%) to whom they are administered. Thus, there is an urgent and compelling need for new diagnostic tests that can accurately predict whether a particular targeted therapy will work for a particular patient. Unfortunately, neither conventional techniques for tumor analysis nor newer detection technologies have proven adequate to meet this need. These techniques typically fail either in their multiplex capability or their ability to preserve the shape and morphology of the tissue section which is often needed for accurate diagnosis.

Newer multiplex technologies such as DNA microarrays or mass spectrometry (MS), as well as older ELISA based techniques, require that samples be homogenized. Yet, a typical biopsy sample would present only a small minority of infiltrating tumor cells of interest. Those cells would be surrounded by numerous other types of cells (normal cells, fibroblasts, lymphocytes, vascular cells etc.) presenting dissimilar gene expression and cell signaling profiles, which are potential biomarkers and drug targets. In a visual field of an invasive tumor, an important issue to a pathologist is whether the tumor cell population is molecularly homogeneous or whether there exist sub-clones within it with distinct transcriptomic or proteomic profiles (Uhlen M et al., 2005, A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol. Cell Proteomics. 4:1920-1932, incorporated herein by reference).

Very small tissue samples such as core needle biopsies are currently being taken from small tumors for performing a tumor-specific molecular diagnosis to enable personalized drug treatments. Additionally, advancement of personalized medicine has been stymied by a paucity of technologies that can effectively derive predictive biomarkers from diseased tissues. In the case of cancer, there is a particular need for techniques that can profile multiple biomarkers in tumor sections, since most of the newer targeted therapies interact with numerous signaling proteins (Faivre S et al., New paradigms in anticancer therapy: targeting multiple signaling pathways with kinase inhibitors. Semin. Oncol. 2006; 407-20, incorporated herein by reference.).

It would therefore be desirable to have an approach for analyzing multiple biomarkers in tissue and other biological samples that overcomes the aforementioned limitations of the LPA method by providing TEMs coated with capture probes that remain more permanently bound thereto, that can be manufactured in an efficient, cost-effective, and reproducible manner, and

that can be engineered for the properties of highly specific ligand binding in predictable and variable ways.

SUMMARY

Disclosed is method of analyzing tissue sections (and other 2-D samples) in a manner that provides information about the presence and expression levels of multiple biomarkers (or other targets) at each location within the tissue section.

In short, the method utilizes a plurality or stack of permeable layers (TEMs or gels) each having a specific oligonucleotide (capture strand) covalently bound thereto so as to create affinity layers. A plurality of antibodies (primary or secondary) are also provided, each of which is conjugated via a cleavable bond to an oligonucleotide (transfer strand) that is complementary to the capture strand. A flourophore or other detectable moiety may be attached to the transfer strand.

These conjugated antibodies are applied to the tissue section (or other sample of interest) and unbound antibodies are washed or otherwise removed. The affinity membrane stack is then applied to the tissue section. The transfer oligonucleotide is then cleaved from the conjugate antibody and migrates through the stack until the transfer oligonucleotide hybridizes to the layer coated with its complementary capture strand.

When used with tissue sections, the layers may then be analyzed by a number of imaging modalities to generate data showing the presence and expression levels of multiple biomarkers (or other targets) at a given location within the tissue section.

Included in the disclosure is a method of covalently binding fixed oligonucleotides to TEM layers that prevents their migration to adjacent layers when the transfer of corresponding transfer oligonucleotides through the stack is underway.

Also disclosed are methods that use capture oligonucleotides conjugated to layers that are formed from transparent, hydrophilic, polymeric hydrogel types of materials, such as those of natural origin (e.g. an agarose) or of synthetic origin (e.g. a polyacrylamide).

Also disclosed are methods of imaging the processed layers following transfer including a method for the serial analysis of an entire stack that need not be separated.

Also disclosed are methods of analyzing images generated with the present invention using software and the like in a manner that, when used with tissue sections, can assist a clinician with medical decision making.

With the foregoing and other objects, advantages, and features of the disclosure that will become hereinafter apparent, the nature of the disclosure may be more clearly understood by

reference to the following detailed description of the disclosure, the appended claims and to the several views illustrated in the drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Other objects and advantages of the present disclosure will be evident from the following detailed description, with like reference numbers referring to like items throughout.

Figure 1 provides enlarged perspective views of the affinity membranes (frame A) as well an enlarged illustration of the antibody-oligonucleotide conjugates (frame B) and use of the same for analysis of tissue sections (frames C-E).

Figure 2 is an enlarged illustration showing typical antibody-oligonucleotide conjugates that may be used with the affinity membranes disclosed herein.

Figure 3 shows the chemical structures of three polymers commonly used as substrates for making commercially available track-etched membranes.

Figure 4 is an illustration showing conversion products after etching of a polyester [poly(ethylene terephthalate)], a polycarbonate, or a polyimide to an oxoacid (B) in the course of the alkaline etching. Two kinds of functional groups are formed from the polymer: carboxylic acid end groups and alcohol end groups.

Figure 5 (A) shows the reaction of an aminated oligodeoxynucleotide which may be primary or secondary with an oxoacid of a track-etched poly(ethylene terephthalate) via carbodiimide condensation. (B) shows an example of that reaction with a polycarbonate membrane.

Figure 6 illustrates the structure of one of oligodeoxynucleotides used in EXAMPLES. The 24-mer (named ATP17) bears a primary amino group, attached through a particular spacer to the 3´ end, and a fluorescent label, Cy5, attached to the oligodeoxynucleotide's 5´ end. Other used modifications are also shown.

Figure 7 is a bar graph with the results of a reaction in which a transfer oligonucleotide was used to probe for a hybridized interaction with a capture oligonucleotide coated on track-etched *poly(ethylene terephthalate)* membranes in the presence (A) or absence (B) of a water-soluble carbodiimide, EDC..

Figure 8 is a bar graph showing the results of oligonucleotide (ATP17) interaction with track-etched *polycarbonate* membranes in the presence (A) or absence (B) of a water-soluble carbodiimide, EDC.

Figure 9 is a bar graph comparing the amounts of capture oligonucleotides that can be adsorbed or covalently coupled to Polyimide, poly(ethylene terephthalate) and polycarbonate track etched membranes.

Figure 10 is a bar graph showing the results of the joint (A) and sequential (B) incubation of the track-etched polycarbonate membranes with ATP17 and EDC. Bars correspond to the red fluorescence intensity detected either in the complete incubation mixture (A1) or in a membrane withdrawn after the incubation from the mixture contained only ATP17 (B1), in membrane sequential washings with MES buffer (2-4), 10% acetonitrile (5-7), 6x SSPE buffer (8), and in a pair of the resulting membranes (9 and 10).

Figure 11 shows fluorescent images of track-etched polycarbonate membranes coated with ATP17 according to the joint (A) or sequential (B) procedure of the oligodeoxynucleotide immobilization. Graphs C and D characterize the fluorescence intensity distribution along diameters of membranes A and B respectively.

Figure 12 is a bar graph comparing ATP25 interactions with the plain track-etched polycarbonate membrane (A) and the membrane with the immobilized complementary ATP60 (B). Bars correspond to the red fluorescence intensity detected in the membranes after washings with 6x SSPE buffer (1), in the last washing with that buffer (2), in following washings with 10% (3-5) and 30% (6) acetonitrile, and in the resulting membranes (7) washed again with the SSPE buffer.

Figure 13 is a bar graph showing results of the thermal dissociation of a complementary duplex formed by the fluorescein-labeled ATP95 and ATP17 immobilized on the track-etched polycarbonate membrane. Bars correspond to the green fluorescence intensity detected in the membrane upon hybridization and washings with 6x SSPE buffer (1), released at 60 °C into 25 mM phosphate buffer, pH 7 (2) and following hot washings with the buffer (3 and 4), and retained on the membrane (5).

Figure 14 illustrates the results of an induced release of a the Cy5-labeled oligodeoxynucleotide induced from an antigen-antibody-streptavidin construct and the oligonucleotide spontaneous distribution in a stack of membranes that contained single membrane with the complementary immobilized oligonucleotide and two membranes with non-complementary oligonucleotides. Fluorescent images of membranes (0-8) and Whatman 3MM paper (9) kept in a stack for 1 hr at 45°C are shown. Numbers below images correspond to positions of the membranes in the stack. 0 - Track-etched, coated with poly(vinylpyrrolidone) polycarbonate membrane; 1 – irregular shaped track-etched polycarbonate membrane with a complex composed of the adsorbed rabbit IgG, goat anti-rabbit Ab conjugated with streptavidin, and transfer oligonucleotide ATP73; layers 2, 4, 6, and 8 – as 0; 3, 5, and 7 – track-etched polycarbonate membranes with the immobilized

capture oligonucleotides ATP62, ATP61, and ATP 60, respectively. The Whatman filter was wetted with 4x SSC buffer contained 0.1% SDS, and 25 mM TCEP.

Figure 15 is an illustration showing serially imaging of a transparent membrane stack that may be optionally used with the affinity membranes disclosed herein.

Figure 16 is a scanning electron micrograph of a typical track-etched membrane.

Figure 17 is a set of image data and quantitative analysis to demonstrate a comparison of non-multiplex to 3-fold multiplex analysis of three target analytes in a pathology tissue section, and includes a quantitative analysis of the intensity of signals in the images.

Figure 18 is a set of image data and quantitative analysis to demonstrate a comparison of 2-fold multiplex to 3-fold multiplex analysis of three target analytes in a pathology tissue section, and includes a quantitative analysis of the intensity of signals in the images.

Figure 19 is a demonstration of the replication of an intensity pattern of an assay target in a tissue specimen on three detection layers after transfer of transfer oligonucleotides, and includes a quantitative analysis of the intensity of signals in the images.

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

1. Overview

A preferred embodiment of the present invention is illustrated in FIG. 1. With reference to frames A and B of FIG. 1 this embodiment preferably includes an affinity membrane stack 12 which comprises multiple layers of track-etched membranes ("TEMs") 14 (a-c) each coated with a different, specific oligonucleotide 16 (a-c), typically a capture strand, which may be referred to as a "capture oligonucleotide" that is covalently bound to membranes 14. A plurality of oligonucleotide / antibodies conjugates 18 (a-c) are also provided, each of which comprises an antibody 20 (a-c), which may be a primary or secondary antibody, attached to a cleavable transfer oligonucleotide 22 (a-c) which is complimentary to capture strand oligonucleotides 16. A fluorescent-tag 21 may be attached to each oligonucleotide 22. While only a three layer stack 12 and three oligonucleotide / antibodies conjugates 18 are illustrated in FIG. 1 it should be appreciated that substantially more layers and conjugates can be employed depending on the number of targets sought to be identified. Ten, 20, or even 30 or more layers can be employed, for example.

With reference to frames C-E of FIG. 1 use of a preferred embodiment of the present invention to analyze multiple biomarkers in a tissue section is illustrated. Oligonucleotide-antibody conjugates 18 (a-c) are applied to a tissue section 24 that is mounted to a glass slide 26. In the illustration shown in FIG. 1, tissue section 24 has three distinct targets or biomarkers 28 (a-c) of interest although 10, 20, or even 30 or more biomarkers (e.g. cell

signaling proteins) can be analyzed using the present invention. After conjugates 18 bind to targets 28 unbound antibodies are washed or otherwise removed from tissue section 24 (not shown). Affinity membrane stack 12 is then applied to tissue section 24. Transfer oligonucleotides 22 (a-c) are then cleaved from conjugate antibodies 20 (a-c) and migrate (upward) through stack 12 (a-c) until transfer oligonucleotides hybridize to their complementary capture strands 16 (a-c) on a particular layer 14 (a-c).

Following transfer and binding of transfer oligonucleotides 22 to the appropriate layer in stack 12 and washing of unbound oligonucleotides, the membranes are separated and scanned or imaged using one of several imaging devices discussed in the sections that follow. The fluorescence intensity per unit area and location on the membrane may then be recorded. These images may then be overlaid upon one another and on a corresponding bright field image of the histochemically stained tissue section using image analysis software and analyzed.

In an alternative embodiment affinity membranes stack 12 may be comprised of generally transparent TEMs enabling them to be imaged together as a stack without the need for separation (FIG. 16) as described in the sections that follow.

It should be readily apparent that while FIG. 1 shows use of the present invention for analysis of tissue sections, affinity membrane stack 12 can be employed for a variety of other applications including use in biosensors to detect pathogens or other environmental analytes.

2. Definitions

The following terms shall have the following meanings as used in this Specification:

"Antibodies" means here in general, any of the following types of analyte-specific reagents: a classical antibody produced in an animal host or a fragment thereof such as a Fab fragment; a variety of recombinant antibody proteins either incorporating or consisting of an antibody fragment or a synthetic ligand selected for its target-binding specificity; synthetic nucleic acid aptamers that performs an analogous function; or in general any kinds of synthetic or semi-synthetic reagents that are engineered or selected to provide an appropriate functionality, namely, to enable the delivery of a signaling oligonucleotide to a specific target molecule in a specimen.

"Antigens" means those target molecules that are used to elicit in vivo, or simulate in vitro, a humoral immune response of an animal, with the intention of obtaining specific antibodies therefrom that recognize one or more of the most characteristic epitopes in that target molecule.

"Capture strand" means an oligonucleotide that is covalently bonded to an assay-specific supportive layer. The capture strand comprises a complementary sequence to the specific transfer strand of an assay in accordance with one embodiment herein.

"Epitopes" means those characteristic portions of an antigenic target molecule to which target-specific antibodies make physical contact in the act of specific binding to that molecule.

"Multiplex" means a capability to perform more than one assay on the same specimen at the same time and, in particular, the use of a set of chemically distinctive physical layers as a substrate upon a set of assays is organized and performed in parallel.

"Oligonucleotides" or "ODNs" means a linear sequence of nucleotides joined by phosphodiester bonds. DNA polymers containing up to 50 nucleotides (or base pairs if double stranded) are generally termed oligonucleotides, and longer polymers are called polynucleotides. "Oligonucleotides" is used synonymously with " polynucleotides" for the present purposes. The oligonucleotides of the invention can range up to a few hundred nucleotides but are generally of a minimum of around 18-25 nucleotides in length if only naturally occurring chemical types are used. The nucleotides can be as short or as long as desired, and they may include protein binding segments such as aptamers, as long as self-hybridization and extrinsic molecular binding activities do not impair the reagent's functionality The oligonucleotides may also comprise peptide nucleic acids (PNAs), locked nucleic acids (LNAs), or other types of chemically modified nucleic acids including a cleavable disulfide bond. The oligonucleotide may be single stranded, or it may incorporate an internal duplex segment that is formed by the hybridization of its own self-complementary sequences. The oligonucleotide also may incorporate fluorescent tags and optional fluorescence quencher units.

"Transfer strand" means an oligonucleotide that is released from an analyte- or target-specific reagent (such as an antibody) that carries information about the position on the specimen surface of the target-specific reagent which is also the position of the target. The transfer strand comprises a complementary sequence to the specific capture strand of an assay.

"Sense" and "antisense" are terms which may be used here solely to distinguish two complementary oligonucleotide sequence tags, without meaning any biological connotations.

"Track-etched membranes" means membranes formed by a process that creates well-defined pores by exposing a dense film to ionizing radiation forming damage tracks. This is followed by etching of the damaged tracks into pores by a strong alkaline solution.

3. *Methods of Preparing and Coating the Membranes*

a. *Membranes*

The membranes employed as substrates are preferably "track-etched' membranes (TEMs). TEMs were invented and patented by General Electric (GE) in the 1960s (see U.S. Patent No. 3,303,085). Methods of making and using TEMs and are described by Hanot et al. in "Industrial applications of ion track technology," *Nucl. Instrum. Methods Phys. Res.* Sect. B, 267: 1019-1022 (2009) and "Expanding the use of track-etched membranes" in *IVD Technology* Nov./Dec/ 2002 as well as on the Internet websites of GE's Water and Healthcare business units. Examples of membranes that may be employed for use with the present invention include the Isopore™ (polycarbonate film membrane available from Millipore (Bedford, Mass), the Poretics® polycarbonate or polyester membranes available from Osmonics (Minnetonka, Minn.) or the Cyclopore™ Polycarbonate or Polyester membranes available from Whatman (Clifton, N.J.). Importantly, the etching process results in pores with carboxylic acid residues or other groups that can be covalently bonded to the oligonucleotides as modified herein.

The pore density of the TEM may be between about 10^7 and 10^9 but is preferably about 10^8. The pore size of the TEM may be between about 0.1 μm and 3 μm but is preferably 0.2 or 0.4 μm. The membrane thickness may be between about 5 μm and 20 μm but is preferably about 10 μm. The total surface area including the interiors of the pores may be between about 10 and 50 cm^2/cm^2 of membrane surface but is preferably about 15 cm^2/cm^2.

In an alternative embodiment transparent TEMs may be employed in lieu of conventional opaque membranes. A representative transparent membrane that may be used is Cyclopore Polycarbonate Thin Clear Membranes 1.0 μm Pore Size (cat. no. 7091-4710) available from GE Healthcare Whatman (www.whatman.com). As illustrated in FIG. 15 transparent membranes 38 (which could also be hydrogel layers) may be used in conjunction with an imaging device (e.g. a confocal microscope 32) that can optically penetrate a stack of membranes 38 and ascertain the location (3 dimensions) of signals. In FIG. 15 (A) confocal microscope 32 is used to perform an x-y scan of a first layer of the layered sample set 38a, which provides a single two-dimensional image 34a of all transfer molecules captured in that layer. Subsequent scans at advancing perpendicular depths (B & C) provide additional second 34b and third 34c image layers which are aligned with respect to the common x-y plane. Digital image stacks of the data set are subsequently analyzed to quantitate the local signal intensity of each assay target, also serving to allow for orientation of all assays with respect

to an image of the target cells (such as tumor cells) present in the same tissue section 36 as obtained after subsequent conventional staining (i.e. with hematoxylin and eosin).

Inert layers may be used to separate some or all assay layers (to facilitate an optical analysis without disassembling the layers), and one or more layers may be used to release reagents that cleave the oligonucleotides from antibodies, or that alter the effective stringency of the oligonucleotide hybridization conditions. The layers may be supported on a firm substrate, which may be both transparent and thin enough so that it does not interfere with a microscopic examination (i.e. less than 120 microns); also the layers may be joined at one or more edges (e.g. by a process of local heating, use of an adhesive or ultrasound, etc.) to facilitate further treatments subsequent to transfer while preserving their alignment.

In another embodiment of the disclosure, the material of the layers can be composed of a hydrogel type of substance such as polyacrylamide or agarose layers. These layers may be transparent, and thus the entire stack of layers can be examined by the use of a confocal microscopy without separation of the layers. The microscopy instrument 32 may be used with transparent hydrogel layers.

b. Oligonucleotides

Capture oligonucleotides 16 are preferably linked to one (or more) primary amino groups through a carbon spacer positioned at the 3' and/or 5' terminals (FIG. 5). The oligonucleotides are preferably between 18 and 28 nucleotides in length (if composed of natural nucleotides only) and most preferably between about 20-24 nt long with an amino linkage having the formula:

$H_2N\text{-}(CH_2)_n\text{-}PO_4$ - ODN

with n= 2-7.

The amino linkage is to either the 3´- or 5´- phosphate end of the oligonucleotide. At the opposite end, some oligonucleotides may also contain a linked fluorophore, Cy5 or fluorescein. Fluorophore-labeled oligonucleotides may be purified by HPLC, or by a standard desalting upon their synthesis. These amino linked oligonucleotides can be ordered from one of many companies or they can be prepared by one who is skilled in the art of oligonucleotide synthesis.

c. Attaching the Oligonucleotides to the Membranes

It is a particular feature of the present invention that the aminated oligonucleotide is attached to the TEM in the presence of a carbodiimide (FIG. 5) or an equivalent condensing agent such as triphenylphosphine / dipyridyl disulfide or triphenylphosphine / carbon tetrachloride

To increase the efficiency and quality of the carbodiimide attachment process the TEMs are preferably pre-treated in the following manner.

First, it may be advantageous to de-gas the TEMs. When these membranes are manufactured, about 90 percent of their surface area is comprised of air-filled pores. De-gassing may be accomplished by submerging the membrane in a base buffer, for example, 0.1 M 4-Morpholineethanesulfonic acid (MES) hemisodium salt solution, using a sterile, nuclease free Molecular Biology grade water, and the membrane degassing proceeds under vacuum for 30 min. This step should be repeated, preferably using a fresh salt solution. Alternatively, other solutions may be used for degassing the membrane. These solutions tend to be base solutions that do not adversely affect the membrane. Another buffer that may be used is a 1-methylimidazole buffer.

The buffer solution system is designed to keep the pH constant in the course of a reaction. In the present case, beside the ability to keep pH 5.8-6.2, the buffer must not react with carbodiimide or intermediates it forms with a carboxylic acid. The base pH range may be broader than that presented supra, depending on the conditions and chemicals used. Quoting from "Bioconjugate Techniques," it should be noted that "other (then MES) buffers may be used as long as they don't contain groups that can participate in the carbodiimide reaction. Generally, it helpful to avoid carboxylate- or amine-containing buffers such as citrate, acetate, glycine, or Tris." Alternatively, the reaction can be conducted without any buffer by controlling the pH with pH-meter and adding HCl manually.

Other methods of degassing may be used, as long as the integrity of the track etched membrane is kept intact, and provided that the solution used does not interfere with the condensation reaction.

Following membrane degassing and oligonucleotide preparation the TEMs are saturated with the oligonucleotides and given time to adsorb to the degassed track etched membranes. (See Example 1)

After several hours, the amino groups of the aminated oligonucleotides are coupled to the carboxyl groups of the track etched membranes via condensation, and a condensing agent.

The preferred condensing agent is a carbodiimide. A number of different carbodiimides may be used for the reaction, including, but not limited to, N-(3-Dimethylaminopropyl)-N´-ethylcarbodiimide hydrochloride, (EDC) (Sigma, Cat. No. E1769). EDC belongs to a class of

cold water-soluble carbodiimides. Other cold water carbodiimides that can be used include CMC (1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide and BDC (1-benzyl-3-dimethylaminopropylcarbodiimide). Other non-carbodiimide condensing agents that may be used include oxidants such as triphenylphosphine or a reductant such as dipyridyl disulfide or carbon tetrachloride, preferably used in an organic solvent.

The period of incubation of the track etched membrane, the aminated oligonucleotides, and the condensing agent can range from six to 14 hours. The number of hours for incubation may vary, as determined by the thickness of the membrane, the type of track etched membrane used, and other variables.

While the condensation reaction is quite thorough, there is the possibility that unreacted aminated oligonucleotides may remain adsorbed to the membrane. This is problematic because when the transfer oligonucleotides 22 pass through the stack they could bind nonspecifically to the wrong layer if their complementary strands are not rigidly bound to their assigned layer. Thus, it is advantageous to remove the adsorbed oligonucleotides by washing them until all or virtually all of the adsorbed oligonucleotides are removed from the track etched membranes, thereby leaving behind only those oligonucleotides covalently bonded to the track etched membrane. To accomplish, this, a rigorous washing is necessary that does not chemically adversely affect the covalently bound oligonucleotides nor the track etched membranes to which the oligonucleotides are bound. Preferably a solvent, such as a polar aprotic solvent, is used. In one embodiment, acetonitrile is used. The acetonitrile can reside in water, or in a phosphate solution. Similarly, the track etched membranes, can be washed in a phosphate solution before and/or after the washing in the acetonitrile solution.

It is also advantageous to saturate the membranes with the aminated oligonucleotide prior to exposure and treatment with the condensing agent. This is accomplished by adding the oligonucleotide solution to the buffer solution in which the membranes were degassed, after which the solution containing the condensing agent is added.

4. Antibody / Oligonucleotide Conjugates

With reference to FIGs. 1 and 2, antibody / oligonucleotide ("Ab/oligo") conjugates 18, 38 or 40 are designed to work with the affinity TEMs 14 by detecting the targets 28 in the tissue sample 24 and releasing a fluorescently tagged transfer oligonucleotide 23, 32 or 36 to bind to a corresponding layer 14 in stack 12 containing capture oligonucleotide 16. A preferred assay specific ligand is an oligonucleotide transfer strand, such that, as in the figure, a specific oligonucleotide transfer strand is attached to a specific antibody, which is in turn is uniquely attached to a specific antigen when applied to tissue 24.

In a preferred embodiment, the oligonucleotide is linked to an antibody domain that is not directly involved in binding of the antibody to its target antigen. Secondary antibody conjugates may be used to detect unconjugated primary antibodies, provided that for the use of multiple secondary antibodies in multiplex mode, an immunological difference between the specificity of the different secondary antibodies would be required (e.g. anti-rabbit vs. anti-mouse). In another embodiment, a secondary antibody can be complexed with a primary antibody before the primary is applied to the tissue section.

FIG. 2A illustrates a method in which the transfer strand 22 is linked to the antibody 20 through a direct covalent bond or a series of such bonds; FIG. 2B exemplifies a method in which there is a noncovalent linkage of the transfer strand 32 to the antibody 20, as exemplified by a biotin-streptavidin noncovalent bond in conjugates 38 and 40.

The link between antibody and oligonucleotide may be through an intermediate molecule such as streptavidin. In one embodiment the antibody 20 is linked to streptavidin 34 to form intermediate conjugate 37 by performing a reaction using a covalent bifunctional crosslinker. Where the bifunctional crosslinker may contain a labile bond that as previously described, which bond may be cleaved by a reducing agent or other means recognized by a person of ordinary skill in the art. Biotinylated oligonucleotides 32 are reacted with this to form complex 38. In a related embodiment a biotinylated primary antibody 30 is used. Biotinylated oligonucleotides 32 are bound to the tetravalent molecule streptavidin to form conjugate 36. The biotinylated oligonucleotides 32 may contain a labile bond, which may be cleaved by a reducing agent or other means recognized by a person having ordinary skill in the art. Conjugate 36 is then reacted with biotinylated antibody 30 to form a complex 40.

The link between the antibody and the transfer oligonucleotide preferably may be provided with a labile bond, which may be either covalent or noncovalent. This bond allows the transfer oligonucleotide to be decoupled from the antibody (or intermediary antibody-binding ligand) so as to solubilize the oligonucleotide separately from the tissue for transfer to the capture strand in a layer. Various chemicals and processes for breaking the antibody-oligonucleotide bond can be used; for example, beta-mercaptoethanol can be used to break a disulfide bond; or one could dissociate a duplex consisting of a linker strand that is covalently linked to the antibody and a complementary transfer strand by increasing the temperature to exceeded the melting temperature of that duplex by at about 10-20° C. Interactions of oligonucleotide-conjugated primary antibodies with tissue and of released transfer oligonucleotides with capture strand-coated membranes is illustrated in FIG. 1.

The migration of the transfer oligonucleotide may be facilitated by cleavage of the bond that attaches the transfer oligonucleotide to the antibody. There are a number of ways known in the art to break the bond, which are surveyed in Bioconjugate Techniques, Second Edition, by G. T. Hermanson, Elsevier 2008 which is incorporated herein by reference in its entirety).

In one embodiment, the bond may be broken by gentle heating (37-55° C.) after the stack has been applied to the bound sample, i.e. during the transfer.

In another embodiment, an ancillary chemical, which could be a stabilized reducing agent such as beta-mercaptoethanol or TCEP (Thermo Pierce cat. no. PI-77720, the manual for which is herein incorporated by reference) is pre-positioned in the layers at a working concentration therein, and upon contact of the outermost layer with the specimen, reacts with the crosslinker, causing rapid and complete cleavage of the crosslinker structure between the antibody and the tagged oligonucleotide within 1-60 minutes.

Alternatively, other linkage structures could be used to enable solubilization of the oligonucleotide from the antibody. Such linkages include but are not limited to a photosensitive linkage, or an oligonucleotide subsequence that could bind a chemical such as an enzyme that would cause localized cleavage within the oligonucleotide. There are numerous chemical structures that would allow for an easily breakable bond between the oligonucleotide and the antibody. Displacement by competitive binding of another oligonucleotide could be used to release the assay-specific ligand.

In the process, one type of specific capture molecule (the capture oligonucleotide strand) is attached to each of said vertically ordered, bioaffinity layers. The assay-specific ligands (transfer oligonucleotides) that are attached to the primary or secondary antibodies (which here exemplify the class of analyte-specific ligands) are released from the tissue sample (by application of a chemical or by mild heating to 37-55°C), and they necessarily diffuse away from the specimen, along a direction of movement which is determined by the vector of their chemical concentration gradient. Each of the assay specific ligands (e.g. a transfer oligonucleotide) passes through the stack until it meets a bioaffinity ligand (the capture oligonucleotide which is to it the antisense strand) to which it specifically binds according to the pertinent type of bioaffinity chemistry (e.g. formation of duplex DNA). It should be noted that in some cases the antibody, still attached to the transfer strand, migrates with the transfer strand, with no adverse effects to the reading of the results.

If a secondary antibody rather than a primary antibody is conjugated with an oligonucleotide, normal care must be taken to avoid cross reactions with different primary antibodies. Either those primary antibodies must be distinguishable by their respective secondary antibodies, or

a complex of a primary antibody and monovalent secondary antibody fragment could be formed prior to application of the mixed antibody complexes to the tissue specimen (not illustrated).

A fluorescent tag 21 may be attached to transfer oligonucleotide 22 for subsequent detection by an imaging instrument such as fluorescent scanner the Typhoon® scanner available from GE Life Sciences. Examples of fluorescent tags that may be used include dyes such Cy5, Cy3, fluorescein, etc. Fluorescent tag 21 is preferably added to the oligonucleotide during stepwise synthesis in a manner commonly known in the art. Whereas in the examples provided herein a single dye was employed it should be appreciated that different colors can be employed which may be advantageous when using the transparent membrane approach (FIG. 15). In lieu of a fluorescent tag various methods may be employed to detect duplexes of the oligonucleotides. A double-strand specific DNA binding dye may be used analogously to a real time PCR detection method, such as SYBR Green I available from Invitrogen, Inc. (Carlsbad, CA).

5. Uses and Applications of Coated Membranes and Conjugates

Oligonucleotide coated affinity membranes 12 and Ab/oligo conjugates 18 may be used in a variety of configurations for a variety testing purposes. One important application, illustrated in FIG.1 and described in Examples 8-10, involves the profiling of multiple biomarkers in a tumor section. For this application a tumor is sectioned and prepared according to standard clinical pathology processes for routine immunohistochemistry (IHC) analysis ('special staining') except that Ab/oligo. conjugates 18 are used in lieu of standard IHC antibodies. (It should be appreciated that virtually any IHC antibody can be conjugated with an oligonucleotide according to the processes described herein.)

Membranes 12 or 14 are then dipped into an oligonucleotide binding buffer containing a reducing / releasing agent, such as one that can cleave the disulfide bonds in the Antibody/oligonucleotide conjugates 18 such as TCEP or beta-mercaptoethanol. The membranes are assembled into a stack 12 in a chosen order as to the identity of the oligonucleotides coated on them, which ordering serves to spatially organize the assays which are to be performed.

Following application and binding of conjugates 18 to tissue section 24 the tissue sample is rinsed with binding buffer without a reducing / releasing agent, and membrane stack 12 is placed directly on top of the tissue section. The excess buffer volume is expressed from the stack by enclosing it between protective layers and applying gentle pressure. In a preferred embodiment, the stack is first assembled and then applied to a specimen surface where a

remnant of wash buffer persists from the last specimen wash step. The excess wash buffer is then expressed from the specimen surface by applying a weight or mechanical pressure to the entire assembly. Both the tissue section and membrane stack are preferably enveloped in a fluid impermeable enclosure (e.g. plastic bag) and kept moist throughout the transfer process.

When membrane stack 12 is applied to the tissue specimen a weight or similar pressuring device such as spring metal clips (not shown) may be applied to express an excess buffer volume and assure uniform contact with the tissue specimen throughout the transfer process. The enclosed stack and slide mounted tissue section are then incubated for an optimal time (between about 10 minutes and 4 hours) and a temperature range of between room temperature and 70° C depending on the requirements of the kinetics of the chemical reactions which are being performed; such requirements could include the concentration of a reducing agent, the diffusion rate of a transfer strand through the type of substance used to form the layered supports, etc.

In an embodiment in which a transfer strand is linked to an antibody through a disulfide bond, the reducing agent functions to cleave the transfer oligonucleotide strands 22 from antibody 20, or transfer strands 32 from complex 38 or 40, or complex 36 from antibody 30. The transfer strands then diffuse upward through the porous membrane stack and bind to the membrane layer 14 supporting the complimentary oligonucleotide strand as shown in FIG. 1 E. Examples 5, 8, 9 and 10 exemplify the use of this method.

Membrane stack 12 is then removed from the enclosure and washed in a buffer. In one embodiment (opaque or translucent membranes) the membranes layers are separated from one another, dried, and arrayed on a flat bed fluorescence scanner (not shown). The scanner then records the location and intensity of fluorescent tags 21 on each membrane. This data can be analyzed using a variety of software programs such as ImageQuant (Molecular Dynamics) to permit the user to quantitate and correlate the assayed multiple biomarker expression levels at given locations within the tissue section.

A fluorescence quencher can be used in the transfer strand to suppress light emission of the fluorescent tag until hybridization to the capture strand occurs on the layer (e.g. FRET); useful fluorescence tags and fluorophore-quencher pairs can be used that are well known to one skilled in the art of fluorescence detection methods (these principles are outlined in the short report, "Fluorescence and Fluorescence Applications", 2005, available online from Integrated DNA Technologies, incorporated herein by reference in its entirety). The membrane stack may be dismantled such that each membrane is read separately, or the entire

stack may be read by use of a confocal microscope if the individual membranes are transparent or translucent.

In some embodiments, an antibody may be provided with a covalently conjugated linker strand that is truncated and forms a low-affinity duplex with a transfer strand; the transfer strand may be subjected to thermal dissociation from the antibody during the transfer step by mild heating (e.g. by 15-30° C). In general, persons who are proficient in the art would recognize many ways to cause dissociation of various types of transfer complexes on the specimen, while providing for the subsequent binding of the transfer molecule to the capture layer.

In some embodiments, an arbitrary number of oligonucleotide sequence tags (for example, United States Patent 5,635,400), sometimes referred to as bio-barcodes, are identified that do not cross-hybridize when mixed in solution. In particular, the transfer oligonucleotides do not cross-hybridize to one another, and they do not cross-hybridize to the capture oligonucleotides of the other transfer oligonucleotides that are attached to the capture layers. These embodiments enable modular expansion of the number of assays per run and a replicable assay development procedure all based on the universality of the oligonucleotide transfer / capture system as here employed. Similar universal multiplex assays have been developed as biochemical techniques, for example these have been used in microarray analysis of PCR products (Favis, R. et al (2000) Universal DNA array detection of small insertions and deletions in BRCA1 and. BRCA2. Nat. Biotechnol. 18, 561–564).

In addition to use with tissue sections, oligonucleotide coated affinity membranes 12 and Ab/oligo conjugates 18 may be used in other testing formats, for example, assays of blood or other bodily fluids in a multi-well plate format in the same manner as that described for the Layered Peptide Arrays as described by Emmert-Buck, et al. in U.S. Patent Application Pub. No. 2009/0215073 A1 (Appl. No. 12/289,736) and in *Clinica Chimica Acta* 376 (2007) 9-16 and the *Journal of Molecular Diagnostics*, Vol. 9, No. 3, July 2007 pgs. 297-304.

Affinity membranes 12 may also be used without conjugates 18, for example to detect DNA or RNA targets in blood or an environmental sample (soil, air, or water) as a component of biosensors. Examples of use of TEMs for various biosensors and diagnostics for which the present invention may be employed are described by Hanot et al. in "Industrial applications of ion track technology," *Nucl. Instrum. Methods Phys. Res.* Sect. B, 267: 1019-1022 (2009) and by Jones et al. in "Expanding the use of track-etched membranes" in *IVD Technology* Nov./Dec/ 2002.

EXAMPLES

Introduction to Examples. The covalent immobilization of single-stranded DNA was demonstrated on commercially available track-etched membranes using oligodeoxynucleotides 20-24 nt long. In the Examples below the membranes were mainly circular discs with a diameter of 6.5 mm that fits the wells of a standard 96 well plate. Before the use, the membranes were submerged into a conjugation buffer and degassed under vacuum with two changes of the buffer. Quantitative and semi-quantitative description of the immobilization is based on the use of fluorescently labeled oligonucleotides and measurement of the fluorescence intensity in reaction mixtures, washings, and on membranes. The measurements and image acquisitions were done on 1420 Multilabel Counter Victor^2 (Wallac, Turku, Finland) and Storm 860 Gel and Blot Imaging System or Typhoon Trio+ Variable Mode Imager (Amersham Biosciences, Sunnyvale, CA). The images were treated with the Molecular Dynamics software ImageQuant, version 5.2.

Membranes, DNA, and Reagents. The track-etched membranes in a sheet format or circular were from two sources: GE Water & Process Technologies (Trevose, PA) and the Belgian company it4ip (Seneffe, Belgium). Specifications of the membranes are presented in TABLE 1. The majority of oligodeoxynucleotides was synthesized by Eurofin MWG Operon (Huntsville, AL) and some were obtained from Integrated DNA Technologies (Coralville, IA). Structures of the oligonucleotides and ways of their purification are shown in TABLE 2. Other principal reagents were N-(3-dimethylaminopropyl)-N´-ethylcarbodiimide (EDC) hydrochloride, 4-morpholineethanesulfonic acid (MES) hemisodium salt, acetonitrile, HPLC grade, and 0.2 M phosphate buffer, pH ~7.4, containing 2.98 M NaCl and 0.02 M EDTA (SSPE buffer 20× concentrate) (Sigma-Aldrich Corp., St. Louis, MO). The oligonucleotide solutions were prepared using DEPC-treated, sterile nuclease-free water (EMD Chemicals, Gibbstown, NJ). Milli-Q water passed through the BioPak ultrafilter (Millipore, Billerica, MA) was used to prepare other solutions and buffers.

TABLE 1

Membrane & Cat. No.	Pore size, μm	Pore Density, cm^{-2}	Thickness, μm	Vendor
Polycarbonate K04SH02500 K04SH81030 K04CP02500*	0.4	1×10^8	10	GE
Polyester T04CP02500	0.4	1×10^8	10	GE
Polycarbonate	0.4	1.6×10^8	10	it4ip

1000M10/811N400/A4				
Poly(ethylene terephthalate) 2000M12/811N400/A4	0.4	1.0×10^8	12	it4ip
Polyimide 3000M12/821N400/A4	0.4	2.0×10^8	12	it4ip

*The membranes are coated with a wetting agent, poly(vinylpyrrolidone).

TABLE 2

Name	**Structure**	hybridizes	Purity	Vendor
ATP17	[Cy5]TGAT$_2$GTAGTATGTAT$_2$GATA$_3$G[AmC7]	ATP95	HPLC	Operon
ATP95	[Fl]CT$_3$ATCA$_2$TACATACTACA$_2$TCA	ATP17	HPLC	Operon
ATP60	[AmC6]TCAT$_3$AC$_2$A$_2$T$_3$AC$_2$A$_2$T	ATP25	desalting	IDT
ATP25	[Cy5-Sp9]AT$_2$G$_2$TA$_3$T$_2$G$_2$TA$_3$TGA	ATP60	PAGE	Operon
ATP73	[Cy5]AT$_2$G$_2$TA$_3$T$_2$G$_2$TA$_3$TGA[ThiSS][BioTEG]	ATP60	HPLC	Operon
ATP61	[AmC6]CT$_3$CT$_3$CT$_3$CT$_3$CT$_3$	ATP72	desalting	Operon
ATP72	[Cy3]AAAGAAAGAAAGAAAGAAAG[AmC7-Q]	ATP61	HPLC	Operon
ATP62	[AmC6]CTACTATACATCT$_2$ACTATACT	ATP74	desalting	Operon
ATP74	[Cy5]AGTATAGTAAGATGTATAGTAG[ThiSS][BioTEG-Q]	ATP62	HPLC	Operon
ATP80	[AminoC6]AAATCATCAATCACTTTAAT	ATP76	desalting	Operon
ATP76	[Cy5]ATTAAAGTGATTGATGATTT[ThiSS][BioTEG-Q]	ATP80	HPLC	Operon
ATP82	[AminoC6]CTACAAACAAACAAACATTAT	ATP77	desalting	Operon
ATP77	[Cy5]ATAATGTTTGTTTGTTTGTAG[ThiSS][BioTEG-Q]	ATP82	HPLC	Operon

Modification abbreviation	**chemical**
[Cy5]	5' CY 5 fluorescent dye
[Fl]	5' Fluorescein fluorescent dye
[AmC6]	5' primary amine with 6 atom spacer
[AmC7]	3' primary amine with 7 atom spacer
[Cy5-Sp9]	5' CY5 with 9 atom spacer
[ThiSS]	Dithiol (disulfide), internal
[BioTEG]	Biotin with 22 atom spacer

Structures of the modifications (Cy5, AmC6, etc.) are shown in FIG. 6.

EXAMPLE 1. Covalent coating of TEMs via their joint incubation with aminated DNA and carbodiimide. Pairs of the membrane discs degassed in 0.1 M MES buffer, pH 6.1 were submerged without overlapping in 500-µl vials in 80 µl of 62.5 µM aminated ODN in 0.125 M MES buffer and kept there at a room temperature light-protected if the ODN had also an attached fluorophore. Upon 2 hour incubation vials received 20 µl of the freshly prepared 0.5 M EDC in water while vials with control membranes received just water. After the overnight reaction the incubation mixtures were withdrawn and the membranes were washed with 100 µl of 6× SSPE buffer (thrice, each time for 30 min) and then with 100 µl of 10% acetonitrile (thrice, each time for 30 min). The coated membranes were stored in 6× SSPE buffer at 4-6 °C. For analytical purposes, the incubation mixtures, washings as well as the resulting membranes (covered with 100 µl of 6× SSPE buffer) were placed in designated wells of 96 well plates and fluorescence intensity was recorded usually at 0.1 s exposures. Such acquired data are presented in FIGS. 7-9.

FIG. 7 describes results of the ATP17 covalent binding to a track-etched polyester membrane. In the procedure used, the membranes were first saturated with the ODN in a prolonged incubation and only then the immobilization was initiated by the carbodiimide addition. By comparing bars 1 and 10 in FIG. 7A, one might conclude that about 10% of the ODN appears covalently bound to the membrane. When passed through the same treatment but without carbodiimide, the membrane retains less than 0.3% of the ODN applied (compare bars 1 and 10 in FIG. 7B). Although polyester (poly(ethylene terephthalate) or PET) is considered as a naturally hydrophilic polymer, data in FIG. 7 show that three consecutive washings with a phosphate buffer (bars 2-4) are less effective in deleting the non-covalently bound ODN from the polyester membrane than single washing with 10% acetonitrile (bars 4 in FIG. 7A&B).

The ATP17 covalent binding to track-etched polycarbonate membranes is characterized in FIG. 8. In contradistinction to polyester, polycarbonate is a highly hydrophobic polymer and its film etching should presumably result not in carboxylic acid end groups but carbonic acid end groups. Still, the procedure provides of the polycarbonate membrane coating (see bar 7 in FIG. 8A) comparable with that achieved with polyester membranes (bar 10 in FIG. 7A). Bars 1 in FIG. 8A&B also show that upon completion of the treatment with or without the use of carbodiimide and regular washings the membranes retain almost the same amount of the ODN. Yet, in the subsequent washings with 10% (bars 2 and 3) and 30% (bars 4 and 5) acetonitrile, the membranes treated without carbodiimide lose nearly 83% of the ODN while the counterparts still retain the most of it (compare bars 7 in FIG. 8A&B).

Data in FIG. 9 are to confirm that the procedure of ODN covalent binding to commercially available TEMs works well irrespectively who manufactured them. As seen in particular, the use of polyester membranes obtained from Belgium company it4ip (bar 2) and GE Water & Process Technologies (bar 4) resulted in their nearly identical coating with ATP17. Considering the intensity of the signals from the bound ODN (gray bars in FIG. 9), it can be concluded that on average the polyester, polycarbonate, and polyimide membranes provide correspondingly high, medium, and small binding capacity. Importantly, that the background signal caused by the ODN binding independent on carbodiimide decreases in the inverse relation; this follows from the comparison of the gray and white bars in FIG. 9.

EXAMPLE 2. Covalent coating of TEMs via their sequential incubation with aminated DNA and carbodiimide. Pairs of the membrane discs degassed in 0.1 M MES buffer, pH 6.1 were submerged without overlapping in 500-µl vials in 100 µl of 50 µM aminated ODN(ATP17) in 0.1 M MES buffer and kept there at a room temperature light-protected if the ODN had also an attached fluorophore. Upon 4 hour incubation the membranes were transferred into the vials with 100 µl of the freshly prepared 0.1 M EDC in 0.1 M MES buffer and the incubation continued overnight. Washing of the membranes was done as described in EXAMPLE 1. The comparison of the two ways of the coating is shown in FIG. 10. This EXAMPLE is to show that the covalent coating of TEMs via carbodiimide condensation could be the same if instead of adding carbodiimide in an ODN solution with a membrane this membrane would be taken out of the solution upon its saturation and placed in another, containing only carbodiimide solution. In such a case, the ODN solution would not be contaminated with added carbodiimide and it could be used multiple times. Based on the comparison of bars 9 and 10 (which refer to pairs of membranes treated in parallel) in FIG. 10A&B it follows that both variants of the coating lead to practically identical products.

EXAMPLE 3. Demonstration of the evenness of the covalent coating. The membrane disks coated with the aminated Cy5-labeled oligonucleotide as in EXAMPLES 1 and 2 and then dried between sheets of Whatman filter paper were placed on the Typhoon Trio+ platen, covered with a regular glass slide and their images were acquired at a normal sensitivity. To avoid saturation of the signal, excitation was done with a green laser (532 nm) instead of the matching red laser (633 nm), emission was registered at 670 nm at a pixel size of 50. Thus obtained fluorescent images of the ATP17-coated polycarbonate membranes are shown in FIG. 11A&B.

Distribution of the oligonucleotide along diameters of the coated disks was revealed with the ImageQuant software and shown in FIG 11 C & D. As seen, both variants of the ODN

covalent binding result in a fairly even coating: the joint incubation provided variations within ± 15% (FIG. 11C) and the sequential incubation gave even better result, ± 5% (FIG. 11D).

EXAMPLE 4. Hybridization and thermal dissociation on a track-etched membrane.

Pair of membrane disks coated with ATP60 as in EXAMPLE 1 and washed with 6× SSPE buffer was incubated at a room temperature for 2 hours in 100 µl of 10 µM complementary Cy5-labeled ATP25 in the buffer. Another pair of plain membrane disks was kept in the identical solution to serve as a control. Upon four 30 min-washings with the buffer the disks were transferred into designated wells of a 96 well plate and intensity of their red fluorescence was measured. The disks were further washed with 10% acetonitrile (thrice), 30% acetonitrile (once), and 6× SSPE buffer. The volume of each washing was 100 µl and intensity of the red fluorescence in the washings and resulting membranes is shown in FIG. 12. The data show that an oligonucleotide (ATP60) covalently bound to polycarbonate membrane preserves its ability to hybridize with a complementary oligonucleotide (ATP25). Although ATP25 can interact with the plain membrane non-specifically (FIG. 12A), 93.5% of it can be washed out by the employed washings (compare bars 1 and 7 in FIG. 12A). The same washings leave 83% of ATP25 on the membrane coated with ATP60 (FIG. 12B, compare bars 1 and 7).

Another pair of membrane disks coated with ATP17 and washed as in EXAMPLE 1 was incubated at 50 °C for 10 min in 100 µl of 25 µM complementary fluorescein-labeled ATP95 in 6× SSPE buffer. This was followed by the 3 hr annealing to a room temperature and washings with the buffer. Upon recording the green fluorescence intensity retained by the disks they were placed into a vial with 100 µl of 25 mM phosphate buffer, pH 7 and heated up to 60 °C. In 15 min the solution from the vial was withdrawn and the heat extraction was performed two times more. Intensity of the green fluorescence in the discs before and after thermal dissociation as well as in the hot extracts is shown in FIG. 13.

It is supposed that a duplex formed in solution by two complementary oligonucleotides can be destroyed (melted) at an elevated temperature. Data in FIG. 13 demonstrate that this is also happening when one of the two is immobilized on a track-etched polycarbonate membrane. The pair of ATP17 and ATP95 has an estimated melting temperature of 54 °C. Bars 2-4 in FIG. 13 relate to amount of ATP95 released from the immobilized duplex into solution at 60 °C while bar 10 shows what remains (less than 7%) on the membrane after the thermal dissociation.

EXAMPLE 5. Hybridization in a stack of track-etched membranes of an oligonucleotide released from an antibody conjugate bound to an antigen. The complex of an antigen with an antibody-streptavidin conjugate and bound biotin-modified ATP73 was prepared on the irregularly shaped piece of a polycarbonate membrane as follows. First, the membrane was coated with rabbit IgG and then blocked with 1% BSA. Second, the membrane was incubated with a conjugate of the goat anti-rabbit antibody and streptavidin that retained ATP73. Upon washing with 4× SSC buffer contained 0.1% SDS the membrane was placed on a supporting polycarbonate PVP-coated disc and covered by a stack in that PVP-coated discs separated the discs with ATP62, ATP61, and ATP60 immobilized as in EXAMPLE 2. The concerted release of the ATP73 fragment and its hybridization with ATP60 were induced by placing on the stack's top a filter paper disc wetted with 4× SSC buffer contained 0.1% SDS and 5 mM Tris[2-carboxyethyl]phosphine hydrochloride (TCEP). After keeping the stack at 45 °C for one hour it was disassembled and the discs were washed thrice with 1x SSC buffer contained 0.25% SDS at 47 °C. Fluorescent images of the discs were registered using Storm 860 scanner (FIG. 14). Intensity of the red fluorescence associated with each disc in the stack was also quantified using the microplate reader.

EXAMPLE 8 – Identification of Multiple Biomarkers From Breast Tumor Sections and Comparison with Immunofluorescence histochemistry

Breast cancer pathology specimens in paraffin blocks were purchased from a commercial tissue bank (Asterand, Inc. or Bioserve, Inc.), 5 µm sections were cut and prepared for staining with conventional immunohistochemistry sample preparation methods. Membranes were prepared by conjugation with oligonucleotides (ODNs) ATP61, ATP80 and ATP82 by the methods of Example 2, and were composed into four replicate stacks, also including an uncoated negative control membrane. Antibodies to assay targets ErbB2, ER (estrogen receptor) and CK7 (cytokeratins 7) were respectively conjugated with CY5 fluorescent-tagged sense ODN, in the same manner as described above and shown in complex 40 on Figure 2B. These conjugates were used to bind conjugates to tissue sections by following a conventional IHC protocol. Either one or three conjugated Abs were used per slide. After the last washing step, coated membranes were equilibrated in release/hybridization buffer (4x SSC/0.1% SDS, 50 mM beta-mercaptoethanol) and applied on the slide for a 30 min incubation at 47^0C. The stacks were dissociated, membranes washed in 2xSSC/0.1% SDS, rinsed in distilled water and dried before being scanned on bed flat scanner (Typhoon) with the results shown (Fig. 17A). The data provide a comparison of the three conditions: a) The amount of conjugate bound on the slide before transfer of transfer strands (first image

column); b) the amount of transfer strands transferred to capture membrane when one conjugate is applied to the tissue (rows 1-3 in image columns 2-5); and the amounts of transfer strands transferred to tissue when all three conjugates are incubated on the tissue together (row 4 in image columns 2-5). The control membrane without any antisense conjugated oligonucleotide had no signal (image column 2). Selected image areas were quantitatively measured for CY5 fluorescence using ImageQuant software and the data were plotted (Fig. 17B). In this experiment, single-staining data were proportional to multiplex data and were more intense.

Example 9. Further comparison of double staining triple staining

The same materials and methods of Example 8 were used, except that either one, two or three conjugates were applied together to one tissue section. The data (Fig. 18) indicate that the capture efficiency of the released ODN-cy5 to complementary membranes is not much affected by the presence of multiple conjugated antibodies. Quantitative analysis (Fig 18B) indicated that the intensity of triple transfers (row 4) was nearly the same as that of double transfers (rows 2 and 3) or single transfers (row 1).

Example 10. Target localization in tissue using triple multiplex mode

The same materials and methods of Example 8 were used. In FIG. 19A, some of the target proteins seem to have been over-expressed in those manually circumscribed areas are imaged both on the slide before transfer (row 1) and on the membranes after transfer, such as: ErbB2 on membrane 60 (row 2) and CK7 on membrane 62 (row 3). Quantitative analysis in FIG. 19B indicates that overall intensities as measured in uniplex and triplex modes were indeed comparable, although those local intensity variations were less obvious.

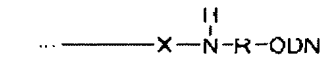
What is claimed is:

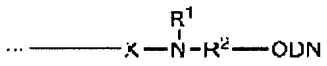
1. A method of analyzing biomolecules in a tissue section comprising the steps of:
providing a stack of membranes, wherein each of said membranes has a different oligonucleotide attached thereto;
providing a set of antibodies that have specific affinity to the biomolecules in the tissue section, wherein each of said antibodies has conjugated thereto a cleavable oligonucleotide that can hybridize to the oligonucleotide attached to said membranes;
applying said antibodies to said tissue section in a manner that permits the antibodies to bind to the biomolecules;
cleaving said oligonucleotides from said antibodies;
transferring said cleaved oligonucleotides from said antibodies to said membrane stack whereupon the transferred oligonucleotide binds to complimentary oligonucleotides attached to the membranes;
detecting the hybridization of the transferred oligonucleotides.

2. An oligonucleotide coated track etched membrane comprising:

a track etched membrane; and
aminated oligonucleotides covalently bound at their amino group to a carboxyl or carbonic group of the track etched membrane.

3. An oligonucleotide coated track etched membrane comprising either of the two structures:





Wherein X is a structural component of the track etched membrane; N is an alkyl amine connecting the oligonucleotide to the structural component of the track etched

membrane; R^1 is a substituent of the alkyl amine; and R^2 is a linker that may be connected to a phosphate group at a 5' or 3' terminal end of the oligonucleotide / oligodeoxynucleotide; and ODN is an oligodeoxynucleotide.

4. A method of capturing oligonucleotides from a sample comprising the steps of:
providing a stack of two or more track-etched membranes, wherein each of said track-etched membrane has a different oligonucleotide attached thereto that can hybridize with an oligonucleotide in the sample;
transferring said sample oligonucleotides to said track-etched membrane stack wherein the sample oligonucleotides hybridize to the oligonucleotides attached to membranes.

5. The method according to claim 4 wherein said oligonucleotide is attached to said membrane by a covalent bond.

6. The method according to claim 4, wherein said track etched membranes are selected from the group consisting of polyimide track etched membranes, polycarbonate track etched membranes, and polyethylene terephthalate track etched membranes.

7. A method of making oligonucleotide coated track etched membranes, said method comprising the steps of:
obtaining track etched membranes;
obtaining aminated oligonucleotides;
obtaining a condensing agent;
reacting said track etched membranes in the presence of said condensing agent, thereby forming said oligonucleotide coated track etched membranes.

8. The method according to claim 7, wherein said condensing agent is a carbodiimide.

9. The method according to claim 8, wherein carbodiimide is selected from the group consisting of Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride, 1-cyclohexyl-3-(2-morpholinoethyl) carbodiimide, and 1-benzyl-3-dimethylaminopropylcarbodiimide.

10. The method according to claim 7, wherein said track etched membranes are selected from the group consisting of polyimide track etched membranes, polycarbonate track etched membranes, and polyethylene terephthalate track etched membranes.

11. The method according to claim 8, further comprising washing said oligonucleotide bound track etched membrane in a solvent having the capability to remove virtually any adsorbed aminated oligonucleotides.

12. The method according to claim 11, wherein said solvent is a polar aprotic solvent.

13. The method according to claim 12, wherein said polar aprotic solvent comprises acetonitrile.

14. The method according claim 7, further comprising the step of degassing said track etched membranes prior to exposing them to said aminated oligonucleotides.

15. The method according to claim 14, wherein said degassing occurs under vacuum pressure.

Figure 1 A

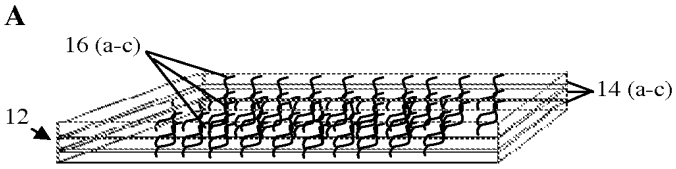


Figure 1 B

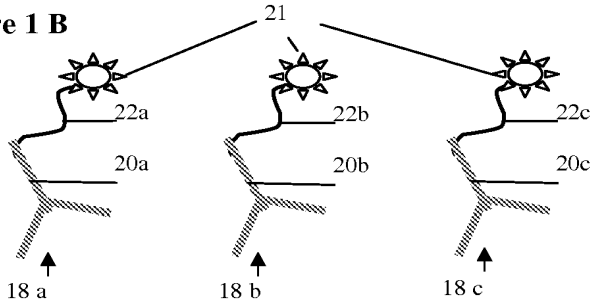


Figure 1 C

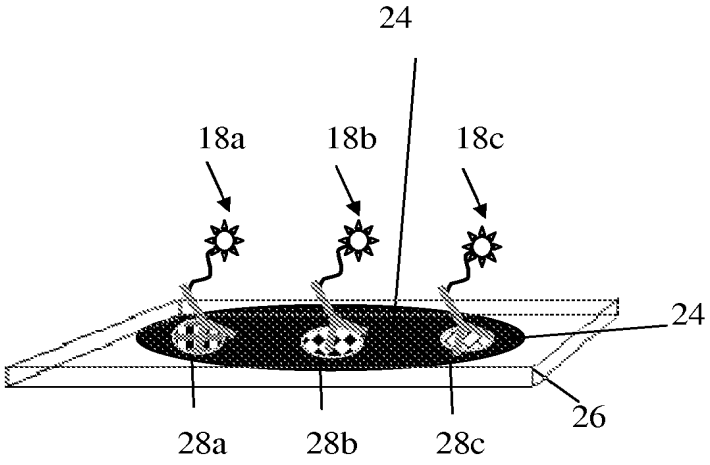


Figure 1 D

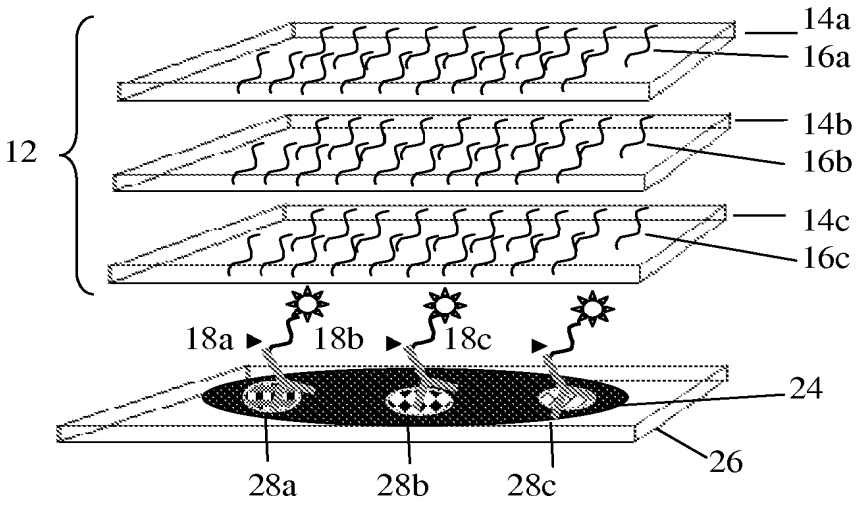


Figure 1 E

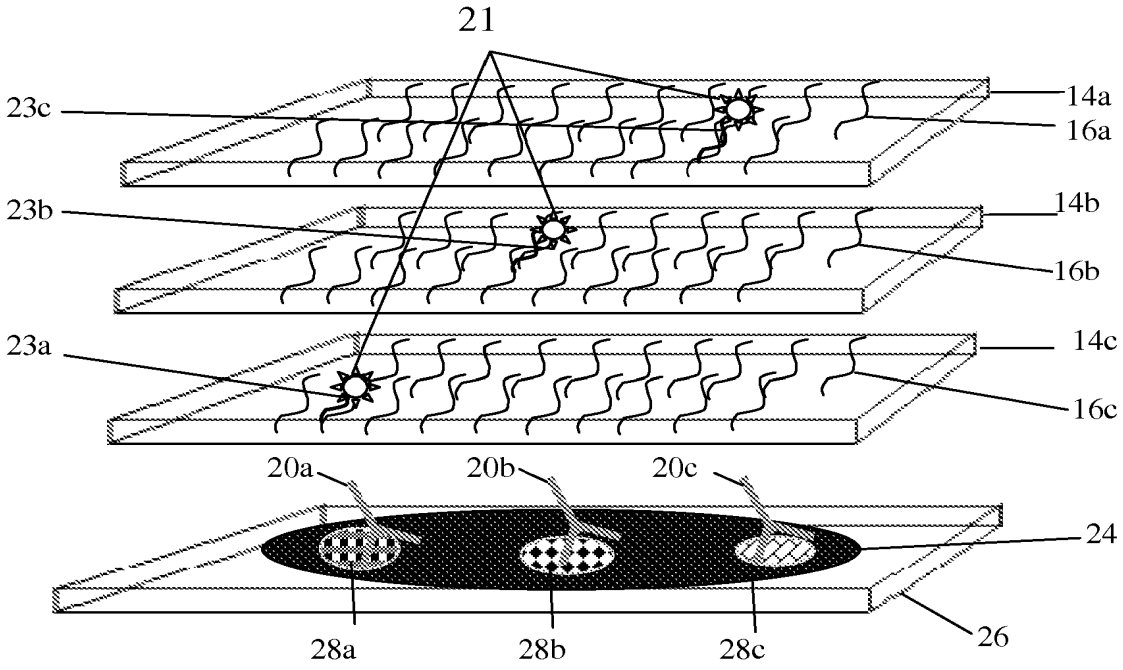


Figure 2 A [types of antibody-oligonucleotide conjugates]

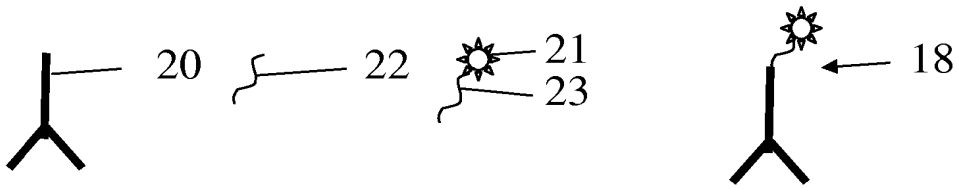


Figure 2 B

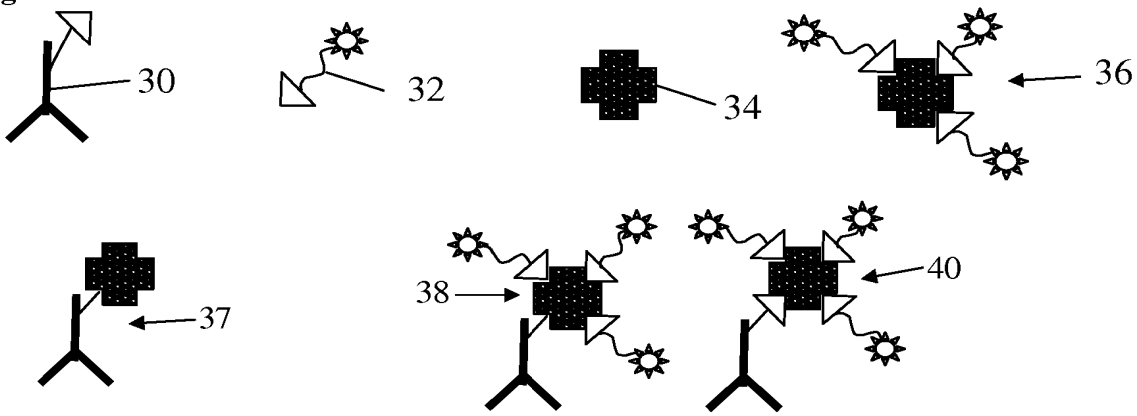
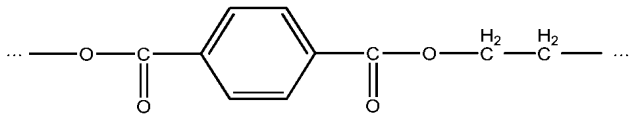
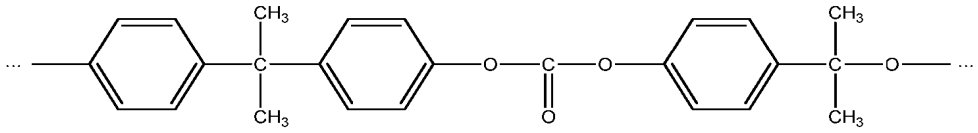


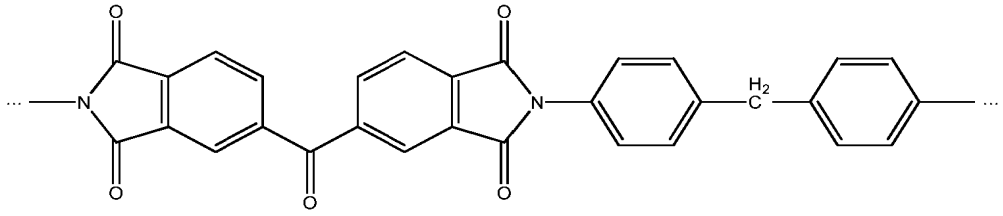
Figure 3



polyester (poly(ethylene terephthalate), PET)



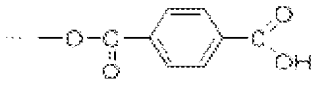
polycarbonate



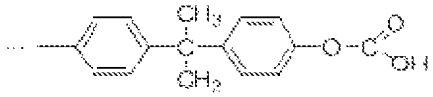
polyimide

Figure 4

PET etch product



Polycarbonate etch product



Polyimide etch product

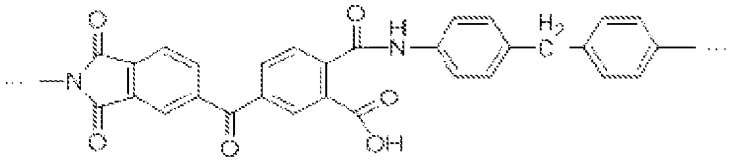
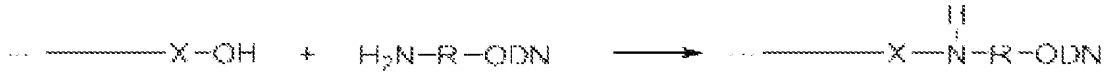


Figure 5 A

1° amine



2° amine



Figure 5 B

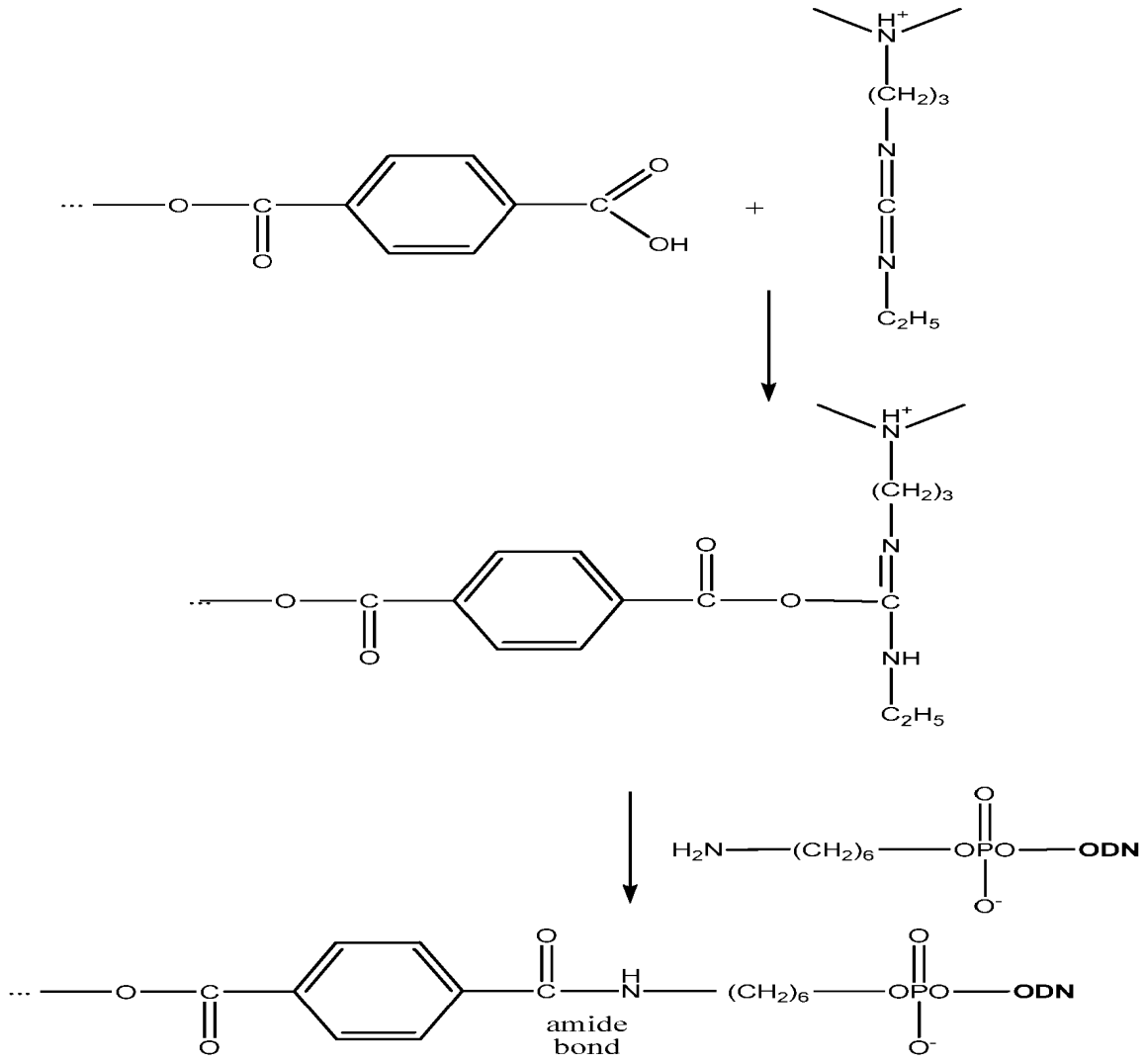
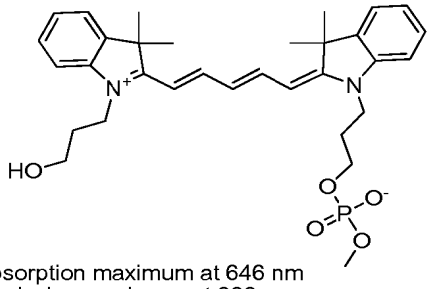


Figure 6

ATP17:

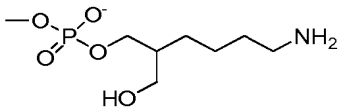
[Cy5] TGA TTG TAG TAT GTA TTG ATA AAG [AmC7-Q]

[Cy5]:

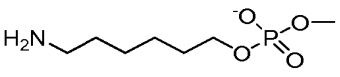


Absorption maximum at 646 nm
Emission maximum at 662 nm

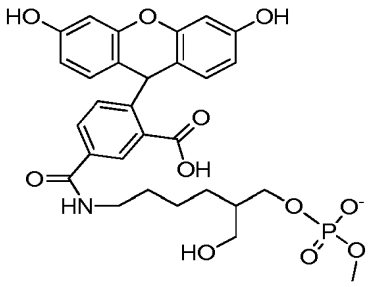
[AmC7]:



[AmC6]:



[Fl]:



Absorption maximum at 495 nm
Emission maximum at 520 nm

[ThiSS][BioTEG]:

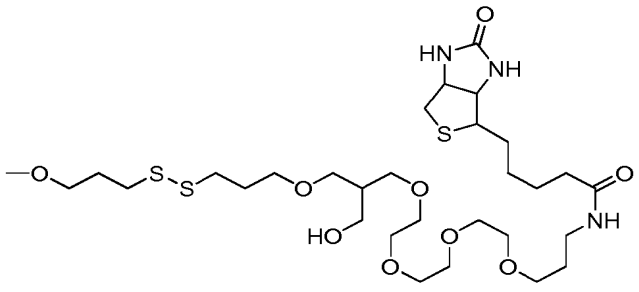


Figure 7A

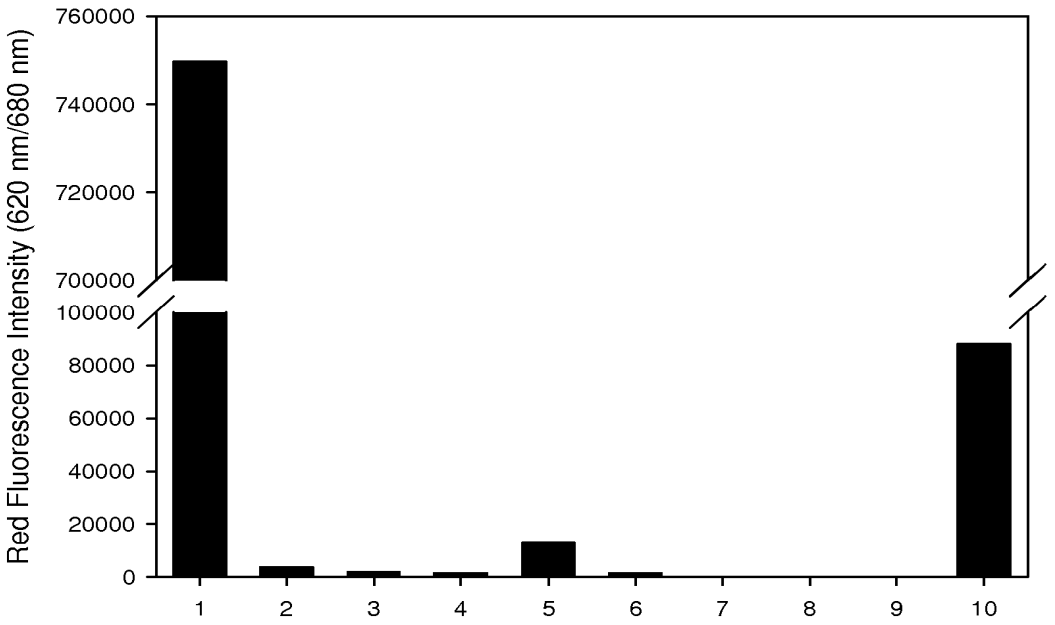


Figure 7B

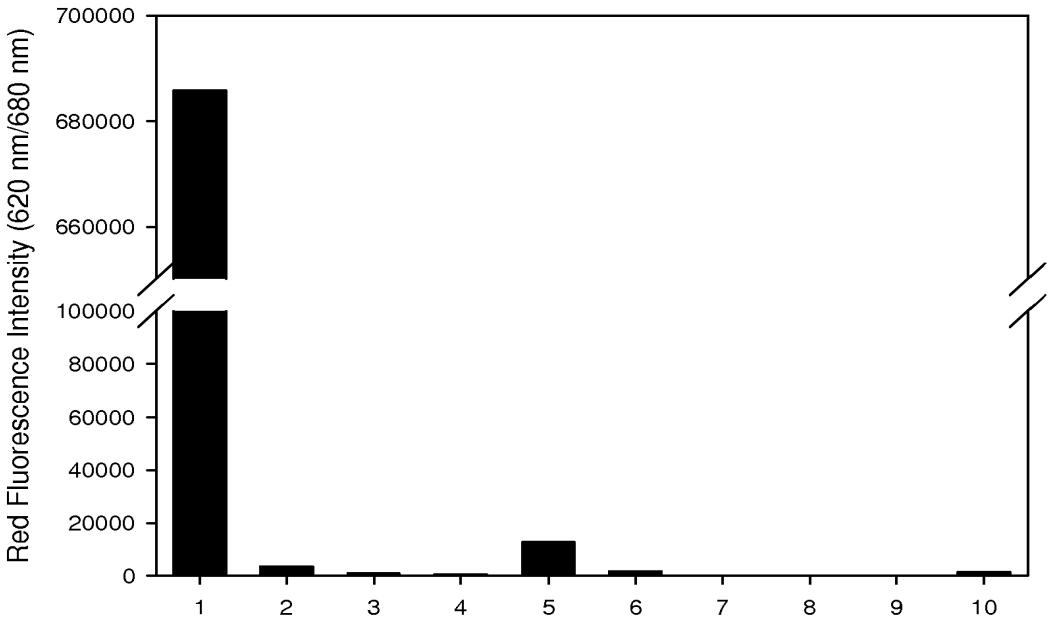


Figure 8A

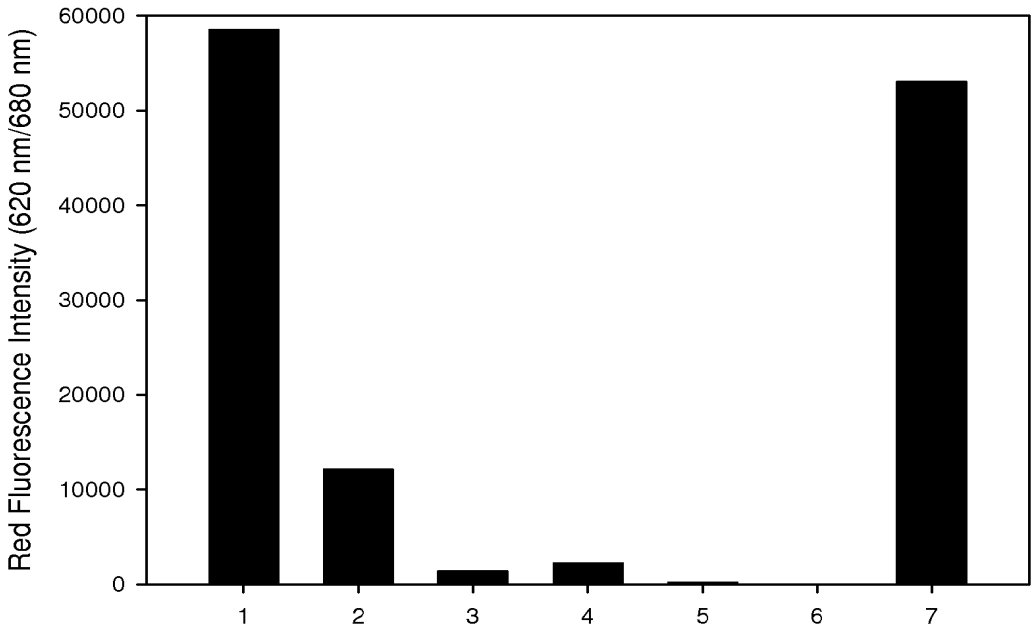


Figure 8B

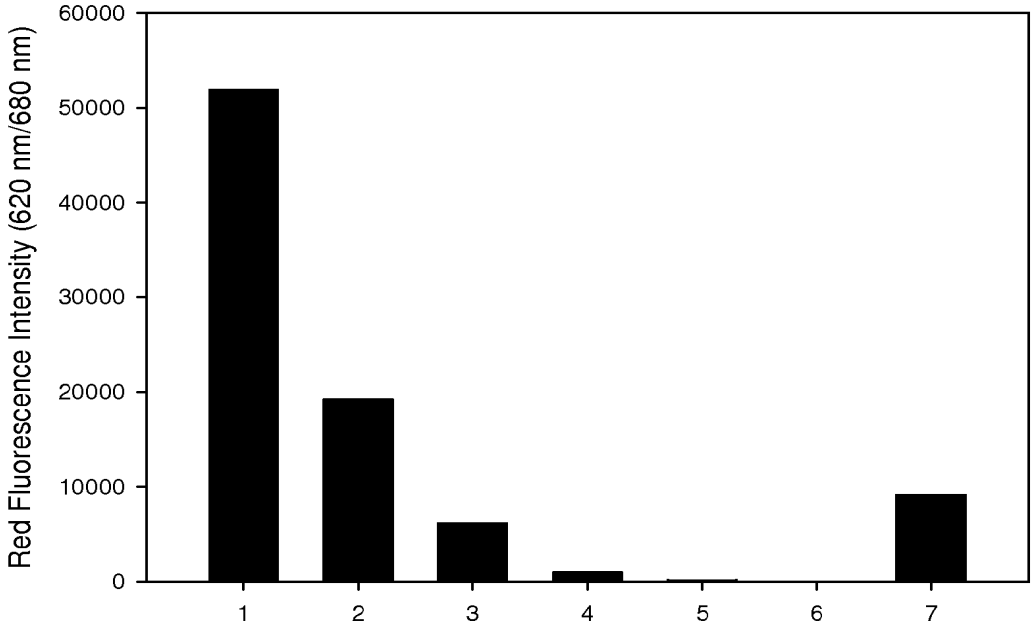


Figure 9

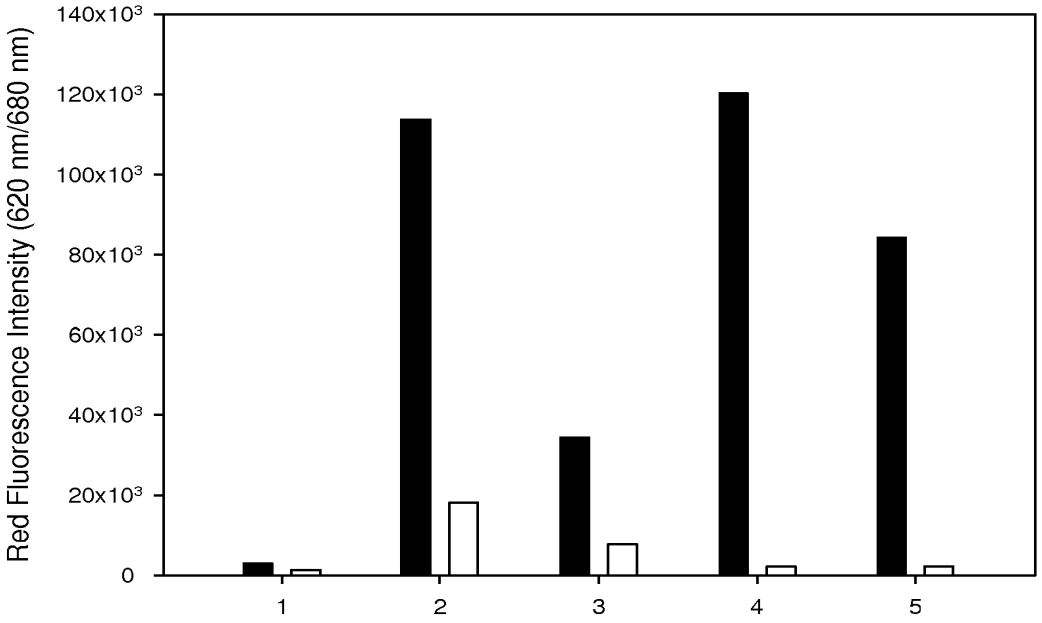


Figure 10A

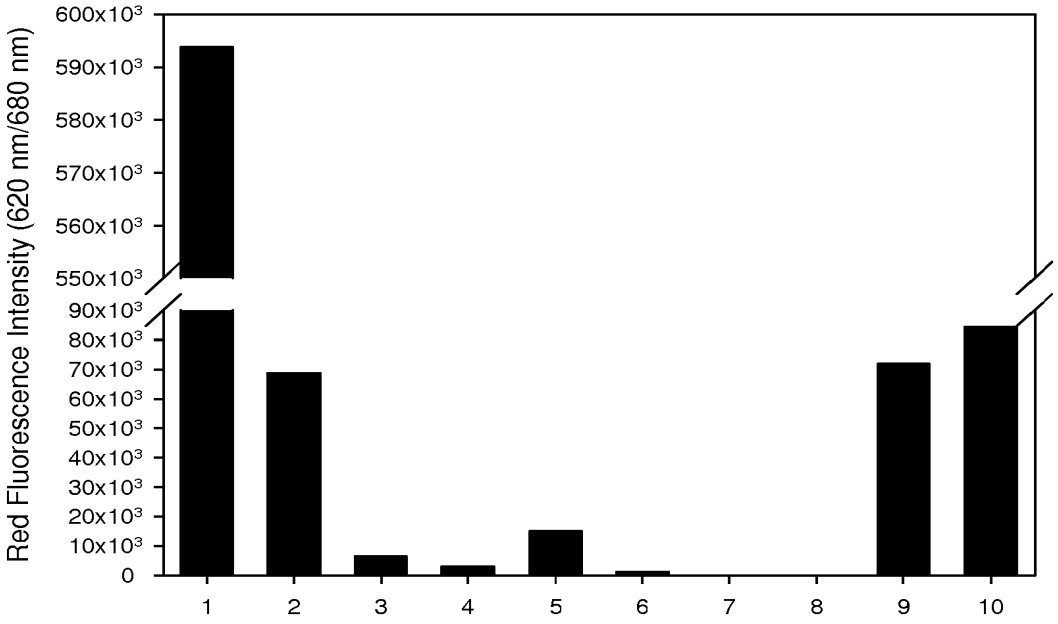


Figure 10B

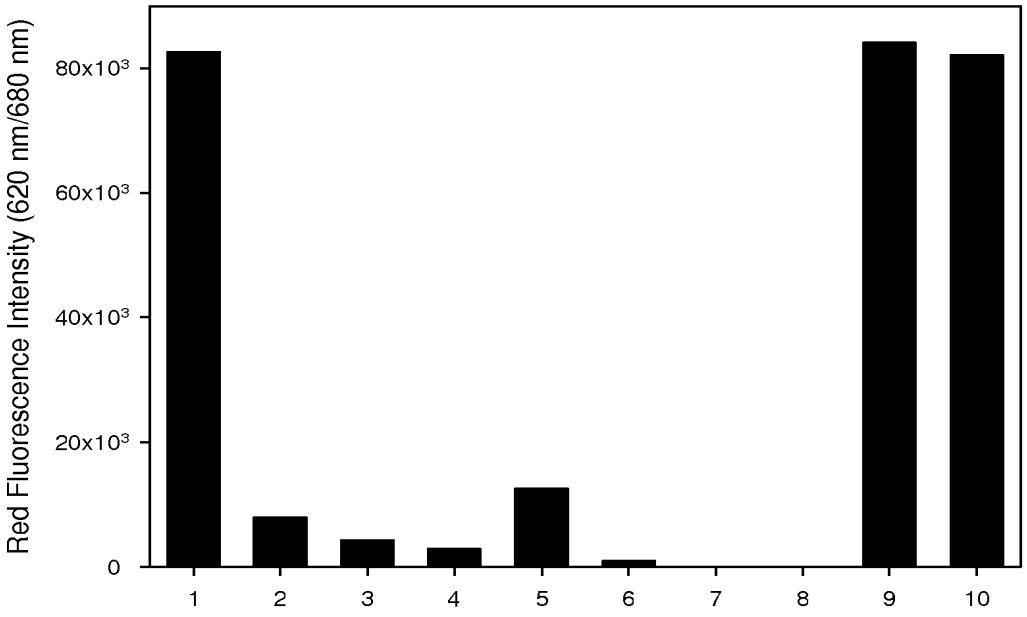


Figure 11 A&B

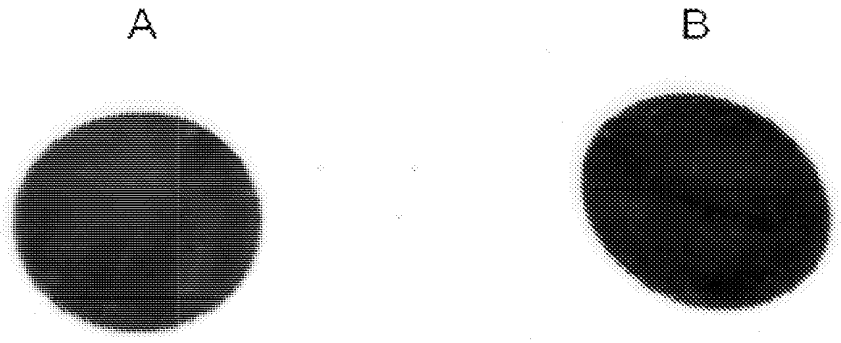


Figure 11C

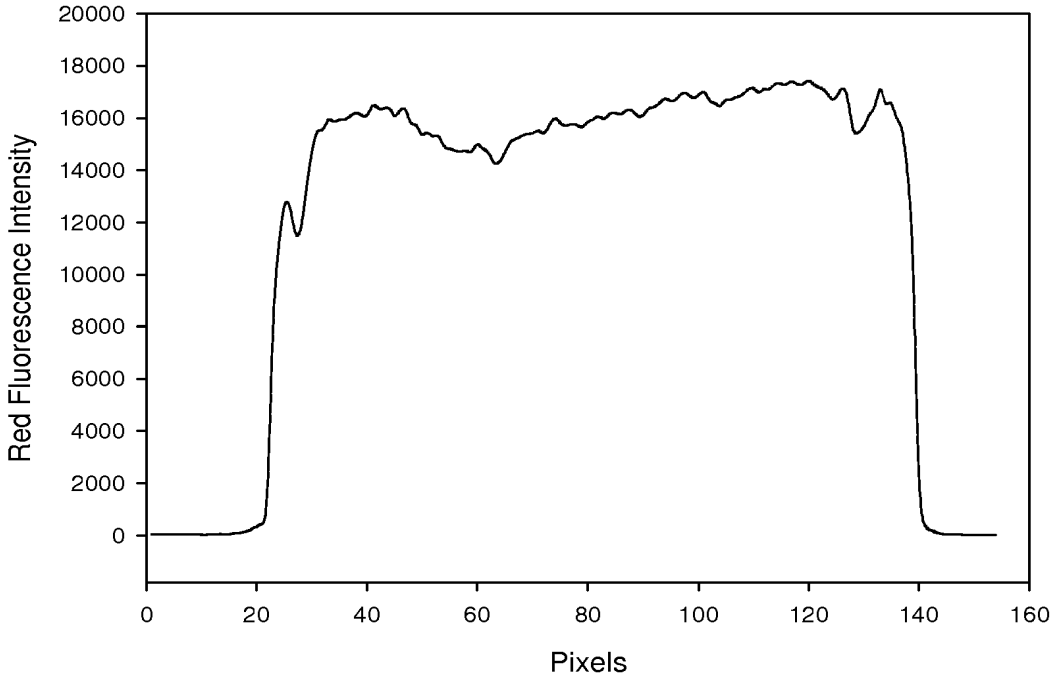


Figure 11D

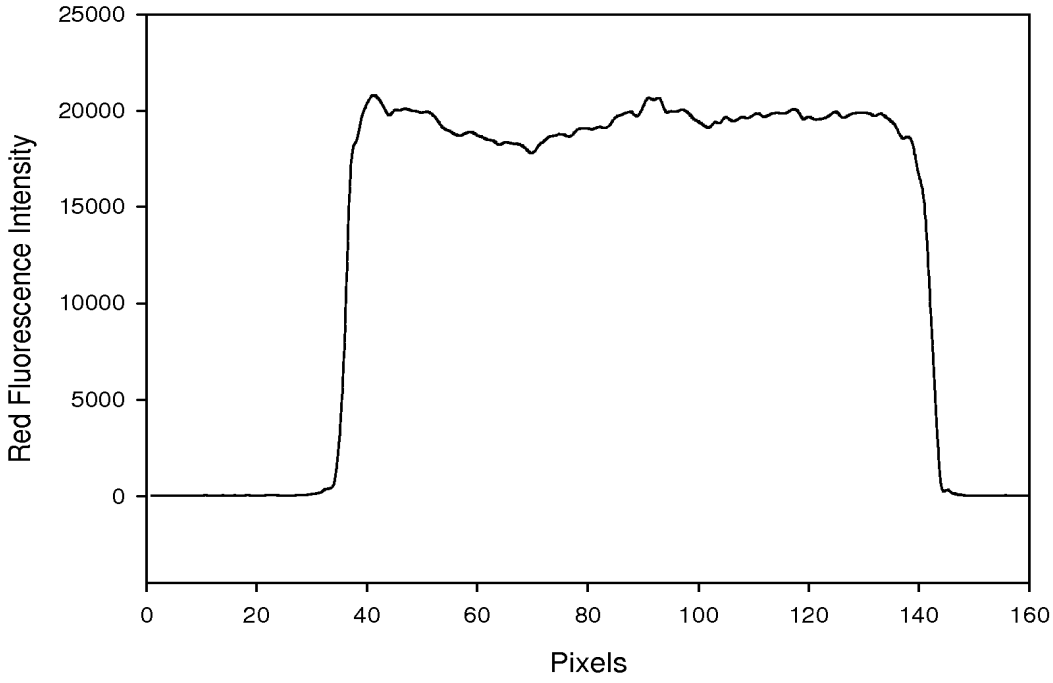


Figure 12A

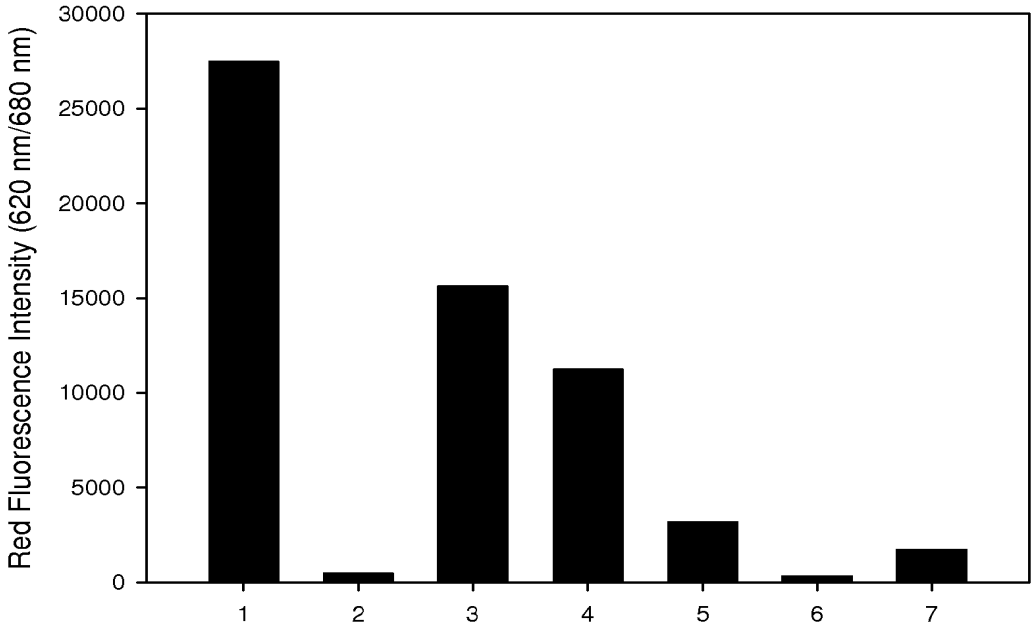


Figure 12B

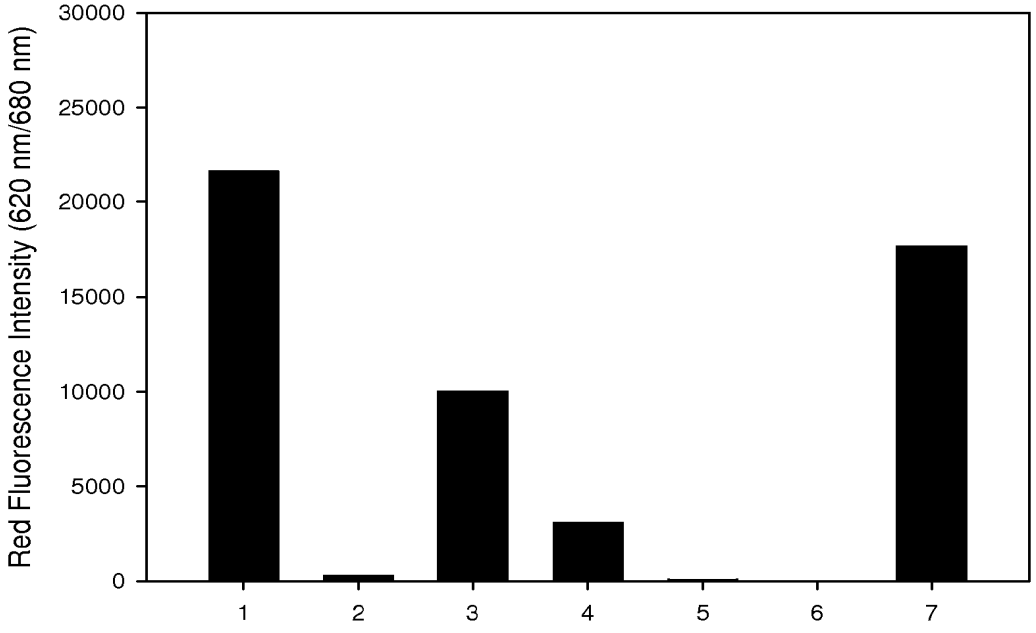


Figure 13

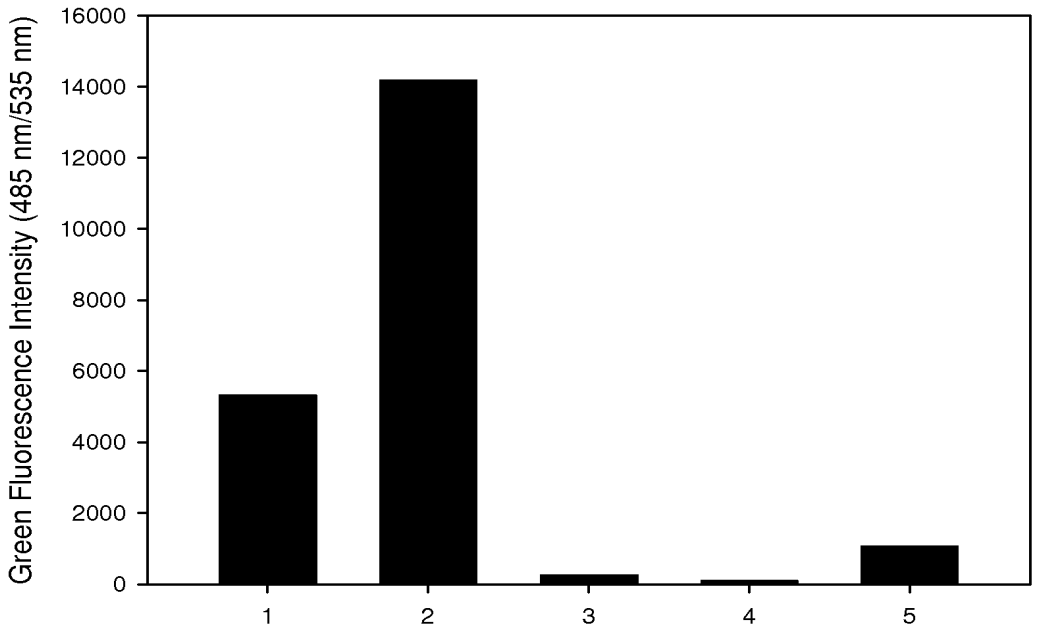


Figure 14

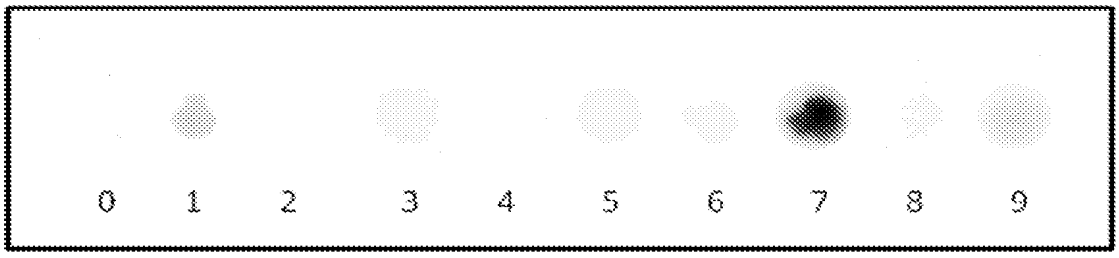


Figure 15

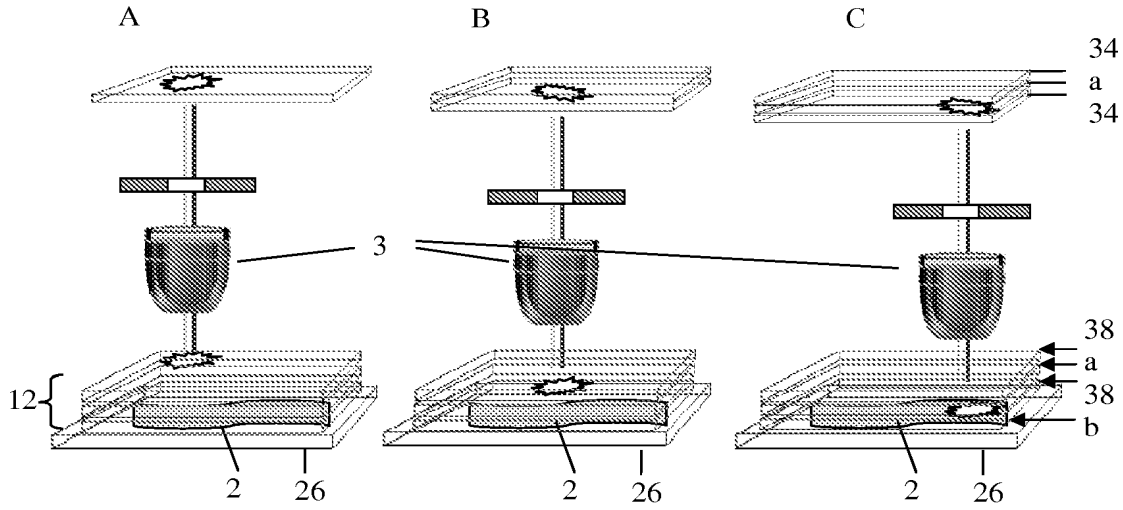


Figure . 16 [types of track-etched membranes used]
A. Polycarbonate B. Polyester

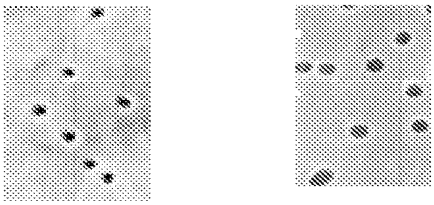
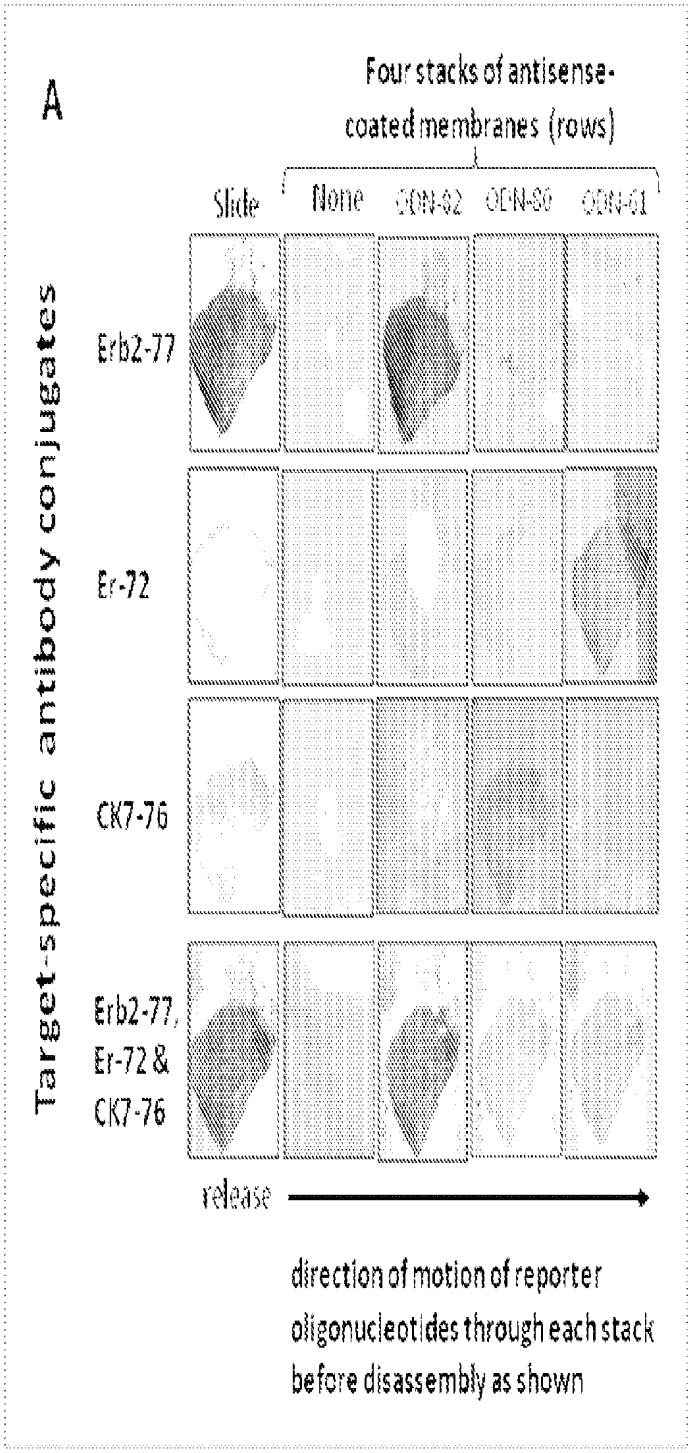


Figure 17


A
Four stacks of antisense-coated membranes (rows)
Slide
None
Target-specific antibody conjugates
Erb2-77
Er-72
CK7-76
Erb2-77, Er-72 & CK7-76
release
direction of motion of reporter oligonucleotides through each stack before disassembly as shown

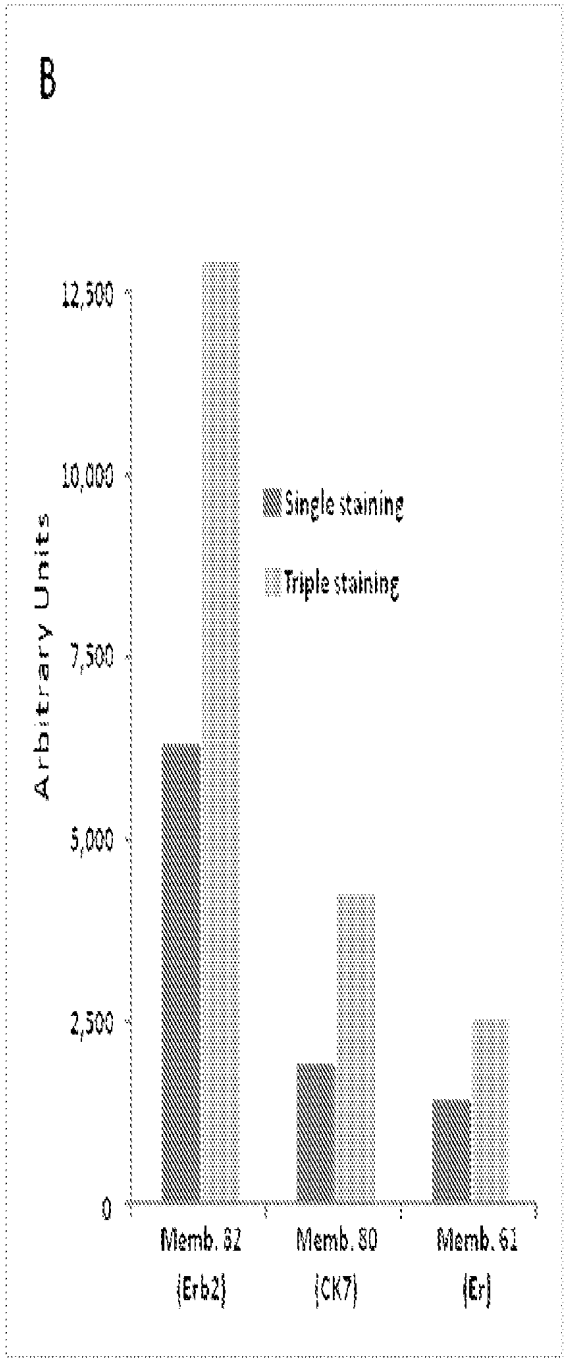
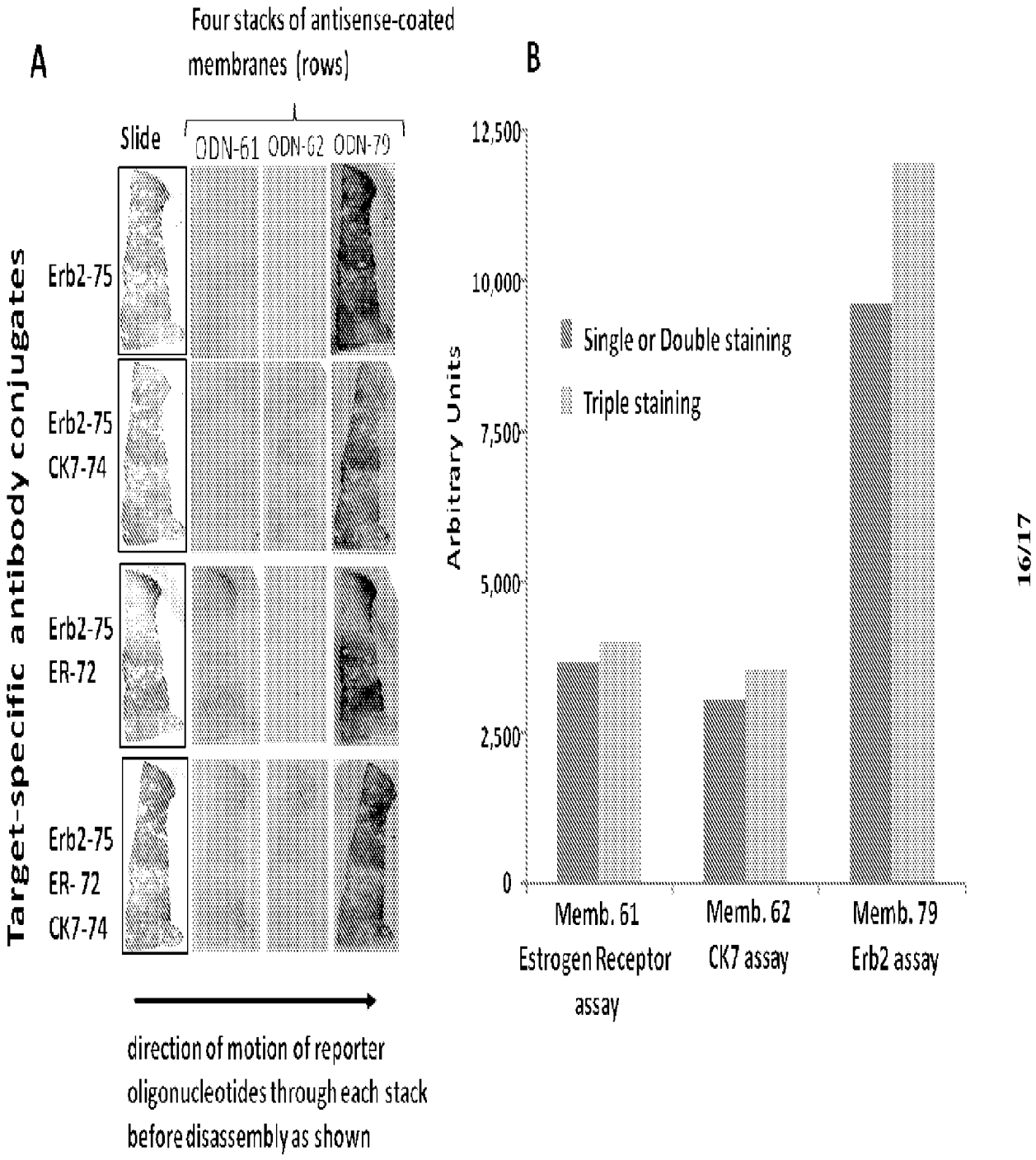
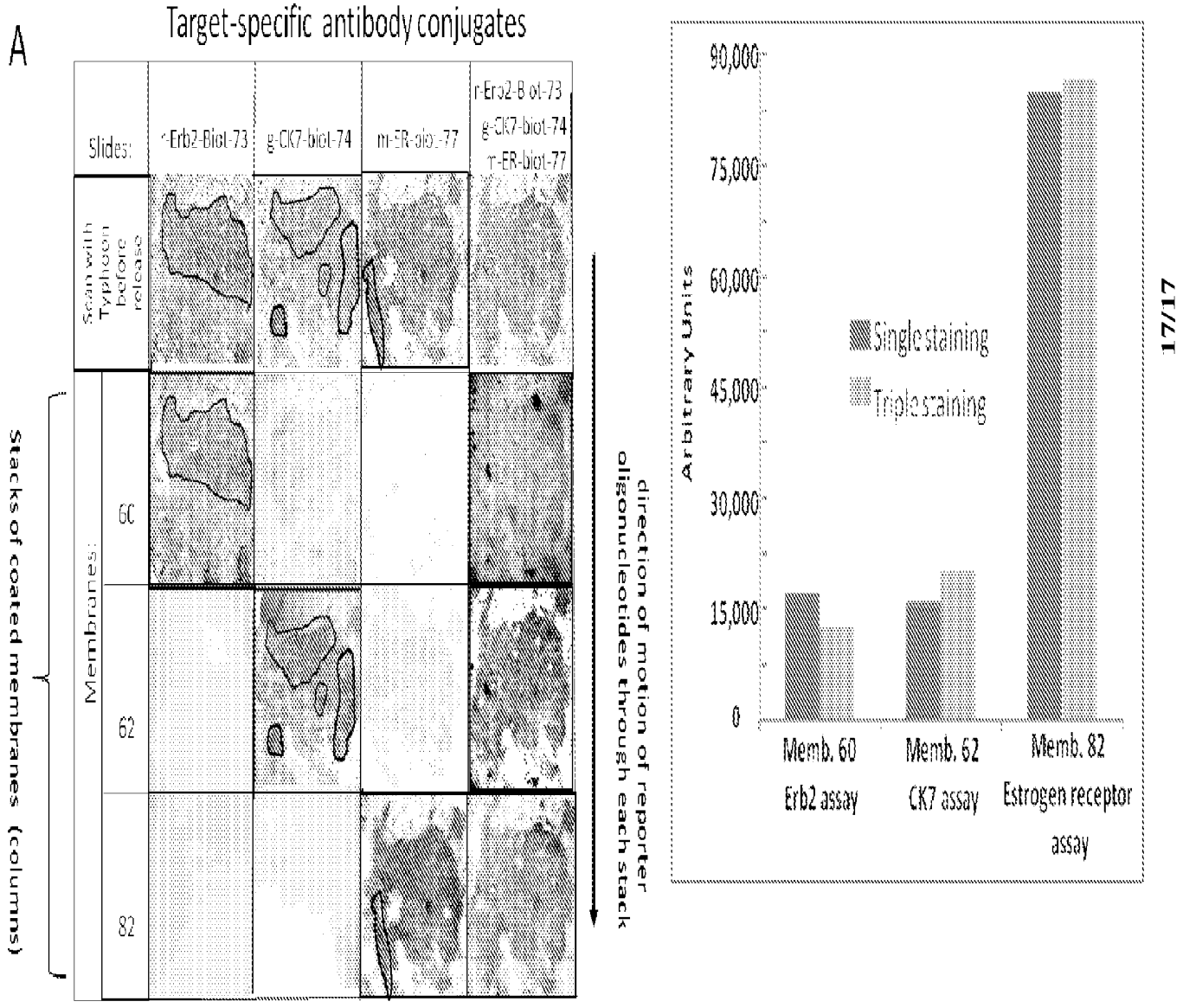

B
Arbitrary Units
12,500
10,000
7,500
5,000
2,500
0
Single staining
Triple staining
Memb. 82 (Erb2)
Memb. 80 (CK7)
Memb. 61 (Er)

Figure 18



16/17

Figure 19



17/17

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2012 (18.10.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/140224 A1

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

(21) International Application Number:
PCT/EP2012/056823

(22) International Filing Date:
13 April 2012 (13.04.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1106254.4 13 April 2011 (13.04.2011) GB

(72) Inventors; and
(71) Applicants : **FRISEN, Jonas** [SE/SE]; Bragevägen 5, SE-114 26 Stockholm (SE). **STÅHL, Patrik** [SE/SE]; Fatburstrappan 18, T3:1, SE-118 26 Stockholm (SE). **LUNDEBERG, Joakim** [SE/SE]; Torsvikssvän 20, SE-181 34 Lidingö (SE).

(74) Agent: **LEATHLEY, Anna**; Dehns, St Bride's House, 10 Salisbury Square, London, Greater London EC4Y 8JD (GB).

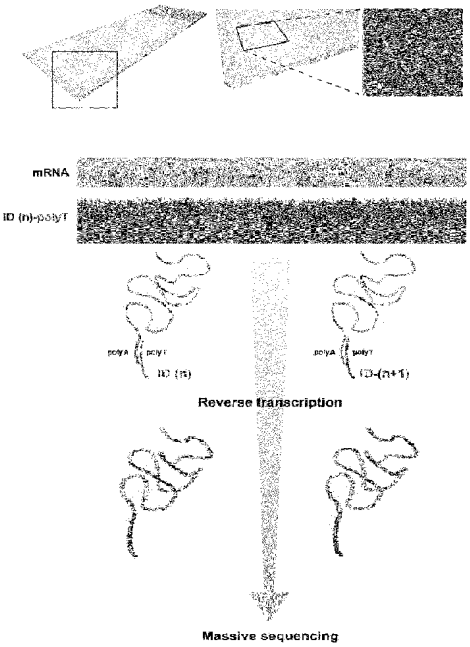
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, MD, 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, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: METHOD AND PRODUCT FOR LOCALISED OR SPATIAL DETECTION OF NUCLEIC ACID IN A TISSUE SAMPLE

Figure 1



(57) Abstract: The present invention relates to methods and products for the localised or spatial detection of nucleic acid in a tissue sample and in particular to a method for localised detection of nucleic acid in a tissue sample comprising: (a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3': (i) a positional domain that corresponds to the position of the capture probe on the array, and (ii) a capture domain; (b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample and allowing nucleic acid of the tissue sample to hybridise to the capture domain in said capture probes; (c) generating DNA molecules from the captured nucleic acid molecules using said capture probes as extension or ligation primers, wherein said extended or ligated DNA molecules are tagged by virtue of the positional domain; (d) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying said tagged DNA; (e) releasing at least part of the tagged DNA molecules and/or their complements or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof; and (f) directly or indirectly analysing the sequence of the released DNA molecules.

WO 2012/140224 A1

Method and product for localised or spatial detection of nucleic acid in a tissue sample

The present invention relates generally to the localised or spatial detection of nucleic acid in a tissue sample. The nucleic acid may be RNA or DNA. Thus, the present invention provides methods for detecting and/or analysing RNA, e.g. RNA transcripts or genomic DNA, so as to obtain spatial information about the localisation, distribution or expression of genes, or indeed about the localisation or distribution of any genomic variation (not necessarily in a gene) in a tissue sample, for example in an individual cell. The present invention thus enables spatial genomics and spatial transcriptomics.

More particularly, the present invention relates to a method for determining and/or analysing a transcriptome or genome and especially the global transcriptome or genome, of a tissue sample. In particular the method relates to a quantitative and/or qualitative method for analysing the distribution, location or expression of genomic sequences in a tissue sample wherein the spatial expression or distribution or location pattern within the tissue sample is retained. Thus, the new method provides a process for performing "spatial transcriptomics" or "spatial genomics", which enables the user to determine simultaneously the expression pattern, or the location/distribution pattern of the genes expressed or genes or genomic loci present in a tissue sample.

The invention is particularly based on array technology coupled with high throughput DNA sequencing technologies, which allows the nucleic acid molecule (e.g. RNA or DNA molecules) in the tissue sample, particularly mRNA or DNA, to be captured and labelled with a positional tag. This step is followed by synthesis of DNA molecules which are sequenced and analysed to determine which genes are expressed in any and all parts of the tissue sample. Advantageously, the individual, separate and specific transcriptome of each cell in the tissue sample may be obtained at the same time. Hence, the methods of the invention may be said to provide highly parallel comprehensive transcriptome signatures from individual cells within a tissue sample without losing spatial information within said investigated tissue sample. The invention also provides an array for performing the method of the invention and methods for making the arrays of the invention.

The human body comprises over 100 trillion cells and is organized into more than 250 different organs and tissues. The development and organization of

complex organs, such as the brain, are far from understood and there is a need to dissect the expression of genes expressed in such tissues using quantitative methods to investigate and determine the genes that control the development and function of such tissues. The organs are in themselves a mixture of differentiated cells that enable all bodily functions, such as nutrient transport, defence etc. to be coordinated and maintained. Consequently, cell function is dependent on the position of the cell within a particular tissue structure and the interactions it shares with other cells within that tissue, both directly and indirectly. Hence, there is a need to disentangle how these interactions influence each cell within a tissue at the transcriptional level.

Recent findings by deep RNA sequencing have demonstrated that a majority of the transcripts can be detected in a human cell line and that a large fraction (75%) of the human protein-coding genes are expressed in most tissues. Similarly, a detailed study of 1% of the human genome showed that chromosomes are ubiquitously transcribed and that the majority of all bases are included in primary transcripts. The transcription machinery can therefore be described as promiscuous at a global level.

It is well-known that transcripts are merely a proxy for protein abundance, because the rates of RNA translation, degradation etc will influence the amount of protein produced from any one transcript. In this respect, a recent antibody-based analysis of human organs and tissues suggests that tissue specificity is achieved by precise regulation of protein levels in space and time, and that different tissues in the body acquire their unique characteristics by controlling not which proteins are expressed but how much of each is produced.

However, in subsequent global studies transcriptome and proteome correlations have been compared demonstrating that the majority of all genes were shown to be expressed. Interestingly, there was shown to be a high correlation between changes in RNA and protein levels for individual gene products which is indicative of the biological usefulness of studying the transcriptome in individual cells in the context of the functional role of proteins.

Indeed, analysis of the histology and expression pattern in tissues is a cornerstone in biomedical research and diagnostics. Histology, utilizing different staining techniques, first established the basic structural organization of healthy organs and the changes that take place in common pathologies more than a

century ago. Developments in this field resulted in the possibility of studying protein distribution by immunohistochemistry and gene expression by *in situ* hybridization.

However, the parallel development of increasingly advanced histological and gene expression techniques has resulted in the separation of imaging and transcriptome analysis and, until the methods of the present invention, there has not been any feasible method available for global transcriptome analysis with spatial resolution.

As an alternative, or in addition, to *in situ* techniques, methods have developed for the *in vitro* analysis of proteins and nucleic acids, i.e. by extracting molecules from whole tissue samples, single cell types, or even single cells, and quantifying specific molecules in said extracts, e.g. by ELISA, qPCR etc.

Recent developments in the analysis of gene expression have resulted in the possibility of assessing the complete transcriptome of tissues using microarrays or RNA sequencing, and such developments have been instrumental in our understanding of biological processes and for diagnostics. However, transcriptome analysis typically is performed on mRNA extracted from whole tissues (or even whole organisms), and methods for collecting smaller tissue areas or individual cells for transcriptome analysis are typically labour intensive, costly and have low precision.

Hence, the majority of gene expression studies based on microarrays or next generation sequencing of RNA use a representative sample containing many cells. Thus the results represent the average expression levels of the investigated genes. The separation of cells that are phenotypically different has been used in some cases together with the global gene expression platforms (Tang F *et al*, Nat Protoc. 2010; 5: 516-35; Wang D & Bodovitz S, Trends Biotechnol. 2010; 28:281-90) and resulted in very precise information about cell-to-cell variations. However, high throughput methods to study transcriptional activity with high resolution in intact tissues have not, until now, been available.

Thus, existing techniques for the analysis of gene expression patterns provide spatial transcriptional information only for one or a handful of genes at a time or offer transcriptional information for all of the genes in a sample at the cost of losing positional information. Hence, it is evident that methods to determine simultaneously, separately and specifically the transcriptome of each cell in a sample are required, i.e. to enable global gene expression analysis in tissue

samples that yields transcriptomic information with spatial resolution, and the present invention addresses this need.

The novel approach of the methods and products of the present invention utilizes now well established array and sequencing technology to yield transcriptional information for all of the genes in a sample, whilst retaining the positional information for each transcript. It will be evident to the person of skill in the art that this represents a milestone in the life sciences. The new technology opens a new field of so-called "spatial transcriptomics", which is likely to have profound consequences for our understanding of tissue development and tissue and cellular function in all multicellular organisms. It will be apparent that such techniques will be particularly useful in our understanding of the cause and progress of disease states and in developing effective treatments for such diseases, e.g. cancer. The methods of the invention will also find uses in the diagnosis of numerous medical conditions.

Whilst initially conceived with the aim of transcriptome analysis in mind, as described in detail below, the principles and methods of the present invention may be applied also to the analysis of DNA and hence for genomic analyses also ("spatial genomics"). Accordingly, at its broadest the invention pertains to the detection and/or analysis of nucleic acid in general.

Array technology, particularly microarrays, arose from research at Stanford University where small amounts of DNA oligonucleotides were successfully attached to a glass surface in an ordered arrangement, a so-called "array", and used it to monitor the transcription of 45 genes (Schena M *et al*, Science. 1995; 270: 368-9, 371).

Since then, researchers around the world have published more than 30,000 papers using microarray technology. Multiple types of microarray have been developed for various applications, e.g. to detect single nucleotide polymorphisms (SNPs) or to genotype or re-sequence mutant genomes, and an important use of microarray technology has been for the investigation of gene expression. Indeed, the gene expression microarray was created as a means to analyze the level of expressed genetic material in a particular sample, with the real gain being the possibility to compare expression levels of many genes simultaneously. Several commercial microarray platforms are available for these types of experiments but it has also been possible to create custom made gene expression arrays.

Whilst the use of microarrays in gene expression studies is now commonplace, it is evident that new and more comprehensive so-called "next-generation DNA sequencing" (NGS) technologies are starting to replace DNA microarrays for many applications, e.g. in-depth transcriptome analysis.

The development of NGS technologies for ultra-fast genome sequencing represents a milestone in the life sciences (Petterson E *et al*, *Genomics.* 2009; 93: 105-11). These new technologies have dramatically decreased the cost of DNA sequencing and enabled the determination of the genome of higher organisms at an unprecedented rate, including those of specific individuals (Wade CM *et al* Science. 2009; 326: 865-7; Rubin J *et al, Nature* 2010; 464: 587-91). The new advances in high-throughput genomics have reshaped the biological research landscape and in addition to complete characterization of genomes it is possible also to study the full transcriptome in a digital and quantitative fashion. The bioinformatics tools to visualize and integrate these comprehensive sets of data have also been significantly improved during recent years.

However, it has surprisingly been found that a unique combination of histological, microarray and NGS techniques can yield comprehensive transcriptional or genomic information from multiple cells in a tissue sample which information is characterised by a two-dimensional spatial resolution. Thus, at one extreme the methods of the present invention can be used to analyse the expression of a single gene in a single cell in a sample, whilst retaining the cell within its context in the tissue sample. At the other extreme, and in a preferred aspect of the invention, the methods can be used to determine the expression of every gene in each and every cell, or substantially all cells, in a sample simultaneously, i.e. the global spatial expression pattern of a tissue sample. It will be apparent that the methods of the invention also enable intermediate analyses to be performed.

In its simplest form, the invention may be illustrated by the following summary. The invention requires reverse transcription (RT) primers, which comprise also unique positional tags (domains), to be arrayed on an object substrate, e.g. a glass slide, to generate an "array". The unique positional tags correspond to the location of the RT primers on the array (the features of the array). Thin tissue sections are placed onto the array and a reverse transcription reaction is performed in the tissue section on the object slide. The RT primers, to which the RNA in the tissue sample binds (or hybridizes), are extended using the bound RNA

as a template to obtain cDNA, which is therefore bound to the surface of the array. As consequence of the unique positional tags in the RT primers, each cDNA strand carries information about the position of the template RNA in the tissue section. The tissue section may be visualised or imaged, e.g. stained and photographed, before or after the cDNA synthesis step to enable the positional tag in the cDNA molecule to be correlated with a position within the tissue sample. The cDNA is sequenced, which results in a transcriptome with exact positional information. A schematic of the process is shown in Figure 1. The sequence data can then be matched to a position in the tissue sample, which enables the visualization, e.g. using a computer, of the sequence data together with the tissue section, for instance to display the expression pattern of any gene of interest across the tissue (Figure 2). Similarly, it would be possible to mark different areas of the tissue section on the computer screen and obtain information on differentially expressed genes between any selected areas of interest. It will be evident that the methods of the invention result in data that is in stark contrast to the data obtained using current methods to study mRNA populations. For example, methods based on *in situ* hybridization provide only relative information of single mRNA transcripts. Thus, the methods of the present invention have clear advantages over current *in situ* technologies. The global gene expression information obtainable from the methods of the invention also allows co-expression information and quantitative estimates of transcript abundance. It will be evident that this is a generally applicable strategy available for the analysis of any tissue in any species, e.g. animal, plant, fungus.

As noted above, and described in more detail below, it will be evident that this basic methodology could readily be extended to the analysis of genomic DNA, e.g. to identify cells within a tissue sample that comprise one or more specific mutations. For instance, the genomic DNA may be fragmented and allowed to hybridise to primers (equivalent to the RT primers described above), which are capable of capturing the fragmented DNA (e.g. an adapter with a sequence that is complementary to the primer may be ligated to the fragmented DNA or the fragmented DNA may be extended e.g. using an enzyme to incorporate additional nucleotides at the end of the sequence, e.g. a poly-A tail, to generate a sequence that is complementary to the primer) and priming the synthesis of complementary strands to the capture molecules. The remaining steps of the analysis may be as described above. Hence, the specific embodiments of the invention described

below in the context of transcriptome analysis may also be employed in methods of analysing genomic DNA, where appropriate.

It will be seen from the above explanation that there is an immense value in coupling positional information to transcriptome or genome information. For instance, it enables global gene expression mapping at high resolution, which will find utility in numerous applications, including e.g. cancer research and diagnostics.

Furthermore, it is evident that the methods described herein differ significantly from the previously described methods for analysis of the global transcriptome of a tissue sample and these differences result in numerous advantages. The present invention is predicated on the surprising discovery that the use of tissue sections does not interfere with synthesis of DNA (e.g. cDNA) primed by primers (e.g. reverse transcription primers) that are coupled to the surface of an array.

Thus, in its first and broadest aspect, the present invention provides a method for localised detection of nucleic acid in a tissue sample comprising:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample and allowing nucleic acid of the tissue sample to hybridise to the capture domain in said capture probes;

(c) generating DNA molecules from the captured nucleic acid molecules using said capture probes as extension or ligation primers, wherein said extended or ligated DNA molecules are tagged by virtue of the positional domain;

(d) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying said tagged DNA;

(e) releasing at least part of the tagged DNA molecules and/or their complements or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof ;

(f) directly or indirectly analysing the sequence of the released DNA molecules.

The methods of the invention represent a significant advance over other methods for spatial transcriptomics known in the art. For example the methods described herein result in a global and spatial profile of all transcripts in the tissue sample. Moreover, the expression of every gene can be quantified for each position or feature on the array, which enables a multiplicity of analyses to be performed based on data from a single assay. Thus, the methods of the present invention make it possible to detect and/or quantify the spatial expression of all genes in single tissue sample. Moreover, as the abundance of the transcripts is not visualised directly, e.g. by fluorescence, akin to a standard microarray, it is possible to measure the expression of genes in a single sample simultaneously even wherein said transcripts are present at vastly different concentrations in the same sample.

Accordingly, in a second and more particular aspect, the present invention can be seen to provide a method for determining and/or analysing a transcriptome of a tissue sample comprising:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample and allowing RNA of the tissue sample to hybridise to the capture domain in said capture probes;

(c) generating cDNA molecules from the captured RNA molecules using said capture probes as RT primers, and optionally amplifying said cDNA molecules;

(d) releasing at least part of the cDNA molecules and/or optionally their amplicons from the surface of the array, wherein said released molecule may be a first strand and/or second strand cDNA molecule or an amplicon thereof and wherein said part includes the positional domain or a complement thereof;

(e) directly or indirectly analysing the sequence of the released molecules.

As described in more detail below, any method of nucleic acid analysis may be used in the analysis step. Typically this may involve sequencing, but it is not necessary to perform an actual sequence determination. For example sequence-specific methods of analysis may be used. For example a sequence-specific amplification reaction may be performed, for example using primers which are specific for the positional domain and/or for a specific target sequence, e.g. a particular target DNA to be detected (i.e. corresponding to a particular cDNA/RNA or gene etc.). An exemplary analysis method is a sequence-specific PCR reaction.

The sequence analysis information obtained in step (e) may be used to obtain spatial information as to the RNA in the sample. In other words the sequence analysis information may provide information as to the location of the RNA in the sample. This spatial information may be derived from the nature of the sequence analysis information determined, for example it may reveal the presence of a particular RNA which may itself be spatially informative in the context of the tissue sample used, and/or the spatial information (e.g. spatial localisation) may be derived from the position of the tissue sample on the array, coupled with the sequencing information. Thus, the method may involve simply correlating the sequence analysis information to a position in the tissue sample e.g. by virtue of the positional tag and its correlation to a position in the tissue sample. However, as described above, spatial information may conveniently be obtained by correlating the sequence analysis data to an image of the tissue sample and this represents one preferred embodiment of the invention. Accordingly, in a preferred embodiment the method also includes a step of:

(f) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (c).

In its broadest sense, the method of the invention may be used for localised detection of a nucleic acid in a tissue sample. Thus, in one embodiment, the method of the invention may be used for determining and/or analysing all of the transcriptome or genome of a tissue sample e.g. the global transcriptome of a tissue sample. However, the method is not limited to this and encompasses

determining and/or analysing all or part of the transcriptome or genome. Thus, the method may involve determining and/or analysing a part or subset of the transcriptome or genome, e.g. a transcriptome corresponding to a subset of genes, e.g. a set of particular genes, for example related to a particular disease or condition, tissue type etc.

Viewed from another aspect, the method steps set out above can be seen as providing a method of obtaining a spatially defined transcriptome or genome, and in particular the spatially defined global transcriptome or genome, of a tissue sample.

Alternatively viewed, the method of the invention may be seen as a method for localised or spatial detection of nucleic acid, whether DNA or RNA in a tissue sample, or for localised or spatial determination and/or analysis of nucleic acid (DNA or RNA) in a tissue sample. In particular, the method may be used for the localised or spatial detection or determination and/or analysis of gene expression or genomic variation in a tissue sample. The localised/spatial detection/determination/analysis means that the RNA or DNA may be localised to its native position or location within a cell or tissue in the tissue sample. Thus for example, the RNA or DNA may be localised to a cell or group of cells, or type of cells in the sample, or to particular regions of areas within a tissue sample. The native location or position of the RNA or DNA (or in other words, the location or position of the RNA or DNA in the tissue sample), e.g. an expressed gene or genomic locus, may be determined.

The invention can also be seen to provide an array for use in the methods of the invention comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain to capture RNA of a tissue sample that is contacted with said array.

In a related aspect, the present invention also provides use of an array, comprising a substrate on which multiple species of capture probe are directly or indirectly immobilized such that each species occupies a distinct position on the

array and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array; and

(ii) a capture domain;

to capture RNA of a tissue sample that is contacted with said array.

Preferably, said use is for determining and/or analysing a transcriptome and in particular the global transcriptome, of a tissue sample and further comprises steps of:

(a) generating cDNA molecules from the captured RNA molecules using said capture probes as RT primers and optionally amplifying said cDNA molecules;

(b) releasing at least part of the cDNA molecules and/or optionally their amplicons from the surface of the array, wherein said released molecule may be a first strand and/or second strand cDNA molecule or an amplicon thereof and wherein said part includes the positional domain or a complement thereof;

(c) directly or indirectly analysing the sequence of the released molecules; and optionally

(d) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (a).

It will be seen therefore that the array of the present invention may be used to capture RNA, e.g. mRNA of a tissue sample that is contacted with said array. The array may also be used for determining and/or analysing a partial or global transcriptome of a tissue sample or for obtaining a spatially defined partial or global transcriptome of a tissue sample. The methods of the invention may thus be considered as methods of quantifying the spatial expression of one or more genes in a tissue sample. Expressed another way, the methods of the present invention may be used to detect the spatial expression of one or more genes in a tissue sample. In yet another way, the methods of the present invention may be used to determine simultaneously the expression of one or more genes at one or more positions within a tissue sample. Still further, the methods may be seen as methods for partial or global transcriptome analysis of a tissue sample with two-dimensional spatial resolution.

The RNA may be any RNA molecule which may occur in a cell. Thus it may be mRNA, tRNA, rRNA, viral RNA, small nuclear RNA (snRNA), small nucleolar

RNA (snoRNA), microRNA (miRNA), small interfering RNA (siRNA), piwi-interacting RNA (piRNA), ribozymal RNA, antisense RNA or non-coding RNA. Preferably however it is mRNA.

Step (c) in the method above (corresponding to step (a) in the preferred statement of use set out above) of generating cDNA from the captured RNA will be seen as relating to the synthesis of the cDNA. This will involve a step of reverse transcription of the captured RNA, extending the capture probe, which functions as the RT primer, using the captured RNA as template. Such a step generates so-called first strand cDNA. As will be described in more detail below, second strand cDNA synthesis may optionally take place on the array, or it may take place in a separate step, after release of first strand cDNA from the array. As also described in more detail below, in certain embodiments second strand synthesis may occur in the first step of amplification of a released first strand cDNA molecule.

Arrays for use in the context of nucleic acid analysis in general, and DNA analysis in particular, are discussed and described below. Specific details and embodiments described herein in relation to arrays and capture probes for use in the context of RNA, apply equally (where appropriate) to all such arrays, including those for use with DNA.

As used herein the term "multiple" means two or more, or at least two, e.g. 3, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 400, 500, 1000, 2000, 5000, 10,000, or more etc. Thus for example, the number of capture probes may be any integer in any range between any two of the aforementioned numbers. It will be appreciated however that it is envisaged that conventional-type arrays with many hundreds, thousands, tens of thousands, hundreds of thousands or even millions of capture probes may be used.

Thus, the methods outlined herein utilise high density nucleic acid arrays comprising "capture probes" for capturing and labelling transcripts from all of the single cells within a tissue sample e.g. a thin tissue sample slice, or "section". The tissue samples or sections for analysis are produced in a highly parallelized fashion, such that the spatial information in the section is retained. The captured RNA (preferably mRNA) molecules for each cell, or "transcriptomes", are transcribed into cDNA and the resultant cDNA molecules are analyzed, for example by high throughput sequencing. The resultant data may be correlated to images of the original tissue samples e.g. sections through so-called barcode sequences (or

ID tags, defined herein as positional domains) incorporated into the arrayed nucleic acid probes.

High density nucleic acid arrays or microarrays are a core component of the spatial transcriptome labelling method described herein. A microarray is a multiplex technology used in molecular biology. A typical microarray consists of an arrayed series of microscopic spots of oligonucleotides (hundreds of thousands of spots, generally tens of thousands, can be incorporated on a single array). The distinct position of each nucleic acid (oligonucleotide) spot (each species of oligonucleotide/nucleic acid molecule) is known as a "feature" (and hence in the methods set out above each species of capture probe may be viewed as a specific feature of the array; each feature occupies a distinct position on the array), and typically each separate feature contains in the region of picomoles (10^{-12} moles) of a specific DNA sequence (a "species"), which are known as "probes" (or "reporters"). Typically, these can be a short section of a gene or other nucleic acid element to which a cDNA or cRNA sample (or "target") can hybridize under high-stringency hybridization conditions. However, as described below, the probes of the present invention differ from the probes of standard microarrays.

In gene expression microarrays, probe-target hybridization is usually detected and quantified by detection of visual signal, e.g. a fluorophore, silver ion, or chemiluminescence-label, which has been incorporated into all of the targets. The intensity of the visual signal correlates to the relative abundance of each target nucleic acid in the sample. Since an array can contain tens of thousands of probes, a microarray experiment can accomplish many genetic tests in parallel.

In standard microarrays, the probes are attached to a solid surface or substrate by a covalent bond to a chemical matrix, e.g. epoxy-silane, amino-silane, lysine, polyacrylamide etc. The substrate typically is a glass, plastic or silicon chip or slide, although other microarray platforms are known, e.g. microscopic beads.

The probes may be attached to the array of the invention by any suitable means. In a preferred embodiment the probes are immobilized to the substrate of the array by chemical immobilization. This may be an interaction between the substrate (support material) and the probe based on a chemical reaction. Such a chemical reaction typically does not rely on the input of energy via heat or light, but can be enhanced by either applying heat, e.g. a certain optimal temperature for a chemical reaction, or light of certain wavelength. For example, a chemical immobilization may take place between functional groups on the substrate and

corresponding functional elements on the probes. Such corresponding functional elements in the probes may either be an inherent chemical group of the probe, e.g. a hydroxyl group or be additionally introduced. An example of such a functional group is an amine group. Typically, the probe to be immobilized comprises a functional amine group or is chemically modified in order to comprise a functional amine group. Means and methods for such a chemical modification are well known.

The localization of said functional group within the probe to be immobilized may be used in order to control and shape the binding behaviour and/or orientation of the probe, e.g. the functional group may be placed at the 5' or 3' end of the probe or within sequence of the probe. A typical substrate for a probe to be immobilized comprises moieties which are capable of binding to such probes, e.g. to amine-functionalized nucleic acids. Examples of such substrates are carboxy, aldehyde or epoxy substrates. Such materials are known to the person skilled in the art. Functional groups, which impart a connecting reaction between probes which are chemically reactive by the introduction of an amine group, and array substrates are known to the person skilled in the art.

Alternative substrates on which probes may be immobilized may have to be chemically activated, e.g. by the activation of functional groups, available on the array substrate. The term "activated substrate" relates to a material in which interacting or reactive chemical functional groups were established or enabled by chemical modification procedures as known to the person skilled in the art. For example, a substrate comprising carboxyl groups has to be activated before use. Furthermore, there are substrates available that contain functional groups that can react with specific moieties already present in the nucleic acid probes.

Alternatively, the probes may be synthesized directly on the substrate. Suitable methods for such an approach are known to the person skilled in the art. Examples are manufacture techniques developed by Agilent Inc., Affymetrix Inc., Roche Nimblegen Inc. or Flexgen BV. Typically, lasers and a set of mirrors that specifically activate the spots where nucleotide additions are to take place are used. Such an approach may provide, for example, spot sizes (i.e. features) of around 30 µm or larger.

The substrate therefore may be any suitable substrate known to the person skilled in the art. The substrate may have any suitable form or format, e.g. it may be flat, curved, e.g. convexly or concavely curved towards the area where the

interaction between the tissue sample and the substrate takes place. Particularly preferred is the where the substrate is a flat, i.e. planar, chip or slide.

Typically, the substrate is a solid support and thereby allows for an accurate and traceable positioning of the probes on the substrate. An example of a substrate is a solid material or a substrate comprising functional chemical groups, e.g. amine groups or amine-functionalized groups. A substrate envisaged by the present invention is a non-porous substrate. Preferred non-porous substrates are glass, silicon, poly-L-lysine coated material, nitrocellulose, polystyrene, cyclic olefin copolymers (COCs), cyclic olefin polymers (COPs), polypropylene, polyethylene and polycarbonate.

Any suitable material known to the person skilled in the art may be used. Typically, glass or polystyrene is used. Polystyrene is a hydrophobic material suitable for binding negatively charged macromolecules because it normally contains few hydrophilic groups. For nucleic acids immobilized on glass slides, it is furthermore known that by increasing the hydrophobicity of the glass surface the nucleic acid immobilization may be increased. Such an enhancement may permit a relatively more densely packed formation. In addition to a coating or surface treatment with poly-L-lysine, the substrate, in particular glass, may be treated by silanation, e.g. with epoxy-silane or amino-silane or by silynation or by a treatment with polyacrylamide.

A number of standard arrays are commercially available and both the number and size of the features may be varied. In the present invention, the arrangement of the features may be altered to correspond to the size and/or density of the cells present in different tissues or organisms. For instance, animal cells typically have a cross-section in the region of 1-100µm, whereas the cross-section of plant cells typically may range from 1-10000µm. Hence, Nimblegen® arrays, which are available with up to 2.1 million features, or 4.2 million features, and feature sizes of 13 micrometers, may be preferred for tissue samples from an animal or fungus, whereas other formats, e.g. with 8x130k features, may be sufficient for plant tissue samples. Commercial arrays are also available or known for use in the context of sequence analysis and in particular in the context of NGS technologies. Such arrays may also be used as the array surface in the context of the present invention e.g. an Illumina bead array. In addition to commercially available arrays, which can themselves be customized, it is possible to make custom or non-standard "in-house" arrays and methods for generating arrays are

well-established. The methods of the invention may utilise both standard and non-standard arrays that comprise probes as defined below.

The probes on a microarray may be immobilized, i.e. attached or bound, to the array preferably via the 5' or 3' end, depending on the chemical matrix of the array. Typically, for commercially available arrays, the probes are attached via a 3' linkage, thereby leaving a free 5' end. However, arrays comprising probes attached to the substrate via a 5' linkage, thereby leaving a free 3' end, are available and may be synthesized using standard techniques that are well known in the art and are described elsewhere herein.

The covalent linkage used to couple a nucleic acid probe to an array substrate may be viewed as both a direct and indirect linkage, in that the although the probe is attached by a "direct" covalent bond, there may be a chemical moiety or linker separating the "first" nucleotide of the nucleic acid probe from the, e.g. glass or silicon, substrate i.e. an indirect linkage. For the purposes of the present invention probes that are immobilized to the substrate by a covalent bond and/or chemical linker are generally seen to be immobilized or attached directly to the substrate.

As will be described in more detail below, the capture probes of the invention may be immobilized on, or interact with, the array directly or indirectly. Thus the capture probes need not bind directly to the array, but may interact indirectly, for example by binding to a molecule which itself binds directly or indirectly to the array (e.g. the capture probe may interact with (e.g. bind or hybridize to) a binding partner for the capture probe, i.e. a surface probe, which is itself bound to the array directly or indirectly). Generally speaking, however, the capture probe will be, directly or indirectly (by one or more intermediaries), bound to, or immobilized on, the array.

The use, method and array of the invention may comprise probes that are immobilized via their 5' or 3' end. However, when the capture probe is immobilized directly to the array substrate, it may be immobilized only such that the 3' end of the capture probe is free to be extended, e.g. it is immobilized by its 5' end. The capture probe may be immobilized indirectly, such that it has a free, i.e. extendible, 3' end.

By extended or extendible 3' end, it is meant that further nucleotides may be added to the most 3' nucleotide of the nucleic acid molecule, e.g. capture probe, to extend the length of the nucleic acid molecule, i.e. the standard polymerization

reaction utilized to extend nucleic acid molecules, e.g. templated polymerization catalyzed by a polymerase.

Thus, in one embodiment, the array comprises probes that are immobilized directly via their 3' end, so-called surface probes, which are defined below. Each species of surface probe comprises a region of complementarity to each species of capture probe, such that the capture probe may hybridize to the surface probe, resulting in the capture probe comprising a free extendible 3' end. In a preferred aspect of the invention, when the array comprises surface probes, the capture probes are synthesized *in situ* on the array.

The array probes may be made up of ribonucleotides and/or deoxyribonucleotides as well as synthetic nucleotide residues that are capable of participating in Watson-Crick type or analogous base pair interactions. Thus, the nucleic acid domain may be DNA or RNA or any modification thereof e.g. PNA or other derivatives containing non-nucleotide backbones. However, in the context of transcriptome analysis the capture domain of the capture probe must capable of priming a reverse transcription reaction to generate cDNA that is complementary to the captured RNA molecules. As described below in more detail, in the context of genome analysis, the capture domain of the capture probe must be capable of binding to the DNA fragments, which may comprise binding to a binding domain that has been added to the fragmented DNA. In some embodiments, the capture domain of the capture probe may prime a DNA extension (polymerase) reaction to generate DNA that is complementary to the captured DNA molecules. In other embodiments, the capture domain may template a ligation reaction between the captured DNA molecules and a surface probe that is directly or indirectly immobilised on the substrate. In yet other embodiments, the capture domain may be ligated to one strand of the captured DNA molecules.

In a preferred embodiment of the invention at least the capture domain of the capture probe comprises or consists of deoxyribonucleotides (dNTPs). In a particularly preferred embodiment the whole of the capture probe comprises or consists of deoxyribonucleotides.

In a preferred embodiment of the invention the capture probes are immobilized on the substrate of the array directly, i.e. by their 5' end, resulting in a free extendible 3' end.

The capture probes of the invention comprise at least two domains, a capture domain and a positional domain (or a feature identification tag or domain;

the positional domain may alternatively be defined as an identification (ID) domain or tag, or as a positional tag). The capture probe may further comprise a universal domain as defined further below. Where the capture probe is indirectly attached to the array surface via hybridization to a surface probe, the capture probe requires a sequence (e.g. a portion or domain) which is complementary to the surface probe. Such a complementary sequence may be complementary to a positional/identification domain and/or a universal domain on the surface probe. In other words the positional domain and/or universal domain may constitute the region or portion of the probe which is complementary to the surface probe. However, the capture probe may also comprise an additional domain (or region, portion or sequence) which is complementary to the surface probe. For ease of synthesis, as described in more detail below, such a surface probe-complementary region may be provided as part, or as an extension of the capture domain (such a part or extension not itself being used for, or capable of, binding to the target nucleic acid, e.g. RNA).

The capture domain is typically located at the 3' end of the capture probe and comprises a free 3' end that can be extended, e.g. by template dependent polymerization. The capture domain comprises a nucleotide sequence that is capable of hybridizing to nucleic acid, e.g. RNA (preferably mRNA) present in the cells of the tissue sample contact with the array.

Advantageously, the capture domain may be selected or designed to bind (or put more generally may be capable of binding) selectively or specifically to the particular nucleic acid, e.g. RNA it is desired to detect or analyse. For example the capture domain may be selected or designed for the selective capture of mRNA. As is well known in the art, this may be on the basis of hybridisation to the poly-A tail of mRNA. Thus, in a preferred embodiment the capture domain comprises a poly-T DNA oligonucleotide, i.e. a series of consecutive deoxythymidine residues linked by phosphodiester bonds, which is capable of hybridizing to the poly-A tail of mRNA. Alternatively, the capture domain may comprise nucleotides which are functionally or structurally analogous to poly-T i.e., are capable of binding selectively to poly-A, for example a poly-U oligonucleotide or an oligonucleotide comprised of deoxythymidine analogues, wherein said oligonucleotide retains the functional property of binding to poly-A. In a particularly preferred embodiment the capture domain, or more particularly the poly-T element of the capture domain, comprises at least 10 nucleotides, preferably at least 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20

nucleotides. In a further embodiment, the capture domain, or more particularly the poly-T element of the capture domain comprises at least 25, 30 or 35 nucleotides.

Random sequences may also be used in the capture of nucleic acid, as is known in the art, e.g. random hexamers or similar sequences, and hence such random sequences may be used to form all or a part of the capture domain. For example, random sequences may be used in conjunction with poly-T (or poly-T analogue etc.) sequences. Thus where a capture domain comprises a poly-T(or a "poly-T-like") oligonucleotide, it may also comprise a random oligonucleotide sequence. This may for example be located 5' or 3' of the poly-T sequence, e.g. at the 3' end of the capture probe, but the positioning of such a random sequence is not critical. Such a construct may facilitate the capturing of the initial part of the poly-A of mRNA. Alternatively, the capture domain may be an entirely random sequence. Degenerate capture domains may also be used, according to principles known in the art.

The capture domain may be capable of binding selectively to a desired sub-type or subset of nucleic acid, e.g. RNA, for example a particular type of RNA such mRNA or rRNA etc. as listed above, or to a particular subset of a given type of RNA, for example, a particular mRNA species e.g. corresponding to a particular gene or group of genes. Such a capture probe may be selected or designed based on sequence of the RNA it is desired to capture. Thus it may be a sequence-specific capture probe, specific for a particular RNA target or group of targets (target group etc). Thus, it may be based on a particular gene sequence or particular motif sequence or common/conserved sequence etc., according to principles well known in the art.

In embodiments where the capture probe is immobilized on the substrate of the array indirectly, e.g. via hybridization to a surface probe, the capture domain may further comprise an upstream sequence (5' to the sequence that hybridizes to the nucleic acid, e.g. RNA of the tissue sample) that is capable of hybridizing to 5' end of the surface probe. Alone, the capture domain of the capture probe may be seen as a capture domain oligonucleotide, which may be used in the synthesis of the capture probe in embodiments where the capture probe is immobilized on the array indirectly.

The positional domain (feature identification domain or tag) of the capture probe is located directly or indirectly upstream, i.e. closer to the 5' end of the capture probe nucleic acid molecule, of the capture domain. Preferably the

positional domain is directly adjacent to the capture domain, i.e. there is no intermediate sequence between the capture domain and the positional domain. In some embodiments the positional domain forms the 5' end of the capture probe, which may be immobilized directly or indirectly on the substrate of the array.

As discussed above, each feature (distinct position) of the array comprises a spot of a species of nucleic acid probe, wherein the positional domain at each feature is unique. Thus, a "species" of capture probe is defined with reference to its positional domain; a single species of capture probe will have the same positional domain. However, it is not required that each member of a species of capture probe has the same sequence in its entirety. In particular, since the capture domain may be or may comprise a random or degenerate sequence, the capture domains of individual probes within a species may vary. Accordingly, in some embodiments where the capture domains of the capture probes are the same, each feature comprises a single probe sequence. However in other embodiments where the capture probe varies, members of a species of probe will not have the exact same sequence, although the sequence of the positional domain of each member in the species will be the same. What is required is that each feature or position of the array carries a capture probe of a single species (specifically each feature or position carries a capture probe which has an identical positional tag, i.e. there is a single positional domain at each feature or position). Each species has a different positional domain which identifies the species. However, each member of a species, may in some cases, as described in more detail herein, have a different capture domain, as the capture domain may be random or degenerate or may have a random or degenerate component. This means that within a given feature, or position, the capture domain of the probes may differ.

Thus in some, but not necessarily in all embodiments, the nucleotide sequence of any one probe molecule immobilized at a particular feature is the same as the other probe molecules immobilized at the same feature, but the nucleotide sequence of the probes at each feature is different, distinct or distinguishable from the probes immobilized at every other feature. Preferably each feature comprises a different species of probe. However, in some embodiments it may be advantageous for a group of features to comprise the same species of probe, i.e. effectively to produce a feature covering an area of the array that is greater than a single feature, e.g. to lower the resolution of the array. In other embodiments of the array, the nucleotide sequence of the positional domain of any one probe molecule

immobilized at a particular feature may be the same as the other probe molecules immobilized at the same feature but the capture domain may vary. The capture domain may nonetheless be designed to capture the same type of molecule, e.g. mRNA in general.

The positional domain (or tag) of the capture probe comprises the sequence which is unique to each feature and acts as a positional or spatial marker (the identification tag). In this way each region or domain of the tissue sample, e.g. each cell in the tissue, will be identifiable by spatial resolution across the array linking the nucleic acid, e.g. RNA (e.g. the transcripts) from a certain cell to a unique positional domain sequence in the capture probe. By virtue of the positional domain a capture probe in the array may be correlated to a position in the tissue sample, for example it may be correlated to a cell in the sample. Thus, the positional domain of the capture domain may be seen as a nucleic acid tag (identification tag).

Any suitable sequence may be used as the positional domain in the capture probes of the invention. By a suitable sequence, it is meant that the positional domain should not interfere with (i.e. inhibit or distort) the interaction between the RNA of the tissue sample and the capture domain of the capture probe. For example, the positional domain should be designed such that nucleic acid molecules in the tissue sample do not hybridize specifically to the positional domain. Preferably, the nucleic acid sequence of the positional domain of the capture probes has less than 80% sequence identity to the nucleic acid sequences in the tissue sample. Preferably, the positional domain of the capture probe has less than 70%, 60%, 50% or less than 40% sequence identity across a substantial part of the nucleic acids molecules in the tissue sample. Sequence identity may be determined by any appropriate method known in the art, e.g. the using BLAST alignment algorithm.

In a preferred embodiment the positional domain of each species of capture probe contains a unique barcode sequence. The barcode sequences may be generated using random sequence generation. The randomly generated sequences may be followed by stringent filtering by mapping to the genomes of all common reference species and with pre-set Tm intervals, GC content and a defined distance of difference to the other barcode sequences to ensure that the barcode sequences will not interfere with the capture of the nucleic acid, e.g. RNA from the tissue sample and will be distinguishable from each other without difficulty.

As mentioned above, and in a preferred embodiment, the capture probe comprises also a universal domain (or linker domain or tag). The universal domain of the capture probe is located directly or indirectly upstream, i.e. closer to the 5' end of the capture probe nucleic acid molecule, of the positional domain. Preferably the universal domain is directly adjacent to the positional domain, i.e. there is no intermediate sequence between the positional domain and the universal domain. In embodiments where the capture probe comprises a universal domain, the domain will form the 5' end of the capture probe, which may be immobilized directly or indirectly on the substrate of the array.

The universal domain may be utilized in a number of ways in the methods and uses of the invention. For example, the methods of the invention comprise a step of releasing (e.g. removing) at least part of the synthesised (i.e. extended or ligated) nucleic acid, e.g. cDNA molecules from the surface of the array. As described elsewhere herein, this may be achieved in a number of ways, of which one comprises cleaving the nucleic acid, e.g. cDNA molecule from the surface of the array. Thus, the universal domain may itself comprise a cleavage domain, i.e. a sequence that can be cleaved specifically, either chemically or preferably enzymatically.

Thus, the cleavage domain may comprise a sequence that is recognised by one or more enzymes capable of cleaving a nucleic acid molecule, i.e. capable of breaking the phosphodiester linkage between two or more nucleotides. For instance, the cleavage domain may comprise a restriction endonuclease (restriction enzyme) recognition sequence. Restriction enzymes cut double-stranded or single stranded DNA at specific recognition nucleotide sequences known as restriction sites and suitable enzymes are well known in the art. For example, it is particularly advantageous to use rare-cutting restriction enzymes, i.e. enzymes with a long recognition site (at least 8 base pairs in length), to reduce the possibility of cleaving elsewhere in the nucleic acid, e.g. cDNA molecule. In this respect, it will be seen that removing or releasing at least part of the nucleic acid, e.g. cDNA molecule requires releasing a part comprising the positional domain of the nucleic acid, e.g. cDNA and all of the sequence downstream of the domain, i.e. all of the sequence that is 3' to the positional domain. Hence, cleavage of the nucleic acid, e.g. cDNA molecule should take place 5' to the positional domain.

By way of example, the cleavage domain may comprise a poly-U sequence which may be cleaved by a mixture of Uracil DNA glycosylase (UDG) and the DNA glycosylase-lyase Endonuclease VIII, commercially known as the USER™ enzyme.

A further example of a cleavage domain can be utilised in embodiments where the capture probe is immobilized to the array substrate indirectly, i.e. via a surface probe. The cleavage domain may comprise one or more mismatch nucleotides, i.e. when the complementary parts of the surface probe and the capture probe are not 100% complementary. Such a mismatch is recognised, e.g. by the MutY and T7 endonuclease I enzymes, which results in cleavage of the nucleic acid molecule at the position of the mismatch.

In some embodiments of the invention, the positional domain of the capture probe comprises a cleavage domain, wherein the said cleavage domain is located at the 5' end of the positional domain.

The universal domain may comprise also an amplification domain. This may be in addition to, or instead of, a cleavage domain. In some embodiments of the invention, as described elsewhere herein, it may be advantageous to amplify the nucleic acid, e.g. cDNA molecules, for example after they have been released (e.g. removed or cleaved) from the array substrate. It will be appreciated however, that the initial cycle of amplification, or indeed any or all further cycles of amplification may also take place *in situ* on the array. The amplification domain comprises a distinct sequence to which an amplification primer may hybridize. The amplification domain of the universal domain of the capture probe is preferably identical for each species of capture probe. Hence a single amplification reaction will be sufficient to amplify all of the nucleic acid, e.g. cDNA molecules (which may or may not be released from the array substrate prior to amplification).

Any suitable sequence may be used as the amplification domain in the capture probes of the invention. By a suitable sequence, it is meant that the amplification domain should not interfere with (i.e. inhibit or distort) the interaction between the nucleic acid, e.g. RNA of the tissue sample and the capture domain of the capture probe. Furthermore, the amplification domain should comprise a sequence that is not the same or substantially the same as any sequence in the nucleic acid, e.g. RNA of the tissue sample, such that the primer used in the amplification reaction can hybridized only to the amplification domain under the amplification conditions of the reaction.

For example, the amplification domain should be designed such that nucleic acid molecules in the tissue sample do not hybridize specifically to the amplification domain or the complementary sequence of the amplification domain. Preferably, the nucleic acid sequence of the amplification domain of the capture probes and the complement thereof has less than 80% sequence identity to the nucleic acid sequences in the tissue sample. Preferably, the positional domain of the capture probe has less than 70%, 60%, 50% or less than 40% sequence identity across a substantial part of the nucleic acid molecules in the tissue sample. Sequence identity may be determined by any appropriate method known in the art, e.g. the using BLAST alignment algorithm.

Thus, alone, the universal domain of the capture probe may be seen as a universal domain oligonucleotide, which may be used in the synthesis of the capture probe in embodiments where the capture probe is immobilized on the array indirectly.

In one representative embodiment of the invention only the positional domain of each species of capture probe is unique. Hence, the capture domains and universal domains (if present) are in one embodiment the same for every species of capture probe for any particular array to ensure that the capture of the nucleic acid, e.g. RNA from the tissue sample is uniform across the array. However, as discussed above, in some embodiments the capture domains may differ by virtue of including random or degenerate sequences.

In embodiments where the capture probe is immobilized on the substrate of the array indirectly, e.g. via hybridisation to a surface probe, the capture probe may be synthesised on the array as described below.

The surface probes are immobilized on the substrate of the array directly by or at, e.g. their 3' end. Each species of surface probe is unique to each feature (distinct position) of the array and is partly complementary to the capture probe, defined above.

Hence the surface probe comprises at its 5' end a domain (complementary capture domain) that is complementary to a part of the capture domain that does not bind to the nucleic acid, e.g. RNA of the tissue sample. In other words, it comprises a domain that can hybridize to at least part of a capture domain oligonucleotide. The surface probe further comprises a domain (complementary positional domain or complementary feature identification domain) that is complementary to the positional domain of the capture probe. The complementary

positional domain is located directly or indirectly downstream (i.e. at the 3' end) of the complementary capture domain, i.e. there may be an intermediary or linker sequence separating the complementary positional domain and the complementary capture domain. In embodiments where the capture probe is synthesized on the array surface, the surface probes of the array always comprise a domain (complementary universal domain) at the 3' end of the surface probe, i.e. directly or indirectly downstream of the positional domain, which is complementary to the universal domain of the capture probe. In other words, it comprises a domain that can hybridize to at least part of the universal domain oligonucleotide.

In some embodiments of the invention the sequence of the surface probe shows 100% complementarity or sequence identity to the positional and universal domains and to the part of the capture domain that does not bind to the nucleic acid, e.g. RNA of the tissue sample. In other embodiments the sequence of the surface probe may show less than 100% sequence identity to the domains of the capture probe, e.g. less than 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91% or 90%. In a particularly preferred embodiment of the invention, the complementary universal domain shares less than 100% sequence identity to the universal domain of the capture probe.

In one embodiment of the invention, the capture probe is synthesized or generated on the substrate of the array. In a representative embodiment (see figure 3), the array comprises surface probes as defined above. Oligonucleotides that correspond to the capture domain and universal domain of the capture probe are contacted with the array and allowed to hybridize to the complementary domains of the surface probes. Excess oligonucleotides may be removed by washing the array under standard hybridization conditions. The resultant array comprises partially single stranded probes, wherein both the 5' and 3' ends of the surface probe are double stranded and the complementary positional domain is single stranded. The array may be treated with a polymerase enzyme to extend the 3' end of the universal domain oligonucleotide, in a template dependent manner, so as to synthesize the positional domain of the capture probe. The 3' end of the synthesized positional domain is then ligated, e.g. using a ligase enzyme, to the 5' end of the capture domain oligonucleotide to generate the capture probe. It will be understood in this regard that the 5' end of the capture domain oligonucleotide is phosphorylated to enable ligation to take place. As each species of surface probe

comprises a unique complementary positional domain, each species of capture probe will comprise a unique positional domain.

The term "hybridisation" or "hybridises" as used herein refers to the formation of a duplex between nucleotide sequences which are sufficiently complementary to form duplexes via Watson-Crick base pairing. Two nucleotide sequences are "complementary" to one another when those molecules share base pair organization homology. "Complementary" nucleotide sequences will combine with specificity to form a stable duplex under appropriate hybridization conditions. For instance, two sequences are complementary when a section of a first sequence can bind to a section of a second sequence in an anti-parallel sense wherein the 3'-end of each sequence binds to the 5'-end of the other sequence and each A, T(U), G and C of one sequence is then aligned with a T(U), A, C and G, respectively, of the other sequence. RNA sequences can also include complementary G=U or U=G base pairs. Thus, two sequences need not have perfect homology to be "complementary" under the invention. Usually two sequences are sufficiently complementary when at least about 90% (preferably at least about 95%) of the nucleotides share base pair organization over a defined length of the molecule. The domains of the capture and surface probes thus contain a region of complementarity. Furthermore the capture domain of the capture probe contains a region of complementarity for the nucleic acid, e.g. RNA (preferably mRNA) of the tissue sample.

The capture probe may also be synthesised on the array substrate using polymerase extension (similarly to as described above) and a terminal transferase enzyme to add a "tail" which may constitute the capture domain. This is described further in Example 7 below. The use of terminal transferases to add nucleotide sequences to the end of an oligonucleotide is known in the art, e.g. to introduce a homopolymeric tail e.g. a poly-T tail. Accordingly, in such a synthesis an oligonucleotide that corresponds to the universal domain of the capture probe may be contacted with the array and allowed to hybridize to the complementary domain of the surface probes. Excess oligonucleotides may be removed by washing the array under standard hybridization conditions. The resultant array comprises partially single stranded probes, wherein the 5' ends of the surface probes are double stranded and the complementary positional domain is single stranded. The array may be treated with a polymerase enzyme to extend the 3' end of the universal domain oligonucleotide, in a template dependent manner, so as to

synthesize the positional domain of the capture probe. The capture domain, e.g. comprising a poly-T sequence may then be introduced using a terminal transferase to add a poly-T tail to generate the capture probe.

The typical array of, and for use in the methods of, the invention may contain multiple spots, or "features". A feature may be defined as an area or distinct position on the array substrate at which a single species of capture probe is immobilized. Hence each feature will comprise a multiplicity of probe molecules, of the same species. It will be understood in this context that whilst it is encompassed that each capture probe of the same species may have the same sequence, this need not necessarily be the case. Each species of capture probe will have the same positional domain (i.e. each member of a species and hence each probe in a feature will be identically "tagged"), but the sequence of each member of the feature (species) may differ, because the sequence of a capture domain may differ. As described above, random or degenerate capture domains may be used. Thus the capture probes within a feature may comprise different random or degenerate sequences. The number and density of the features on the array will determine the resolution of the array, i.e. the level of detail at which the transcriptome or genome of the tissue sample can be analysed. Hence, a higher density of features will typically increase the resolution of the array.

As discussed above, the size and number of the features on the array of the invention will depend on the nature of the tissue sample and required resolution. Thus, if it is desirable to determine a transcriptome or genome only for regions of cells within a tissue sample (or the sample contains large cells) then the number and/or density of features on the array may be reduced (i.e. lower than the possible maximum number of features) and/or the size of the features may be increased (i.e. the area of each feature may be greater than the smallest possible feature), e.g. an array comprising few large features. Alternatively, if it is desirable to determine a transcriptome or genome of individual cells within a sample, it may be necessary to use the maximum number of features possible, which would necessitate using the smallest possible feature size, e.g. an array comprising many small features.

Whilst single cell resolution may be a preferred and advantageous feature of the present invention, it is not essential to achieve this, and resolution at the cell group level is also of interest, for example to detect or distinguish a particular cell type or tissue region, e.g. normal vs tumour cells.

In representative embodiments of the invention, an array may contain at least 2, 5, 10, 50, 100, 500, 750, 1000, 1500, 3000, 5000, 10000, 20000, 40000, 50000, 75000, 100000, 150000, 200000, 300000, 400000, 500000, 750000, 800000, 1000000, 1200000, 1500000, 1750000, 2000000, 2100000. 3000000, 3500000, 4000000 or 4200000 features. Whilst 4200000 represents the maximum number of features presently available on a commercial array, it is envisaged that arrays with features in excess of this may be prepared and such arrays are of interest in the present invention. As noted above, feature size may be decreased and this may allow greater numbers of features to be accommodated within the same or a similar area. By way of example. these features may be comprised in an area of less than about $20cm^2$, $10cm^2$, $5cm^2$, $1cm^2$, $1mm^2$, or $100\mu m^2$.

Thus, in some embodiments of the invention the area of each feature may be from about 1 μm^2, 2 μm^2, 3 μm^2, 4 μm^2, 5 μm^2, 10 μm^2, 12 μm^2, 15 μm^2, 20 μm^2, 50 μm^2, 75 μm^2, 100 μm^2, 150 μm^2, 200 μm^2, 250 μm^2, 300 μm^2, 400 μm^2, or 500 μm^2.

It will be evident that a tissue sample from any organism could be used in the methods of the invention, e.g. plant, animal or fungal. The array of the invention allows the capture of any nucleic acid, e.g. mRNA molecules, which are present in cells that are capable of transcription and/or translation. The arrays and methods of the invention are particularly suitable for isolating and analysing the transcriptome or genome of cells within a sample, wherein spatial resolution of the transcriptomes or genomes is desirable, e.g. where the cells are interconnected or in contact directly with adjacent cells. However, it will be apparent to a person of skill in the art that the methods of the invention may also be useful for the analysis of the transcriptome or genome of different cells or cell types within a sample even if said cells do not interact directly, e.g. a blood sample. In other words, the cells do not need to present in the context of a tissue and can be applied to the array as single cells (e.g. cells isolated from a non-fixed tissue). Such single cells, whilst not necessarily fixed to a certain position in a tissue, are nonetheless applied to a certain position on the array and can be individually identified. Thus, in the context of analysing cells that do not interact directly, or are not present in a tissue context, the spatial properties of the described methods may be applied to obtaining or retrieving unique or independent transcriptome or genome information from individual cells.

The sample may thus be a harvested or biopsied tissue sample, or possibly a cultured sample. Representative samples include clinical samples e.g. whole blood or blood-derived products, blood cells, tissues, biopsies, or cultured tissues or cells etc. including cell suspensions. Artificial tissues may for example be prepared from cell suspension (including for example blood cells). Cells may be captured in a matrix (for example a gel matrix e.g. agar, agarose, etc) and may then be sectioned in a conventional way. Such procedures are known in the art in the context of immunohistochemistry (see e.g. Andersson *et al* 2006, J. Histochem. Cytochem. 54(12): 1413-23. Epub 2006 Sep 6).

The mode of tissue preparation and how the resulting sample is handled may effect the transcriptomic or genomic analysis of the methods of the invention. Moreover, various tissue samples will have different physical characteristics and it is well within the skill of a person in the art to perform the necessary manipulations to yield a tissue sample for use with the methods of the invention. However, it is evident from the disclosures herein that any method of sample preparation may be used to obtain a tissue sample that is suitable for use in the methods of the invention. For instance any layer of cells with a thickness of approximately 1 cell or less may be used in the methods of the invention. In one embodiment, the thickness of the tissue sample may be less than 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 or 0.1 of the cross-section of a cell. However, since as noted above, the present invention is not limited to single cell resolution and hence it is not a requirement that the tissue sample has a thickness of one cell diameter or less; thicker tissue samples may if desired be used. For example cryostat sections may be used, which may be e.g. 10-20 μm thick.

The tissue sample may be prepared in any convenient or desired way and the invention is not restricted to any particular type of tissue preparation. Fresh, frozen, fixed or unfixed tissues may be used. Any desired convenient procedure may be used for fixing or embedding the tissue sample, as described and known in the art. Thus any known fixatives or embedding materials may be used.

As a first representative example of a tissue sample for use in the invention, the tissue may prepared by deep freezing at temperature suitable to maintain or preserve the integrity (i.e. the physical characteristics) of the tissue structure, e.g. less than -20°C and preferably less than -25, -30, -40, -50, -60, -70 or -80 °C. The frozen tissue sample may be sectioned, i.e. thinly sliced, onto the array surface by any suitable means. For example, the tissue sample may be prepared using a

chilled microtome, a cryostat, set at a temperature suitable to maintain both the structural integrity of the tissue sample and the chemical properties of the nucleic acids in the sample, e.g. to less than -15°C and preferably less than -20 or -25°C. Thus, the sample should be treated so as to minimize the degeneration or degradation of the nucleic acid, e.g. RNA in the tissue. Such conditions are well-established in the art and the extent of any degradation may be monitored through nucleic acid extraction, e.g. total RNA extraction and subsequent quality analysis at various stages of the preparation of the tissue sample.

In a second representative example, the tissue may be prepared using standard methods of formalin-fixation and paraffin-embedding (FFPE), which are well-established in the art. Following fixation of the tissue sample and embedding in a paraffin or resin block, the tissue samples may sectioned, i.e. thinly sliced, onto the array. As noted above, other fixatives and/or embedding materials can be used.

It will be apparent that the tissue sample section will need to be treated to remove the embedding material e.g. to deparaffinize, i.e. to remove the paraffin or resin, from the sample prior to carrying out the methods of the invention. This may be achieved by any suitable method and the removal of paraffin or resin or other material from tissue samples is well established in the art, e.g. by incubating the sample (on the surface of the array) in an appropriate solvent e.g. xylene, e.g. twice for 10 minutes, followed by an ethanol rinse, e.g. 99.5% ethanol for 2 minutes, 96% ethanol for 2 minutes, and 70% ethanol for 2 minutes.

It will be evident to the skilled person that the RNA in tissue sections prepared using methods of FFPE or other methods of fixing and embedding is more likely to be partially degraded than in the case of frozen tissue. However, without wishing to be bound by any particular theory, it is believed that this may be advantageous in the methods of the invention. For instance, if the RNA in the sample is partially degraded the average length of the RNA polynucleotides will be less and more randomized than a non-degraded sample. It is postulated therefore that partially degraded RNA would result in less bias in the various processing steps, described elsewhere herein, e.g. ligation of adaptors (amplification domains), amplification of the cDNA molecules and sequencing thereof.

Hence, in one embodiment of the invention the tissue sample, i.e. the section of the tissue sample contacted with the array, is prepared using FFPE or other methods of fixing and embedding. In other words the sample may be fixed, e.g. fixed and embedded. In an alternative embodiment of the invention the tissue

sample is prepared by deep-freezing. In another embodiment a touch imprint of a tissue may be used, according to procedures known in the art. In other embodiments an unfixed sample may be used.

The thickness of the tissue sample section for use in the methods of the invention may be dependent on the method used to prepare the sample and the physical characteristics of the tissue. Thus, any suitable section thickness may be used in the methods of the invention. In representative embodiments of the invention the thickness of the tissue sample section will be at least 0.1μm, further preferably at least 0.2, 0.3, 0.4, 0.5, 0.7, 1.0, 1.5, 2, 3, 4, 5, 6, 7, 8, 9 or 10μm. In other embodiments the thickness of the tissue sample section is at least 10, 12, 13 , 14, 15, 20, 30, 40 or 50μm. However, the thickness is not critical and these are representative values only. Thicker samples may be used if desired or convenient e.g. 70 or 100 μm or more. Typically, the thickness of the tissue sample section is between 1-100 μm, 1-50 μm , 1-30 μm, 1-25 μm, 1-20 μm, 1-15 μm, 1-10 μm, 2-8μm, 3-7μm or 4-6μm, but as mentioned above thicker samples may be used.

On contact of the tissue sample section with the array, e.g. following removal of the embedding material e.g. deparrafinization, the nucleic acid, e.g. RNA molecules in the tissue sample will bind to the immobilized capture probes on the array. In some embodiments it may be advantageous to facilitate the hybridization of the nucleic acid, e.g. RNA molecules to the capture probes. Typically, facilitating the hybridization comprises modifying the conditions under which hybridization occurs. The primary conditions that can be modified are the time and temperature of the incubation of the tissue section on the array prior to the reverse transcription step, which is described elsewhere herein.

For instance, on contacting the tissue sample section with the array, the array may be incubated for at least 1 hour to allow the nucleic acid, e.g. RNA to hybridize to the capture probes. Preferably the array may be incubated for at least 2, 3, 5, 10, 12, 15, 20, 22 or 24 hours or until the tissue sample section has dried. The array incubation time is not critical and any convenient or desired time may be used. Typical array incubations may be up to 72 hours. Thus, the incubation may occur at any suitable temperature, for instance at room temperature, although in a preferred embodiment the tissue sample section is incubated on the array at a temperature of at least 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36 or 37°C. Incubation temperatures of up to 55°C are commonplace in the art. In a particularly preferred embodiment the tissue sample section is allowed to dry on the array at

37°C for 24 hours. Once the tissue sample section has dried the array may be stored at room temperature before performing the reverse transcription step. It will be understood that the if the tissue sample section is allowed to dry on the surface of the array, it will need to be rehydrated before further manipulation of the captured nucleic acid can be achieved, e.g. the step of reverse transcribing the captured RNA.

Hence, the method of the invention may comprise a further step of rehydrating the tissue sample after contacting the sample with the array.

In some embodiments it may be advantageous to block (e.g. mask or modify) the capture probes prior to contacting the tissue sample with the array, particularly when the nucleic acid in the tissue sample is subject to a process of modification prior to its capture on the array. Specifically, it may be advantageous to block or modify the free 3' end of the capture probe. In a particular embodiment, the nucleic acid in the tissue sample, e.g. fragmented genomic DNA, may be modified such that it can be captured by the capture probe. For instance, and as described in more detail below, an adaptor sequence (comprising a binding domain capable of binding to the capture domain of the capture probe) may be added to the end of the nucleic acid, e.g. fragmented genomic DNA. This may be achieved by, e.g. ligation of an adaptor or extension of the nucleic acid, e.g. using an enzyme to incorporate additional nucleotides at the end of the sequence, e.g. a poly-A tail. It is necessary to block or modify the capture probes, particularly the free 3' end of the capture probe, prior to contacting the tissue sample with the array to avoid modification of the capture probes, e.g. to avoid the addition of a poly-A tail to the free 3' end of the capture probes. Preferably the incorporation of a blocking domain may be incorporated into the capture probe when it is synthesised. However, the blocking domain may be incorporated to the capture probe after its synthesis.

In some embodiments the capture probes may be blocked by any suitable and reversible means that would prevent modification of the capture domains during the process of modifying the nucleic acid of the tissue sample, which occurs after the tissue sample has been contacted with the array. In other words, the capture probes may be reversibly masked or modified such that the capture domain of the capture probe does not comprise a free 3' end, i.e. such that the 3' end is removed or modified, or made inaccessible so that the capture probe is not susceptible to the process which is used to modify the nucleic acid of the tissue

sample, e.g. ligation or extension, or the additional nucleotides may be removed to reveal and/or restore the 3' end of the capture domain of the capture probe.

For example, blocking probes may be hybridised to the capture probes to mask the free 3' end of the capture domain, e.g. hairpin probes or partially double stranded probes, suitable examples of which are known in the art. The free 3' end of the capture domain may be blocked by chemical modification, e.g. addition of an azidomethyl group as a chemically reversible capping moiety such that the capture probes do not comprise a free 3' end. Suitable alternative capping moieties are well known in the art, e.g. the terminal nucleotide of the capture domain could be a reversible terminator nucleotide, which could be included in the capture probe during or after probe synthesis.

Alternatively or additionally, the capture domain of the capture probe could be modified so as to allow the removal of any modifications of the capture probe, e.g. additional nucleotides, that occur when the nucleic acid molecules of the tissue sample are modified. For instance, the capture probes may comprise an additional sequence downstream of the capture domain, i.e. 3' to capture domain, namely a blocking domain. This could be in the form of, e.g. a restriction endonuclease recognition sequence or a sequence of nucleotides cleavable by specific enzyme activities, e.g. uracil. Following the modification of the nucleic acid of the tissue sample, the capture probes could be subjected to an enzymatic cleavage, which would allow the removal of the blocking domain and any of the additional nucleotides that are added to the 3' end of the capture probe during the modification process. The removal of the blocking domain would reveal and/or restore the free 3' end of the capture domain of the capture probe. The blocking domain could be synthesised as part of the capture probe or could be added to the capture probe *in situ* (i.e. as a modification of an existing array), e.g. by ligation of the blocking domain.

The capture probes may be blocked using any combination of the blocking mechanisms described above.

Once the nucleic acid of the tissue sample, e.g. fragmented genomic DNA, has been modified to enable it to hybridise to the capture domain of the capture probe, the capture probe must be unblocked, e.g. by dissociation of the blocking oligonucleotide, removal of the capping moiety and/or blocking domain.

In order to correlate the sequence analysis or transcriptome or genome information obtained from each feature of the array with the region (i.e. an area or

cell) of the tissue sample the tissue sample is oriented in relation to the features on the array. In other words, the tissue sample is placed on the array such that the position of a capture probe on the array may be correlated with a position in the tissue sample. Thus it may be identified where in the tissue sample the position of each species of capture probe (or each feature of the array) corresponds. In other words, it may be identified to which location in the tissue sample the position of each species of capture probe corresponds. This may be done by virtue of positional markers present on the array, as described below. Conveniently, but not necessarily, the tissue sample may be imaged following its contact with the array. This may be performed before or after the nucleic acid of the tissue sample is processed, e.g. before or after the cDNA generation step of the method, in particular the step of generating the first strand cDNA by reverse transcription. In a preferred embodiment the tissue sample is imaged prior to the release of the captured and synthesised (i.e. extended or ligated) DNA, e.g. cDNA, from the array. In a particularly preferred embodiment the tissue is imaged after the nucleic acid of the tissue sample has been processed, e.g. after the reverse transcription step, and any residual tissue is removed (e.g. washed) from the array prior to the release of molecules, e.g. of the cDNA from the array. In some embodiments, the step of processing the captured nucleic acid, e.g. the reverse transcription step, may act to remove residual tissue from the array surface, e.g. when using tissue preparing by deep-freezing. In such a case, imaging of the tissue sample may take place prior to the processing step, e.g. the cDNA synthesis step. Generally speaking, imaging may take place at any time after contacting the tissue sample with the area, but before any step which degrades or removes the tissue sample. As noted above, this may depend on the tissue sample.

Advantageously, the array may comprise markers to facilitate the orientation of the tissue sample or the image thereof in relation to the features of the array. Any suitable means for marking the array may be used such that they are detectable when the tissue sample is imaged. For instance, a molecule, e.g. a fluorescent molecule, that generates a signal, preferably a visible signal, may be immobilized directly or indirectly on the surface of the array. Preferably, the array comprises at least two markers in distinct positions on the surface of the array, further preferably at least 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100 markers. Conveniently several hundred or even several thousand markers may be used. The markers may be provided in a pattern, for example make up an outer edge of the

array, e.g. an entire outer row of the features of an array. Other informative patterns may be used, e.g. lines sectioning the array. This may facilitate aligning an image of the tissue sample to an array, or indeed generally in correlating the features of the array to the tissue sample. Thus, the marker may be an immobilized molecule to which a signal giving molecule may interact to generate a signal. In a representative example, the array may comprise a marker feature, e.g. a nucleic acid probe immobilized on the substrate of array, to which a labelled nucleic acid may hybridize. For instance, the labelled nucleic acid molecule, or marker nucleic acid, may be linked or coupled to a chemical moiety capable of fluorescing when subjected to light of a specific wavelength (or range of wavelengths), i.e. excited. Such a marker nucleic acid molecule may be contacted with the array before, contemporaneously with or after the tissue sample is stained in order to visualize or image the tissue sample. However, the marker must be detectable when the tissue sample is imaged. Thus, in a preferred embodiment the marker may be detected using the same imaging conditions used to visualize the tissue sample.

In a particularly preferred embodiment of the invention, the array comprises marker features to which a labelled, preferably fluorescently labelled, marker nucleic acid molecule, e.g. oligonucleotide, is hybridized.

The step of imaging the tissue may use any convenient histological means known in the art, e.g. light, bright field, dark field, phase contrast, fluorescence, reflection, interference, confocal microscopy or a combination thereof. Typically the tissue sample is stained prior to visualization to provide contrast between the different regions, e.g. cells, of the tissue sample. The type of stain used will be dependent on the type of tissue and the region of the cells to be stained. Such staining protocols are known in the art. In some embodiments more than one stain may be used to visualize (image) different aspects of the tissue sample, e.g. different regions of the tissue sample, specific cell structures (e.g. organelles) or different cell types. In other embodiments, the tissue sample may be visualized or imaged without staining the sample, e.g. if the tissue sample contains already pigments that provide sufficient contrast or if particular forms of microscopy are used.

In a preferred embodiment, the tissue sample is visualized or imaged using fluorescence microscopy.

The tissue sample, i.e. any residual tissue that remains in contact with the array substrate following the reverse transcription step and optionally imaging, if

imaging is desired and was not carried out before reverse transcription, preferably is removed prior to the step of releasing the cDNA molecules from the array. Thus, the methods of the invention may comprise a step of washing the array. Removal of the residual tissue sample may be performed using any suitable means and will be dependent on the tissue sample. In the simplest embodiment, the array may be washed with water. The water may contain various additives, e.g. surfactants (e.g. detergents), enzymes etc to facilitate to removal of the tissue. In some embodiments, the array is washed with a solution comprising a proteinase enzyme (and suitable buffer) e.g. proteinase K. In other embodiments, the solution may comprise also or alternatively cellulase, hemicelluase or chitinase enzymes, e.g. if the tissue sample is from a plant or fungal source. In further embodiments, the temperature of the solution used to wash the array may be, e.g. at least 30°C, preferably at least 35, 40, 45, 50 or 55°C. It will be evident that the wash solution should minimize the disruption of the immobilized nucleic acid molecules. For instance, in some embodiments the nucleic acid molecules may be immobilized on the substrate of the array indirectly, e.g. via hybridization of the capture probe and the RNA and/or the capture probe and the surface probe, thus the wash step should not interfere with the interaction between the molecules immobilized on the array, i.e. should not cause the nucleic acid molecules to be denatured.

Following the step of contacting the array with a tissue sample, under conditions sufficient to allow hybridization to occur between the nucleic acid, e.g. RNA (preferably mRNA), of the tissue sample to the capture probes, the step of securing (acquiring) the hybridized nucleic acid takes place. Securing or acquiring the captured nucleic acid involves a covalent attachment of a complementary strand of the hybridized nucleic acid to the capture probe (i.e. via a nucleotide bond, a phosphodiester bond between juxtaposed 3'-hydroxyl and 5'-phosphate termini of two immediately adjacent nucleotides), thereby tagging or marking the captured nucleic acid with the positional domain specific to the feature on which the nucleic acid is captured.

In some embodiments, securing the hybridized nucleic acid, e.g. a single stranded nucleic acid, may involve extending the capture probe to produce a copy of the captured nucleic acid, e.g. generating cDNA from the captured (hybridized) RNA. It will be understood that this refers to the synthesis of a complementary strand of the hybridized nucleic acid, e.g. generating cDNA based on the captured RNA template (the RNA hybridized to the capture domain of the capture probe).

Thus, in an initial step of extending the capture probe, e.g. the cDNA generation, the captured (hybridized) nucleic acid, e.g. RNA acts as a template for the extension, e.g. reverse transcription, step. In other embodiments, as described below, securing the hybridized nucleic acid, e.g. partially double stranded DNA, may involve covalently coupling the hybridized nucleic acid, e.g. fragmented DNA, to the capture probe, e.g. ligating to the capture probe the complementary strand of the nucleic acid hybridized to the capture probe, in a ligation reaction.

Reverse transcription concerns the step of synthesizing cDNA (complementary or copy DNA) from RNA, preferably mRNA (messenger RNA), by reverse transcriptase. Thus cDNA can be considered to be a copy of the RNA present in a cell at the time at which the tissue sample was taken, i.e. it represents all or some of the genes that were expressed in said cell at the time of isolation.

The capture probe, specifically the capture domain of the capture probe, acts as a primer for producing the complementary strand of the nucleic acid hybridized to the capture probe, e.g. a primer for reverse transcription. Hence, the nucleic acid, e.g. cDNA, molecules generated by the extension reaction, e.g. reverse transcription reaction, incorporate the sequence of the capture probe, i.e. the extension reaction, e.g. reverse transcription reaction, may be seen as a way of labelling indirectly the nucleic acid, e.g. transcripts, of the tissue sample that are in contact with each feature of the array. As mentioned above, each species of capture probe comprises a positional domain (feature identification tag) that represents a unique sequence for each feature of the array. Thus, all of the nucleic acid, e.g. cDNA, molecules synthesized at a specific feature will comprise the same nucleic acid "tag".

The nucleic acid, e.g. cDNA, molecules synthesized at each feature of the array may represent the genome of, or genes expressed from, the region or area of the tissue sample in contact with that feature, e.g. a tissue or cell type or group or sub-group thereof, and may further represent genes expressed under specific conditions, e.g. at a particular time, in a specific environment, at a stage of development or in response to stimulus etc. Hence, the cDNA at any single feature may represent the genes expressed in a single cell, or if the feature is in contact with the sample at a cell junction, the cDNA may represent the genes expressed in more than one cell. Similarly, if a single cell is in contact with multiple features, then each feature may represent a proportion of the genes expressed in said cell. Similarly, in embodiments in which the captured nucleic acid is DNA, any single

feature may be representative of the genome of a single cell or more than one cell. Alternatively, the genome of a single cell may be represented by multiple features.

The step of extending the capture probe, e.g. reverse transcription, may be performed using any suitable enzymes and protocol of which many exist in the art, as described in detail below. However, it will be evident that it is not necessary to provide a primer for the synthesis of the first nucleic acid, e.g. cDNA, strand because the capture domain of the capture probe acts as the primer, e.g. reverse transcription primer.

Preferably, in the context of the present invention the secured nucleic acid (i.e. the nucleic acid covalently attached to the capture probe), e.g. cDNA is treated to comprise double stranded DNA. However, in some embodiments, the captured DNA may already comprise double stranded DNA, e.g. where partially double stranded fragmented DNA is ligated to the capture probe. Treatment of the captured nucleic acid to produce double stranded DNA may be achieved in a single reaction to generate only a second DNA, e.g. cDNA, strand, i.e. to produce double stranded DNA molecules without increasing the number of double stranded DNA molecules, or in an amplification reaction to generate multiple copies of the second strand, which may be in the form of single stranded DNA (e.g. linear amplification) or double stranded DNA, e.g. cDNA (e.g. exponential amplification).

The step of second strand DNA, e.g. cDNA, synthesis may take place *in situ* on the array, either as a discrete step of second strand synthesis, for example using random primers as described in more detail below, or in the initial step of an amplification reaction. Alternatively, the first strand DNA, e.g. cDNA (the strand comprising, i.e. incorporating, the capture probe) may be released from the array and second strand synthesis, whether as a discrete step or in an amplification reaction may occur subsequently, e.g. in a reaction carried out in solution.

Where second strand synthesis takes place on the array (i.e. *in situ*) the method may include an optional step of removing the captured nucleic acid, e.g. RNA before the second strand synthesis, for example using an RNA digesting enzyme (RNase) e.g. RNase H. Procedures for this are well known and described in the art. However, this is generally not necessary, and in most cases the RNA degrades naturally. Removal of the tissue sample from the array will generally remove the RNA from the array. RNase H can be used if desired to increase the robustness of RNA removal.

For instance, in tissue samples that comprise large amounts of RNA, the step of generating the double stranded cDNA may yield a sufficient amount of cDNA that it may be sequenced directly (following release from the array). In this case, second strand cDNA synthesis may be achieved by any means known in the art and as described below. The second strand synthesis reaction may be performed on the array directly, i.e. whilst the cDNA is immobilized on the array, or preferably after the cDNA has been released from the array substrate, as described below.

In other embodiments it will be necessary to enhance, i.e. amplify, the amount of secured nucleic acid, e.g. synthesized cDNA to yield quantities that are sufficient for DNA sequencing. In this embodiment, the first strand of the secured nucleic acid, e.g. cDNA molecules, which comprise also the capture probe of the features of the array, acts as a template for the amplification reaction, e.g. a polymerase chain reaction. The first reaction product of the amplification will be a second strand of DNA, e.g. cDNA, which itself will act as a template for further cycles of the amplification reaction.

In either of the above described embodiments, the second strand of DNA, e.g. cDNA, will comprise a complement of the capture probe. If the capture probe comprises a universal domain, and particularly an amplification domain within the universal domain, then this may be used for the subsequent amplification of the DNA, e.g. cDNA, e.g. the amplification reaction may comprise a primer with the same sequence as the amplification domain, i.e. a primer that is complementary (i.e. hybridizes) to the complement of the amplification domain. In view of the fact that the amplification domain is upstream of the positional domain of the capture probe (in the secured nucleic acid, e.g. the first cDNA strand), the complement of the positional domain will be incorporated in the second strand of the DNA, e.g. cDNA molecules.

In embodiments where the second strand of DNA, e.g. cDNA is generated in a single reaction, the second strand synthesis may be achieved by any suitable means. For instance, the first strand cDNA, preferably, but not necessarily, released from the array substrate, may be incubated with random primers, e.g. hexamer primers, and a DNA polymerase, preferably a strand displacement polymerase, e.g. klenow (exo), under conditions sufficient for templated DNA synthesis to occur. This process will yield double stranded cDNA molecules of varying lengths and is unlikely to yield full-length cDNA molecules, i.e. cDNA

molecules that correspond to entire mRNA from which they were synthesized. The random primers will hybridise to the first strand cDNA molecules at a random position, i.e. within the sequence rather than at the end of the sequence.

If it is desirable to generate full-length DNA, e.g. cDNA, molecules, i.e. molecules that correspond to the whole of the captured nucleic acid, e.g. RNA molecule (if the nucleic acid, e.g. RNA, was partially degraded in the tissue sample then the captured nucleic acid, e.g. RNA, molecules will not be "full-length" transcripts or the same length as the initial fragments of genomic DNA), then the 3' end of the secured nucleic acid, e.g. first stand cDNA, molecules may be modified. For example, a linker or adaptor may be ligated to the 3' end of the cDNA molecules. This may be achieved using single stranded ligation enzymes such as T4 RNA ligase or Circligase™ (Epicentre Biotechnologies).

Alternatively, a helper probe (a partially double stranded DNA molecule capable of hybridising to the 3' end of the first strand cDNA molecule), may be ligated to the 3' end of the secured nucleic acid, e.g. first strand cDNA, molecule using a double stranded ligation enzyme such as T4 DNA ligase. Other enzymes appropriate for the ligation step are known in the art and include, e.g. Tth DNA ligase, Taq DNA ligase, *Thermococcus* sp. (strain 9°N) DNA ligase (9°N™ DNA ligase, New England Biolabs), and Ampligase™ (Epicentre Biotechnologies). The helper probe comprises also a specific sequence from which the second strand DNA, e.g. cDNA, synthesis may be primed using a primer that is complementary to the part of the helper probe that is ligated to the secured nucleic acid, e.g. first cDNA strand. A further alternative comprises the use of a terminal transferase active enzyme to incorporate a polynucleotide tail, e.g. a poly-A tail, at the 3' end of the secured nucleic acid, e.g. first strand of cDNA, molecules. The second strand synthesis may be primed using a poly-T primer, which may also comprise a specific amplification domain for further amplification. Other methods for generating "full-length" double stranded DNA, e.g. cDNA, molecules (or maximal length second strand synthesis) are well-established in the art.

In some embodiments, second strand synthesis may use a method of template switching, e.g. using the SMART™ technology from Clontech®. SMART (Switching Mechanism at 5' End of RNA Template) technology is well established in the art and is based that the discovery that reverse transcriptase enzymes, e.g. Superscript® II (Invitrogen), are capable of adding a few nucleotides at the 3' end of an extended cDNA molecule, i.e. to produce a DNA/RNA hybrid with a single

stranded DNA overhang at the 3' end. The DNA overhang may provide a target sequence to which an oligonucleotide probe can hybridise to provide an additional template for further extension of the cDNA molecule. Advantageously, the oligonucleotide probe that hybridises to the cDNA overhang contains an amplification domain sequence, the complement of which is incorporated into the synthesised first strand cDNA product. Primers containing the amplification domain sequence, which will hybridise to the complementary amplification domain sequence incorporated into the cDNA first strand, can be added to the reaction mix to prime second strand synthesis using a suitable polymerase enzyme and the cDNA first strand as a template. This method avoids the need to ligate adaptors to the 3' end of the cDNA first strand. Whilst template switching was originally developed for full-length mRNAs, which have a 5' cap structure, it has since been demonstrated to work equally well with truncated mRNAs without the cap structure. Thus, template switching may be used in the methods of the invention to generate full length and/or partial or truncated cDNA molecules. Thus, in a preferred embodiment of the invention, the second strand synthesis may utilise, or be achieved by, template switching. In a particularly preferred embodiment, the template switching reaction, i.e. the further extension of the cDNA first strand to incorporate the complementary amplification domain, is performed *in situ* (whilst the capture probe is still attached, directly or indirectly, to the array). Preferably, the second strand synthesis reaction is also performed *in situ*.

In embodiments where it may be necessary or advantageous to enhance, enrich or amplify the DNA, e.g. cDNA molecules, amplification domains may be incorporated in the DNA, e.g. cDNA molecules. As discussed above, a first amplification domain may be incorporated into the secured nucleic acid molecules, e.g. the first strand of the cDNA molecules, when the capture probe comprises a universal domain comprising an amplification domain. In these embodiments, the second strand synthesis may incorporate a second amplification domain. For example, the primers used to generate the second strand cDNA, e.g. random hexamer primers, poly-T primer, the primer that is complementary to the helper probe, may comprise at their 5' end an amplification domain, i.e. a nucleotide sequence to which an amplification primer may hybridize. Thus, the resultant double stranded DNA may comprise an amplification domain at or towards each 5' end of the double stranded DNA, e.g. cDNA molecules. These amplification domains may be used as targets for primers used in an amplification reaction, e.g.

PCR. Alternatively, the linker or adaptor which is ligated to the 3' end of the secured nucleic acid molecules, e.g. first strand cDNA molecules, may comprise a second universal domain comprising a second amplification domain. Similarly, a second amplification domain may be incorporated into the first strand cDNA molecules by template switching.

In embodiments where the capture probe does not comprise a universal domain, particularly comprising an amplification domain, the second strand of the cDNA molecules may be synthesised in accordance with the above description. The resultant double stranded DNA molecules may be modified to incorporate an amplification domain at the 5' end of the first DNA, e.g. cDNA strand (a first amplification domain) and, if not incorporated in the second strand DNA, e.g. cDNA synthesis step, at the 5' end of the second DNA, e.g. cDNA strand (a second amplification domain). Such amplification domains may be incorporated, e.g. by ligating double stranded adaptors to the ends of the DNA, e.g. cDNA molecules. Enzymes appropriate for the ligation step are known in the art and include, e.g. Tth DNA ligase, Taq DNA ligase, *Thermococcus* sp. (strain 9°N) DNA ligase (9°N™ DNA ligase, New England Biolabs), Ampligase™ (Epicentre Biotechnologies) and T4 DNA ligase. In a preferred embodiment the first and second amplification domains comprise different sequences.

From the above, it is therefore apparent that universal domains, which may comprise an amplification domain, may be added to the secured (i.e. extended or ligated) DNA molecules, for example to the cDNA molecules, or their complements (e.g. second strand) by various methods and techniques and combinations of such techniques known in the art e.g. by use of primers which include such a domain, ligation of adaptors, use of terminal transferase enzymes and/or by template switching methods. As is clear from the discussion herein, such domains may be added before or after release of the DNA molecules from the array.

It will be apparent from the above description that all of the DNA, e.g. cDNA molecules from a single array that have been synthesized by the methods of the invention may all comprise the same first and second amplification domains. Consequently, a single amplification reaction, e.g. PCR, may be sufficient to amplify all of the DNA, e.g. cDNA molecules. Thus in a preferred embodiment, the method of the invention may comprise a step of amplifying the DNA, e.g. cDNA molecules. In one embodiment the amplification step is performed after the release of the DNA, e.g. cDNA molecules from the substrate of the array. In other embodiments

amplification may be performed on the array (i.e. *in situ* on the array). It is known in the art that amplification reactions may be carried out on arrays and on-chip thermocyclers exist for carrying out such reactions. Thus, in one embodiment arrays which are known in the art as sequencing platforms or for use in any form of sequence analysis (e.g. in or by next generation sequencing technologies) may be used as the basis of the arrays of the present invention (e.g. Illumina bead arrays etc.)

For the synthesis of the second strand of DNA, e.g. cDNA it is preferable to use a strand displacement polymerase (e.g. Φ29 DNA polymerase, Bst (exo^-) DNA polymerase, klenow (exo^-) DNA polymerase) if the cDNA released from the substrate of the array comprises a partially double stranded nucleic acid molecule. For instance, the released nucleic acids will be at least partially double stranded (e.g. DNA:DNA, DNA:RNA or DNA:DNA/RNA hybrid) in embodiments where the capture probe is immobilized indirectly on the substrate of the array via a surface probe and the step of releasing the DNA, e.g. cDNA molecules comprises a cleavage step. The strand displacement polymerase is necessary to ensure that the second cDNA strand synthesis incorporates the complement of the positional domain (feature identification domain) into the second DNA, e.g. cDNA strand.

It will be evident that the step of releasing at least part of the DNA, e.g. cDNA molecules or their amplicons from the surface or substrate of the array may be achieved using a number of methods. The primary aim of the release step is to yield molecules into which the positional domain of the capture probe (or its complement) is incorporated (or included), such that the DNA, e.g. cDNA molecules or their amplicons are "tagged" according to their feature (or position) on the array. The release step thus removes DNA, e.g. cDNA molecules or amplicons thereof from the array, which DNA, e.g. cDNA molecules or amplicons include the positional domain or its complement (by virtue of it having been incorporated into the secured nucleic acid, e.g. the first strand cDNA by, e.g. extension of the capture probe, and optionally copied in the second strand DNA if second strand synthesis takes place on the array, or copied into amplicons if amplification takes place on the array). Hence, in order to yield sequence analysis data that can be correlated with the various regions in the tissue sample it is essential that the released molecules comprise the positional domain of the capture probe (or its complement).

Since the released molecule may be a first and/or second strand DNA, e.g. cDNA molecule or amplicon, and since the capture probe may be immobilised

indirectly on the array, it will be understood that whilst the release step may comprise a step of cleaving a DNA, e.g. cDNA molecule from the array, the release step does not require a step of nucleic acid cleavage; a DNA, e.g. cDNA molecule or an amplicon may simply be released by denaturing a double-stranded molecule, for example releasing the second cDNA strand from the first cDNA strand, or releasing an amplicon from its template or releasing the first strand cDNA molecule (i.e. the extended capture probe) from a surface probe. Accordingly, a DNA, e.g. cDNA molecule may be released from the array by nucleic acid cleavage and/or by denaturation (e.g. by heating to denature a double-stranded molecule). Where amplification is carried out *in situ* on the array, this will of course encompass releasing amplicons by denaturation in the cycling reaction.

In some embodiments, the DNA, e.g. cDNA molecules are released by enzymatic cleavage of a cleavage domain, which may be located in the universal domain or positional domain of the capture probe. As mentioned above, the cleavage domain must be located upstream (at the 5' end) of the positional domain, such that the released DNA, e.g. cDNA molecules comprise the positional (identification) domain. Suitable enzymes for nucleic acid cleavage include restriction endonucleases, e.g. RsaI. Other enzymes, e.g. a mixture of Uracil DNA glycosylase (UDG) and the DNA glycosylase-lyase Endonuclease VIII (USER™ enzyme) or a combination of the MutY and T7 endonuclease I enzymes, are preferred embodiments of the methods of the invention.

In an alternative embodiment, the DNA, e.g. cDNA molecules may be released from the surface or substrate of the array by physical means. For instance, in embodiments where the capture probe is indirectly immobilized on the substrate of the array, e.g. via hybridization to the surface probe, it may be sufficient to disrupt the interaction between the nucleic acid molecules. Methods for disrupting the interaction between nucleic acid molecules, e.g. denaturing double stranded nucleic acid molecules, are well known in the art. A straightforward method for releasing the DNA, e.g. cDNA molecules (i.e. of stripping the array of the synthesized DNA, e.g. cDNA molecules) is to use a solution that interferes with the hydrogen bonds of the double stranded molecules. In a preferred embodiment of the invention, the DNA, e.g. cDNA molecules may be released by applying heated water, e.g. water or buffer of at least 85°C, preferably at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99°C. As an alternative or addition to the use of a temperature sufficient to disrupt the hydrogen bonding, the solution may comprise salts, surfactants etc.

that may further destabilize the interaction between the nucleic acid molecules, resulting in the release of the DNA, e.g. cDNA molecules.

It will be understood that the application of a high temperature solution, e.g. 90-99°C water may be sufficient to disrupt a covalent bond used to immobilize the capture probe or surface probe to the array substrate. Hence, in a preferred embodiment, the DNA, e.g. cDNA molecules may be released by applying hot water to the array to disrupt covalently immobilized capture or surface probes.

It is implicit that the released DNA, e.g. cDNA molecules (the solution comprising the released DNA, e.g. cDNA molecules) are collected for further manipulation, e.g. second strand synthesis and/or amplification. Nevertheless, the method of the invention may be seen to comprise a step of collecting or recovering the released DNA, e.g. cDNA molecules. As noted above, in the context of *in situ* amplification the released molecules may include amplicons of the secured nucleic acid, e.g. cDNA.

In embodiments of methods of the invention, it may be desirable to remove any unextended or unligated capture probes. This may be, for example, after the step of releasing DNA molecules from the array. Any desired or convenient method may be used for such removal including, for example, use of an enzyme to degrade the unextended or unligated probes, e.g. exonuclease.

The DNA, e.g. cDNA molecules, or amplicons, that have been released from the array, which may have been modified as discussed above, are analysed to investigate (e.g. determine their sequence, although as noted above actual sequence determination is not required - any method of analysing the sequence may be used). Thus, any method of nucleic acid analysis may be used. The step of sequence analysis may identify the positional domain and hence allow the analysed molecule to be localised to a position in the tissue sample. Similarly, the nature or identity of the analysed molecule may be determined. In this way the nucleic acid, e.g. RNA at given position in the array, and hence in the tissue sample may be determined. Hence the analysis step may include or use any method which identifies the analysed molecule (and hence the "target" molecule) and its positional domain. Generally such a method will be a sequence-specific method. For example, the method may use sequence-specific primers or probes, particularly primers or probes specific for the positional domain and/or for a specific nucleic acid molecule to be detected or analysed e.g. a DNA molecule corresponding to a

nucleic acid, e.g. RNA or cDNA molecule to be detected. Typically in such a method sequence-specific amplification primers e.g. PCR primers may be used.

In some embodiments it may be desirable to analyse a subset or family of target related molecules, e.g. all of the sequences that encode a particular group of proteins which share sequence similarity and/or conserved domains, e.g. a family of receptors. Hence, the amplification and/or analysis methods described herein may use degenerate or gene family specific primers or probes that hybridise to a subset of the captured nucleic acids or nucleic acids derived therefrom, e.g. amplicons. In a particularly preferred embodiment, the amplification and/or analysis methods may utilise a universal primer (i.e. a primer common to all of the captured sequences) in combination with a degenerate or gene family specific primer specific for a subset of target molecules.

Thus in one embodiment, amplification-based, especially PCR-based methods of sequence analysis are used.

However, the steps of modifying and/or amplifying the released DNA, e.g. cDNA molecules may introduce additional components into the sample, e.g. enzymes, primers, nucleotides etc. Hence, the methods of the invention may further comprise a step of purifying the sample comprising the released DNA, e.g. cDNA molecules or amplicons prior to the sequence analysis, e.g. to remove oligonucleotide primers, nucleotides, salts etc that may interfere with the sequencing reactions. Any suitable method of purifying the DNA, e.g. cDNA molecules may be used.

As noted above, sequence analysis of the released DNA molecules may be direct or indirect. Thus the sequence analysis substrate (which may be viewed as the molecule which is subjected to the sequence analysis step or process) may directly be the molecule which is released from the array or it may be a molecule which is derived therefrom. Thus, for example in the context of sequence analysis step which involves a sequencing reaction, the sequencing template may be the molecule which is released from the array or it may be a molecule derived therefrom. For example, a first and/or second strand DNA, e.g. cDNA molecule released from the array may be directly subjected to sequence analysis (e.g. sequencing), i.e. may directly take part in the sequence analysis reaction or process (e.g. the sequencing reaction or sequencing process, or be the molecule which is sequenced or otherwise identified). In the context of *in situ* amplification the released molecule may be an amplicon. Alternatively, the released molecule

may be subjected to a step of second strand synthesis or amplification before sequence analysis (e.g. sequencing or identification by other means). The sequence analysis substrate (e.g. template) may thus be an amplicon or a second strand of a molecule which is directly released from the array.

Both strands of a double stranded molecule may be subjected to sequence analysis (e.g. sequenced) but the invention is not limited to this and single stranded molecules (e.g. cDNA) may be analysed (e.g. sequenced). For example various sequencing technologies may be used for single molecule sequencing, e.g. the Helicos or Pacbio technologies, or nanopore sequencing technologies which are being developed. Thus, in one embodiment the first strand of DNA, e.g. cDNA may be subjected to sequencing. The first strand DNA, e.g. cDNA may need to be modified at the 3' end to enable single molecule sequencing. This may be done by procedures analogous to those for handling the second DNA, e.g. cDNA strand. Such procedures are known in the art.

In a preferred aspect of the invention the sequence analysis will identify or reveal a portion of captured nucleic acid, e.g. RNA sequence and the sequence of the positional domain. The sequence of the positional domain (or tag) will identify the feature to which the nucleic acid, e.g. mRNA molecule was captured. The sequence of the captured nucleic acid, e.g. RNA molecule may be compared with a sequence database of the organism from which the sample originated to determine the gene to which it corresponds. By determining which region (e.g. cell) of the tissue sample was in contact with the feature, it is possible to determine which region of the tissue sample was expressing said gene (or contained the gene, e.g. in the case of spatial genomics). This analysis may be achieved for all of the DNA, e.g. cDNA molecules generated by the methods of the invention, yielding a spatial transcriptome or genome of the tissue sample.

By way of a representative example, sequencing data may be analysed to sort the sequences into specific species of capture probe, i.e. according to the sequence of the positional domain. This may be achieved by, e.g. using the FastX toolkit FASTQ Barcode splitter tool to sort the sequences into individual files for the respective capture probe positional domain (tag) sequences. The sequences of each species, i.e. from each feature, may be analyzed to determine the identity of the transcripts. For instance, the sequences may be identified using e.g. Blastn software, to compare the sequences to one or more genome databases, preferably the database for the organism from which the tissue sample was obtained. The

identity of the database sequence with the greatest similarity to the sequence generated by the methods of the invention will be assigned to said sequence. In general, only hits with a certainty of at least $1e^{-6}$, preferably $1e^{-7}$, $1e^{-8}$, or $1e^{-9}$ will be considered to have been successfully identified.

It will be apparent that any nucleic acid sequencing method may be utilised in the methods of the invention. However, the so-called "next generation sequencing" techniques will find particular utility in the present invention. High-throughput sequencing is particularly useful in the methods of the invention because it enables a large number of nucleic acids to be partially sequenced in a very short period of time. In view of the recent explosion in the number of fully or partially sequenced genomes, it is not essential to sequence the full length of the generated DNA, e.g. cDNA molecules to determine the gene to which each molecule corresponds. For example, the first 100 nucleotides from each end of the DNA, e.g. cDNA molecules should be sufficient to identify both the feature to which the nucleic acid, e.g. mRNA was captured (i.e. its location on the array) and the gene expressed. The sequence reaction from the "capture probe end" of the DNA, e.g. cDNA molecules yields the sequence of the positional domain and at least about 20 bases, preferably 30 or 40 bases of transcript specific sequence data. The sequence reaction from the "non-capture probe end" may yield at least about 70 bases, preferably 80, 90, or 100 bases of transcript specific sequence data.

As a representative example, the sequencing reaction may be based on reversible dye-terminators, such as used in the Illumina™ technology. For example, DNA molecules are first attached to primers on, e.g. a glass or silicon slide and amplified so that local clonal colonies are formed (bridge amplification). Four types of ddNTPs are added, and non-incorporated nucleotides are washed away. Unlike pyrosequencing, the DNA can only be extended one nucleotide at a time. A camera takes images of the fluorescently labelled nucleotides then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing a next cycle. This may be repeated until the required sequence data is obtained. Using this technology, thousands of nucleic acids may be sequenced simultaneously on a single slide.

Other high-throughput sequencing techniques may be equally suitable for the methods of the invention, e.g. pyrosequencing. In this method the DNA is amplified inside water droplets in an oil solution (emulsion PCR), with each droplet containing a single DNA template attached to a single primer-coated bead that then

forms a clonal colony. The sequencing machine contains many picolitre-volume wells each containing a single bead and sequencing enzymes. Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA and the combined data are used to generate sequence read-outs.

An example of a technology in development is based on the detection of hydrogen ions that are released during the polymerisation of DNA. A microwell containing a template DNA strand to be sequenced is flooded with a single type of nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide it is incorporated into the growing complementary strand. This causes the release of a hydrogen ion that triggers a hypersensitive ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence multiple nucleotides will be incorporated in a single cycle. This leads to a corresponding number of released hydrogen ions and a proportionally higher electronic signal.

Thus, it is clear that future sequencing formats are slowly being made available, and with shorter run times as one of the main features of those platforms it will be evident that other sequencing technologies will be useful in the methods of the invention.

An essential feature of the present invention, as described above, is a step of securing a complementary strand of the captured nucleic acid molecules to the capture probe, e.g. reverse transcribing the captured RNA molecules. The reverse transcription reaction is well known in the art and in representative reverse transcription reactions, the reaction mixture includes a reverse transcriptase, dNTPs and a suitable buffer. The reaction mixture may comprise other components, e.g. RNase inhibitor(s). The primers and template are the capture domain of the capture probe and the captured RNA molecules are described above. In the subject methods, each dNTP will typically be present in an amount ranging from about 10 to 5000 μM, usually from about 20 to 1000 μM. It will be evident that an equivalent reaction may be performed to generate a complementary strand of a captured DNA molecule, using an enzyme with DNA polymerase activity. Reactions of this type are well known in the art and are described in more detail below.

The desired reverse transcriptase activity may be provided by one or more distinct enzymes, wherein suitable examples are: M-MLV, MuLV, AMV, HIV, ArrayScript™, MultiScribe™, ThermoScript™, and SuperScript® I, II, and III enzymes.

The reverse transcriptase reaction may be carried out at any suitable temperature, which will be dependent on the properties of the enzyme. Typically, reverse transcriptase reactions are performed between 37-55°C, although temperatures outside of this range may also be appropriate. The reaction time may be as little as 1, 2, 3, 4 or 5 minutes or as much as 48 hours. Typically the reaction will be carried out for between 5-120 minutes, preferably 5-60, 5-45 or 5-30 minutes or 1-10 or 1-5 minutes according to choice. The reaction time is not critical and any desired reaction time may be used.

As indicated above, certain embodiments of the methods include an amplification step, where the copy number of generated DNA, e.g. cDNA molecules is increased, e.g., in order to enrich the sample to obtain a better representation of the nucleic acids, e.g. transcripts captured from the tissue sample. The amplification may be linear or exponential, as desired, where representative amplification protocols of interest include, but are not limited to: polymerase chain reaction (PCR); isothermal amplification, etc.

The polymerase chain reaction (PCR) is well known in the art, being described in U.S. Pat. Nos.: 4,683,202; 4,683,195; 4,800,159; 4,965,188 and 5,512,462, the disclosures of which are herein incorporated by reference. In representative PCR amplification reactions, the reaction mixture that includes the above released DNA, e.g. cDNA molecules from the array, which are combined with one or more primers that are employed in the primer extension reaction, e.g., the PCR primers that hybridize to the first and/or second amplification domains (such as forward and reverse primers employed in geometric (or exponential) amplification or a single primer employed in a linear amplification). The oligonucleotide primers with which the released DNA, e.g. cDNA molecules (hereinafter referred to as template DNA for convenience) is contacted will be of sufficient length to provide for hybridization to complementary template DNA under annealing conditions (described in greater detail below). The length of the primers will depend on the length of the amplification domains, but will generally be at least 10 bp in length, usually at least 15 bp in length and more usually at least 16 bp in length and may be as long as 30 bp in length or longer, where the length of the primers will generally range from 18 to 50 bp in length, usually from about 20 to 35 bp in length. The template DNA may be contacted with a single primer or a set of two primers (forward and reverse primers), depending on whether primer extension, linear or exponential amplification of the template DNA is desired.

In addition to the above components, the reaction mixture produced in the subject methods typically includes a polymerase and deoxyribonucleoside triphosphates (dNTPs). The desired polymerase activity may be provided by one or more distinct polymerase enzymes. In many embodiments, the reaction mixture includes at least a Family A polymerase, where representative Family A polymerases of interest include, but are not limited to: Thermus aquaticus polymerases, including the naturally occurring polymerase (Taq) and derivatives and homologues thereof, such as Klentaq (as described in Barnes et al, Proc. Natl. Acad. Sci USA (1994) 91:2216-2220); Thermus thermophilus polymerases, including the naturally occurring polymerase (Tth) and derivatives and homologues thereof, and the like. In certain embodiments where the amplification reaction that is carried out is a high fidelity reaction, the reaction mixture may further include a polymerase enzyme having 3'-5' exonuclease activity, e.g., as may be provided by a Family B polymerase, where Family B polymerases of interest include, but are not limited to: Thermococcus litoralis DNA polymerase (Vent) as described in Perler et al., Proc. Natl. Acad. Sci. USA (1992) 89:5577-5581; Pyrococcus species GB-D (Deep Vent); Pyrococcus furiosus DNA polymerase (Pfu) as described in Lundberg et al., Gene (1991) 108:1-6, Pyrococcus woesei (Pwo) and the like. Where the reaction mixture includes both a Family A and Family B polymerase, the Family A polymerase may be present in the reaction mixture in an amount greater than the Family B polymerase, where the difference in activity will usually be at least 10-fold, and more usually at least about 100-fold. Usually the reaction mixture will include four different types of dNTPs corresponding to the four naturally occurring bases present, i.e. dATP, dTTP, dCTP and dGTP. In the subject methods, each dNTP will typically be present in an amount ranging from about 10 to 5000 μM, usually from about 20 to 1000 μM.

The reaction mixtures prepared in the reverse transcriptase and/or amplification steps of the subject methods may further include an aqueous buffer medium that includes a source of monovalent ions, a source of divalent cations and a buffering agent. Any convenient source of monovalent ions, such as KCl, K-acetate, NH_4-acetate, K-glutamate, NH_4Cl, ammonium sulphate, and the like may be employed. The divalent cation may be magnesium, manganese, zinc and the like, where the cation will typically be magnesium. Any convenient source of magnesium cation may be employed, including $MgCl_2$, Mg-acetate, and the like. The amount of Mg^{2+} present in the buffer may range from 0.5 to 10 mM, but will

preferably range from about 3 to 6 mM, and will ideally be at about 5 mM. Representative buffering agents or salts that may be present in the buffer include Tris, Tricine, HEPES, MOPS and the like, where the amount of buffering agent will typically range from about 5 to 150 mM, usually from about 10 to 100 mM, and more usually from about 20 to 50 mM, where in certain preferred embodiments the buffering agent will be present in an amount sufficient to provide a pH ranging from about 6.0 to 9.5, where most preferred is pH 7.3 at 72 °C. Other agents which may be present in the buffer medium include chelating agents, such as EDTA, EGTA and the like.

In preparing the reverse transcriptase, DNA extension or amplification reaction mixture of the steps of the subject methods, the various constituent components may be combined in any convenient order. For example, in the amplification reaction the buffer may be combined with primer, polymerase and then template DNA, or all of the various constituent components may be combined at the same time to produce the reaction mixture.

As discussed above, a preferred embodiment of the invention the DNA, e.g. cDNA molecules may be modified by the addition of amplification domains to the ends of the nucleic acid molecules, which may involve a ligation reaction. A ligation reaction is also required for the *in situ* synthesis of the capture probe on the array, when the capture probe is immobilized indirectly on the array surface.

As is known in the art, ligases catalyze the formation of a phosphodiester bond between juxtaposed 3'-hydroxyl and 5'-phosphate termini of two immediately adjacent nucleic acids. Any convenient ligase may be employed, where representative ligases of interest include, but are not limited to: Temperature sensitive and thermostable ligases. Temperature sensitive ligases include, but are not limited to, bacteriophage T4 DNA ligase, bacteriophage T7 ligase, and E. coli ligase. Thermostable ligases include, but are not limited to, Taq ligase, Tth ligase, and Pfu ligase. Thermostable ligase may be obtained from thermophilic or hyperthermophilic organisms, including but not limited to, prokaryotic, eukaryotic, or archael organisms. Certain RNA ligases may also be employed in the methods of the invention.

In this ligation step, a suitable ligase and any reagents that are necessary and/or desirable are combined with the reaction mixture and maintained under conditions sufficient for ligation of the relevant oligonucleotides to occur. Ligation reaction conditions are well known to those of skill in the art. During ligation, the

reaction mixture in certain embodiments may be maintained at a temperature ranging from about 4°C to about 50°C, such as from about 20°C to about 37°C for a period of time ranging from about 5 seconds to about 16 hours, such as from about 1 minute to about 1 hour. In yet other embodiments, the reaction mixture may be maintained at a temperature ranging from about 35°C to about 45°C, such as from about 37°C to about 42°C, e.g., at or about 38°C, 39°C, 40°C or 41°C, for a period of time ranging from about 5 seconds to about 16 hours, such as from about 1 minute to about 1 hour, including from about 2 minutes to about 8 hours. In a representative embodiment, the ligation reaction mixture includes 50 mM Tris pH7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 mg/ml BSA, 0.25 units/ml RNase inhibitor, and T4 DNA ligase at 0.125 units/ml. In yet another representative embodiment, 2.125 mM magnesium ion, 0.2 units/ml RNase inhibitor; and 0.125 units/ml DNA ligase are employed. The amount of adaptor in the reaction will be dependent on the concentration of the DNA, e.g. cDNA in the sample and will generally be present at between 10-100 times the molar amount of DNA, e.g. cDNA.

By way of a representative example the method of the invention may comprise the following steps:

(a) contacting an array with a tissue sample, wherein the array comprises a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

such that RNA of the tissue sample hybridises to said capture probes;

(b) imaging the tissue sample on the array;

(c) reverse transcribing the captured mRNA molecules to generate cDNA molecules;

(d) washing the array to remove residual tissue;

(e) releasing at least part of the cDNA molecules from the surface of the array;

(f) performing second strand cDNA synthesis on the released cDNA molecules;

and

(g) analysing the sequence of (e.g. sequencing) the cDNA molecules.

By way of an alternative representative example the method of the invention may comprise the following steps:

(a) contacting an array with a tissue sample, wherein the array comprises a substrate on which at least two species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

such that RNA of the tissue sample hybridises to said capture probes;

(b) optionally rehydrating the tissue sample;

(c) reverse transcribing the captured mRNA molecules to generate first strand cDNA molecules and optionally synthesising second strand cDNA molecules;

(d) imaging the tissue sample on the array;

(e) washing the array to remove residual tissue;

(f) releasing at least part of the cDNA molecules from the surface of the array;

(g) amplifying the released cDNA molecules;

and

(h) analysing the sequence of (e.g. sequencing) the amplified cDNA molecules.

By way of yet a further representative example the method of the invention may comprise the following steps:

(a) contacting an array with a tissue sample, wherein the array comprises a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a reverse

transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

such that RNA of the tissue sample hybridises to said capture probes;

(b) optionally imaging the tissue sample on the array;

(c) reverse transcribing the captured mRNA molecules to generate cDNA molecules;

(d) optionally imaging the tissue sample on the array if not already performed as step (b):

(e) washing the array to remove residual tissue;

(f) releasing at least part of the cDNA molecules from the surface of the array;

(g) performing second strand cDNA synthesis on the released cDNA molecules;

(h) amplifying the double stranded cDNA molecules;

(i) optionally purifying the cDNA molecules to remove components that may interfere with the sequencing reaction;

and

(j) analysing the sequence of (e.g. sequencing) the amplified cDNA molecules.

The present invention includes any suitable combination of the steps in the above described methods. It will be understood that the invention also encompasses variations of these methods, for example where amplification is performed *in situ* on the array. Also encompassed are methods which omit the imaging step.

The invention may also be seen to include a method for making or producing an array (i) for use in capturing mRNA from a tissue sample that is contacted with said array; or (ii) for use in determining and/or analysing a (e.g. the partial or global) transcriptome of a tissue sample, said method comprising immobilizing, directly or indirectly, multiple species of capture probe to an array substrate, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array; and

(ii) a capture domain.

The method of producing an array of the invention may be further defined such that each species of capture probe is immobilized as a feature on the array.

The method of immobilizing the capture probes on the array may be achieved using any suitable means as described herein. Where the capture probes are immobilized on the array indirectly the capture probe may be synthesized on the array. Said method may comprise any one or more of the following steps:

(a) immobilizing directly or indirectly multiple surface probes to an array substrate, wherein the surface probes comprise:

(i) a domain capable of hybridizing to part of the capture domain oligonucleotide (a part not involved in capturing the nucleic acid, e.g. RNA);

(ii) a complementary positional domain; and

(iii) a complementary universal domain;

(b) hybridizing to the surface probes immobilized on the array capture domain oligonucleotides and universal domain oligonucleotides;

(c) extending the universal domain oligonucleotides, by templated polymerisation, to generate the positional domain of the capture probe; and

(d) ligating the positional domain to the capture domain oligonucleotide to produce the capture oligonucleotide.

Ligation in step (d) may occur simultaneously with extension in step (c). Thus it need not be carried out in a separate step, although this is course encompassed if desired.

The features of the array produced by the above method of producing the array of the invention, may be further defined in accordance with the above description.

Although the invention is described above with reference to detection or analysis of RNA, and transcriptome analysis or detection, it will be appreciated that the principles described can be applied analogously to the detection or analysis of DNA in cells and to genomic studies. Thus, more broadly viewed, the invention can be seen as being generally applicable to the detection of nucleic acids in general and in a further more particular aspect, as providing methods for the analysis or detection of DNA. Spatial information may be valuable also in a genomics context i.e. detection and/or analysis of a DNA molecule with spatial resolution. This may

be achieved by genomic tagging according to the present invention. Such localised or spatial detection methods may be useful for example in the context of studying genomic variations in different cells or regions of a tissue, for example comparing normal and diseased cells or tissues (e.g. normal vs tumour cells or tissues) or in studying genomic changes in disease progression etc. For example, tumour tissues may comprise a heterogeneous population of cells which may differ in the genomic variants they contain (e.g. mutations and/or other genetic aberrations, for example chromosomal rearrangements, chromosomal amplifications/deletions/insertions etc.). The detection of genomic variations, or different genomic loci, in different cells in a localised way may be useful in such a context, e.g. to study the spatial distribution of genomic variations. A principal utility of such a method would be in tumour analysis. In the context of the present invention, an array may be prepared which is designed, for example, to capture the genome of an entire cell on one feature. Different cells in the tissue sample may thus be compared. Of course the invention is not limited to such a design and other variations may be possible, wherein the DNA is detected in a localised way and the position of the DNA captured on the array is correlated to a position or location in the tissue sample.

Accordingly, in a more general aspect, the present invention can be seen to provide a method for localised detection of nucleic acid in a tissue sample comprising:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample and allowing nucleic acid of the tissue sample to hybridise to the capture domain in said capture probes;

(c) generating DNA molecules from the captured nucleic acid molecules using said capture probes as extension or ligation primers, wherein said extended or ligated DNA molecules are tagged by virtue of the positional domain;

(d) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying said tagged DNA;

(e) releasing at least part of the tagged DNA molecules and/or their complements or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof ;

(f) directly or indirectly analysing the sequence of (e.g. sequencing) the released DNA molecules.

As described in more detail above, any method of nucleic acid analysis may be used in the analysis step. Typically this may involve sequencing, but it is not necessary to perform an actual sequence determination. For example sequence-specific methods of analysis may be used. For example a sequence-specific amplification reaction may be performed, for example using primers which are specific for the positional domain and/or for a specific target sequence, e.g. a particular target DNA to be detected (i.e. corresponding to a particular cDNA/RNA or gene or gene variant or genomic locus or genomic variant etc.). An exemplary analysis method is a sequence-specific PCR reaction.

The sequence analysis (e.g. sequencing) information obtained in step (f) may be used to obtain spatial information as to the nucleic acid in the sample. In other words the sequence analysis information may provide information as to the location of the nucleic acid in the sample. This spatial information may be derived from the nature of the sequence analysis information obtained e.g. from a sequence determined or identified, for example it may reveal the presence of a particular nucleic acid molecule which may itself be spatially informative in the context of the tissue sample used, and/or the spatial information (e.g. spatial localisation) may be derived from the position of the tissue sample on the array, coupled with the sequence analysis information. However, as described above, spatial information may conveniently be obtained by correlating the sequence analysis data to an image of the tissue sample and this represents one preferred embodiment of the invention.

Accordingly, in a preferred embodiment the method also includes a step of:

(g) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (c).

The primer extension reaction referred to in step (a) may be defined as a polymerase-catalysed extension reaction and acts to acquire a complementary strand of the captured nucleic acid molecule that is covalently attached to the capture probe, i.e. by synthesising the complementary strand utilising the capture probe as a primer and the captured nucleic acid as a template. In other words it may be any primer extension reaction carried out by any polymerase enzyme. The nucleic acid may be RNA or it may be DNA. Accordingly the polymerase may be any polymerase. It may be a reverse transcriptase or it may be a DNA polymerase. The ligation reaction may be carried out by any ligase and acts to secure the complementary strand of the captured nucleic acid molecule to the capture probe, i.e. wherein the captured nucleic acid molecule (hybridised to the capture probe) is partially double stranded and the complementary strand is ligated to the capture probe.

One preferred embodiment of such a method is the method described above for the determination and/or analysis of a transcriptome, or for the detection of RNA. In alternative preferred embodiment the detected nucleic acid molecule is DNA. In such an embodiment the invention provides a method for localised detection of DNA in a tissue sample comprising:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample and allowing DNA of the tissue sample to hybridise to the capture domain in said capture probes;

(c) fragmenting DNA in said tissue sample, wherein said fragmentation is carried out before, during or after contacting the array with the tissue sample in step (b);

(d) extending said capture probes in a primer extension reaction using the captured DNA fragments as templates to generate extended DNA molecules, or ligating the captured DNA fragments to the capture probes in a ligation reaction to generate ligated DNA molecules, wherein said extended or ligated DNA molecules are tagged by virtue of the positional domain;

(e) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying said tagged DNA;

(f) releasing at least part of the tagged DNA molecules and/or their complements and/or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof ;

(g) directly or indirectly analysing the sequence of the released DNA molecules.

The method may further include a step of:

(h) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (d).

In the context of spatial genomics, where the target nucleic acid is DNA the inclusion of imaging and image correlation steps may in some circumstances be preferred.

In embodiments in which DNA is captured, the DNA may be any DNA molecule which may occur in a cell. Thus it may be genomic, i.e. nuclear, DNA, mitochondrial DNA or plastid DNA, e.g. chloroplast DNA. In a preferred embodiment, the DNA is genomic DNA.

It will be understood that where fragmentation is carried out after the contacting in step (b), i.e. after the tissue sample is placed on the array, fragmentation occurs before the DNA is hybridised to the capture domain. In other words the DNA fragments are hybridised (or more particularly, allowed to hybridise) to the capture domain in said capture probes.

Advantageously, but not necessarily, in a particular embodiment of this aspect of the invention, the DNA fragments of the tissue sample may be provided with a binding domain to enable or facilitate their capture by the capture probes on the array. Accordingly, the binding domain is capable of hybridising to the capture domain of the capture probe. Such a binding domain may thus be regarded as a complement of the capture domain (i.e. it may be viewed as a complementary capture domain), although absolute complementarity between the capture and binding domains is not required, merely that the binding domain is sufficiently

complementary to allow a productive hybridisation to take place, i.e. that the DNA fragments in the tissue sample are able to hybridise to the capture domain of the capture probes. Provision of such a binding domain may ensure that DNA in the sample does not bind to the capture probes until after the fragmentation step. The binding domain may be provided to the DNA fragments by procedures well known in the art, for example by ligation of adaptor or linker sequences which may contain the binding domain. For example a linker sequence with a protruding end may be used. The binding domain may be present in the single-stranded portion of such a linker, such that following ligation of the linker to the DNA fragments, the single-stranded portion containing the binding domain is available for hybridisation to the capture domain of the capture probes. Alternatively and in a preferred embodiment, the binding domain may be introduced by using a terminal transferase enzyme to introduce a polynucleotide tail e.g. a homopolymeric tail such as a poly-A domain. This may be carried out using a procedure analogous to that described above for introducing a universal domain in the context of the RNA methods. Thus, in advantageous embodiments a common binding domain may be introduced. In other words, a binding domain which is common to all the DNA fragments and which may be used to achieve the capture of the fragments on the array.

Where a tailing reaction is carried out to introduce a (common) binding domain, the capture probes on the array may be protected from the tailing reaction, i.e. the capture probes may be blocked or masked as described above. This may be achieved for example by hybridising a blocking oligonucleotide to the capture probe e.g. to the protruding end (e.g. single stranded portion) of the capture probe. Where the capture domain comprises a poly-T sequence for example, such a blocking oligonucleotide may be a poly-A oligonucleotide. The blocking oligonucleotide may have a blocked 3' end (i.e. an end incapable of being extended, or tailed). The capture probes may also be protected, i.e. blocked, by chemical and/or enzymatic modifications, as described in detail above.

Where the binding domain is provided by ligation of a linker as described above, it will be understood that rather than extending the capture probe to generate a complementary copy of the captured DNA fragment which comprises the positional tag of the capture probe primer, the DNA fragment may be ligated to the 3' end of the capture probe. As noted above ligation requires that the 5' end to be ligated is phosphorylated. Accordingly, in one embodiment, the 5' end of the added linker, namely the end which is to be ligated to the capture probe (i.e. the

non-protruding end of the linker added to the DNA fragments) will be phosphorylated. In such a ligation embodiment, it will accordingly be seen that a linker may be ligated to double stranded DNA fragments, said linker having a single stranded protruding 3' end which contains the binding domain. Upon contact with the array, the protruding end hybridises to the capture domain of the capture probes. This hybridisation brings the 3' end of the capture probe into juxtaposition for ligation to the 5' (non-protruding) end of the added linker. The capture probe, and hence the positional domain, is thus incorporated into the captured DNA fragment by this ligation. Such an embodiment is shown schematically in Figure 21.

Thus, the method of this aspect of the invention may in a more particular embodiment comprise:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample;

(c) fragmenting DNA in said tissue sample, wherein said fragmentation is carried out before, during or after contacting the array with the tissue sample in step (b);

(d) providing said DNA fragments with a binding domain which is capable of hybridising to said capture domain;

(e) allowing said DNA fragments to hybridise to the capture domain in said capture probes;

(f) extending said capture probes in a primer extension reaction using the captured DNA fragments as templates to generate extended DNA molecules, or ligating the captured DNA fragments to the capture probes in a ligation reaction to generate ligated DNA molecules, wherein said extended or ligated DNA molecules are tagged by virtue of the positional domain;

(g) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying the tagged DNA;

(h) releasing at least part of the tagged DNA molecules and/or their complements and/or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof;

(i) directly or indirectly analysing the sequence of the released DNA molecules.

The method may optionally include a further step of

(j) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (f).

In the methods of nucleic acid or DNA detection set out above, the optional step of generating a complementary copy of the tagged nucleic acid/DNA or of amplifying the tagged DNA, may involve the use of a strand displacing polymerase enzyme, according to the principles explained above in the context of the RNA/transcriptome analysis/detection methods. Suitable strand displacing polymerases are discussed above. This is to ensure that the positional domain is copied into the complementary copy or amplicon. This will particularly be the case where the capture probe is immobilized on the array by hybridisation to a surface probe.

However, the use of a strand displacing polymerase in this step is not essential. For example a non-strand displacing polymerase may be used together with ligation of an oligonucleotide which hybridises to the positional domain. Such a procedure is analogous to that described above for the synthesis of capture probes on the array.

In one embodiment, the method of the invention may be used for determining and/or analysing all of the genome of a tissue sample e.g. the global genome of a tissue sample. However, the method is not limited to this and encompasses determining and/or analysing all or part of the genome. Thus, the method may involve determining and/or analysing a part or subset of the genome, e.g. a partial genome corresponding to a subset or group of genes or of chromosomes, e.g. a set of particular genes or chromosomes or a particular region or part of the genome, for example related to a particular disease or condition, tissue type etc. Thus, the method may be used to detect or analyse genomic sequences or genomic loci from tumour tissue as compared to normal tissue, or even within different types of cell in a tissue sample. The presence or absence, or

the distribution or location of different genomic variants or loci in different cells, groups of cells, tissues or parts or types of tissue may be examined.

Viewed from another aspect, the method steps set out above can be seen as providing a method of obtaining spatial information regarding the nucleic acids, e.g. genomic sequences, variants or loci of a tissue sample. Put another way, the methods of the invention may be used for the labelling (or tagging) of genomes, particularly individual or spatially distributed genomes.

Alternatively viewed, the method of the invention may be seen as a method for spatial detection of DNA in a tissue sample, or a method for detecting DNA with spatial resolution, or for localised or spatial determination and/or analysis of DNA in a tissue sample. In particular, the method may be used for the localised or spatial detection or determination and/or analysis of genes or genomic sequences or genomic variants or loci (e.g. distribution of genomic variants or loci) in a tissue sample. The localised/spatial detection/determination/analysis means that the DNA may be localised to its native position or location within a cell or tissue in the tissue sample. Thus for example, the DNA may be localised to a cell or group of cells, or type of cells in the sample, or to particular regions of areas within a tissue sample. The native location or position of the DNA (or in other words, the location or position of the DNA in the tissue sample), e.g. a genomic variant or locus, may be determined.

It will be seen therefore that the array of the present invention may be used to capture nucleic acid, e.g. DNA of a tissue sample that is contacted with said array. The array may also be used for determining and/or analysing a partial or global genome of a tissue sample or for obtaining a spatially defined partial or global genome of a tissue sample. The methods of the invention may thus be considered as methods of quantifying the spatial distribution of one or more genomic sequences (or variants or loci) in a tissue sample. Expressed another way, the methods of the present invention may be used to detect the spatial distribution of one or more genomic sequences or genomic variants or genomic loci in a tissue sample. In yet another way, the methods of the present invention may be used to determine simultaneously the location or distribution of one or more genomic sequences or genomic variants or genomic loci at one or more positions within a tissue sample. Still further, the methods may be seen as methods for partial or global analysis of the nucleic acid e.g. DNA of a tissue sample with spatial resolution e.g. two-dimensional spatial resolution.

The invention can also be seen to provide an array for use in the methods of the invention comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as an extension or ligation primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain to capture nucleic acid of a tissue sample that is contacted with said array.

In one aspect the nucleic acid molecule to be captured is DNA. The capture domain may be specific to a particular DNA to be detected, or to a particular class or group of DNAs, e.g. by virtue of specific hybridisation to a specific sequence of motif in the target DNA e.g. a conserved sequence, by analogy to the methods described in the context of RNA detection above. Alternatively the DNA to be captured may be provided with a binding domain, e.g. a common binding domain as described above, which binding domain may be recognised by the capture domain of the capture probes. Thus, as noted above, the binding domain may for example be a homopolymeric sequence e.g. poly-A. Again such a binding domain may be provided according to or analogously to the principles and methods described above in relation to the methods for RNA/transcriptome analysis or detection. In such a case, the capture domain may be complementary to the binding domain introduced into the DNA molecules of the tissue sample.

As also described in the RNA context above, the capture domain may be a random or degenerate sequence. Thus, DNA may be captured non-specifically by binding to a random or degenerate capture domain or to a capture domain which comprises at least partially a random or degenerate sequence.

In a related aspect, the present invention also provides use of an array, comprising a substrate on which multiple species of capture probe are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array; and

(ii) a capture domain;

to capture nucleic acid, e.g. DNA or RNA, of a tissue sample that is contacted with said array.

Preferably, said use is for localised detection of nucleic acid in a tissue sample and further comprises steps of:

(a) generating DNA molecules from the captured nucleic acid molecules using said capture probes as extension or ligation primers, wherein said extended or ligated molecules are tagged by virtue of the positional domain;

(b) optionally generating a complementary strand of said tagged nucleic acid and/or amplifying said tagged nucleic acid;

(c) releasing at least part of the tagged DNA molecules and/or their complements or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof;

(d) directly or indirectly analysing the sequence of the released DNA molecules; and optionally

(e) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (a).

The step of fragmenting DNA in a tissue sample may be carried out using any desired procedure known in the art. Thus physical methods of fragmentation may be used e.g. sonication or ultrasound treatment. Chemical methods are also known. Enzymatic methods of fragmentation may also be used, e.g. with endonucleases, for example restriction enzymes. Again methods and enzymes for this are well known in the art. Fragmentation may be done before during or after preparing the tissue sample for placing on an array, e.g. preparing a tissue section. Conveniently, fragmentation may be achieved in the step of fixing tissue. Thus for example, formalin fixation will result in fragmentation of DNA. Other fixatives may produce similar results.

In terms of the detail of preparing and using the arrays in these aspects of the invention, it will understood that the description and detail given above in the context of RNA methods applies analogously to the more general nucleic acid detection and DNA detection methods set out herein. Thus, all aspects and details discussed above apply analogously. For example, the discussion of reverse transcriptase primers and reactions etc may be applied analogously to any aspect of the extension primers, polymerase reactions etc. referred to above. Likewise, references and to first and second strand cDNA synthesis may be applied

analogously to the tagged DNA molecule and its complement. Methods of sequence analysis as discussed above may be used.

By way of example, the capture domain may be as described for the capture probes above. A poly-T or poly-T-containing capture domain may be used for example where the DNA fragments are provided with a binding domain comprising a poly-A sequence.

The capture probes/tagged DNA molecules (i.e. the tagged extended or ligated molecules) may be provided with universal domains as described above, e.g. for amplification and/or cleavage.

The invention will be further described with reference to the following non-limiting Examples with reference to the following drawings in which:

Figure 1 shows the overall concept using arrayed "barcoded" oligo-dT probes to capture mRNA from tissue sections for transcriptome analysis.

Figure 2 shows the a schematic for the visualization of transcript abundance for corresponding tissue sections.

Figure 3 shows 3' to 5' surface probe composition and synthesis of 5' to 3' oriented capture probes that are indirectly immobilized at the array surface.

Figure 4 shows a bar chart demonstrating the efficiency of enzymatic cleavage (USER or RsaI) from in-house manufactured arrays and by 99°C water from Agilent manufactured arrays, as measured by hybridization of fluorescently labelled probes to the array surface after probe release.

Figure 5 shows a fluorescent image captured after 99°C water mediated release of DNA surface probes from commercial arrays manufactured by Agilent. A fluorescent detection probe was hybridized after hot water treatment. Top array is an untreated control.

Figure 6 shows a fixated mouse brain tissue section on top of the transcriptome capture array post cDNA synthesis and treated with cytoplasmic (top) and nucleic stains (middle), respectively, and merged image showing both stains (bottom).

Figure 7 shows a table that lists the reads sorted for their origin across the low density in-house manufactured DNA-capture array as seen in the schematic representation.

Figure 8 shows a FFPE mouse brain tissue with nucleic and Map2 specific stains using a barcoded microarray.

Figure 9 shows FFPE mouse brain olfactory bulb with nucleic stain (white) and visible morphology.

Figure 10 shows FFPE mouse brain olfactory bulb (approx 2x2mm) with nucleic stain (white), overlaid with theoretical spotting pattern for low resolution array.

Figure 11 shows FFPE mouse brain olfactory bulb (approx 2x2mm) with nucleic stain (white), overlaid with theoretical spotting pattern for medium-high resolution array.

Figure 12 shows FFPE mouse brain olfactory bulb zoomed in on glomerular area (top right of Figure 9).

Figure 13 shows the resulting product from a USER release using a random hexamer primer (R6) coupled to the B_handle (B_R6) during amplification; product as depicted on a bioanalyzer.

Figure 14 shows the resulting product from a USER release using a random octamer primer (R8) coupled to the B_handle (B_R8) during amplification; product as depicted on a bioanalyzer.

Figure 15 shows the results of an experiment performed on FFPE brain tissue covering the whole array. ID5 (left) and ID20 (right) amplified with ID specific and gene specific primers (B2M exon 4) after synthesis and release of cDNA from surface; ID5 and ID20 amplified.

Figure 16 shows a schematic illustration of the principle of the method described in Example 4, i.e. use of microarrays with immobilized DNA oligos (capture probes) carrying spatial labeling tag sequences (positional domains). Each feature of oligos of the microarray carries a 1) a unique labeling tag (positional domain) and 2) a capture sequence (capture domain).

Figure 17 shows the results of the spatial genomics protocol described in Example 5 carried out with genomic DNA prefragmented to mean size of 200 bp. Internal products amplified on array labeled and synthesized DNA. The detected peak is of expected size.

Figure 18 shows the results of the spatial genomics protocol described in Example 5 carried out with genomic DNA prefragmented to mean size of 700 bp. Internal products amplified on array labeled and synthesized DNA. The detected peak is of expected size.

Figure 19 shows the results of the spatial genomics protocol described in Example 5 carried out with genomic DNA prefragmented to mean size of 200 bp.

Products amplified with one internal primer and one universal sequence contained in the surface oligo. Amplification carried out on array labeled and synthesized DNA. The expected product is a smear given that the random fragmentation and terminal transferase labeling of genomic DNA will generate a very diverse sample pool.

Figure 20 shows the results of the spatial genomics protocol described in Example 5 carried out with genomic DNA prefragmented to mean size of 700 bp. Products amplified with one internal primer and one universal sequence contained in the surface oligo. Amplification carried out on array labeled and synthesized DNA. The expected product is a smear given that the random fragmentation and terminal transferase labeling of genomic DNA will generate a very diverse sample pool.

Figure 21 shows a schematic illustration of the ligation of a linker to a DNA fragment to introduce a binding domain for hybridisation to a poly-T capture domain, and subsequent ligation to the capture probe.

Figure 22 shows the composition of 5' to 3' oriented capture probes used on high-density capture arrays.

Figure 23 shows the frame of the high-density arrays, which is used to orientate the tissue sample, visualized by hybridization of fluorescent marker probes.

Figure 24 shows capture probes cleaved and non-cleaved from high-density array, wherein the frame probes are not cleaved since they do not contain uracil bases. Capture probes were labelled with fluorophores coupled to poly-A oligonucleotides.

Figure 25 shows a bioanalyzer image of a prepared sequencing library with transcripts captured from mouse olfactory bulb.

Figure 26 shows a Matlab visualization of captured transcripts from total RNA extracted from mouse olfactory bulb.

Figure 27 shows Olfr (olfactory receptor) transcripts as visualized across the capture array using Matlab visualization after capture from mouse olfactory bulb tissue.

Figure 28 shows a pattern of printing for in-house 41-ID-tag microarrays.

Figure 29 shows a spatial genomics library generated from a A431 specific translocation after capture of poly-A tailed genomic fragments on capture array.

Figure 30 shows the detection of A431 specific translocation after capture of spiked 10% and 50% poly-A tailed A431 genomic fragments into poly-A tailed U2OS genomic fragments on capture array.

Figure 31 shows a Matlab visualization of captured ID-tagged transcripts from mouse olfactory bulb tissue on 41-ID-tag in-house arrays overlaid with the tissue image. For clarity, the specific features on which particular genes were identified have been circled.

Example 1

Preparation of the array

The following experiments demonstrate how oligonucleotide probes may be attached to an array substrate by either the 5' or 3' end to yield an array with capture probes capable of hybridizing to mRNA.

Preparation of in-house printed microarray with 5' to 3' oriented probes

20 RNA-capture oligonucleotides with individual tag sequences (Tag 1-20, Table 1 were spotted on glass slides to function as capture probes. The probes were synthesized with a 5'-terminus amino linker with a C6 spacer. All probes where synthesized by Sigma-Aldrich (St. Louis, MO, USA). The RNA-capture probes were suspended at a concentration of 20 µM in 150 mM sodium phosphate, pH 8.5 and were spotted using a Nanoplotter NP2.1/E (Gesim, Grosserkmannsdorf, Germany) onto CodeLink™ Activated microarray slides (7.5cm x 2.5cm; Surmodics, Eden Prairie, MN, USA). After printing, surface blocking was performed according to the manufacturer's instructions. The probes were printed in 16 identical arrays on the slide, and each array contained a pre-defined printing pattern. The 16 sub-arrays were separated during hybridization by a 16-pad mask (ChipClip™ Schleicher & Schuell BioScience, Keene, NH, USA).

Table 1

Name	Sequence	5' mod	3' mod	Length
Sequences for free 3' capture probes				
TAP-ID1	UUAAGTACAAATCTCGACTGCCACTCTGAACCTTCTCCTTCTCCTTCACCTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 1)	Amino-C6		72
Enzymatic recog	UUAAGTACAA (SEQ ID NO: 2)			10
Universal amp handle P	ATCTCGACTGCCACTCTGAA (SEQ ID NO: 3)			20
ID1	CCTTCTCCTTCTCCTTCACC (SEQ ID NO: 4)			20
Capture sequence	TTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 5)			22
ID1	CCTTCTCCTTCTCCTTCACC (SEQ ID NO: 6)			20
ID2	CCTTGCTGCTTCTCCTCCTC (SEQ ID NO: 7)			20
ID3	ACCTCCTCCGCCTCCTCCTC (SEQ ID NO: 8)			20
ID4	GAGACATACCACCAAGAGAC (SEQ ID NO: 9)			20
ID5	GTCCTCTATTCCGTCACCAT (SEQ ID NO: 10)			20
ID6	GACTGAGCTCGAACATATGG (SEQ ID NO: 11)			20
ID7	TGGAGGATTGACACAGAACG (SEQ ID NO: 12)			20
ID8	CCAGCCTCTCCATTACATCG (SEQ ID NO: 13)			20
ID9	AAGATCTACCAGCCAGCCAG (SEQ ID NO: 14)			20
ID10	CGAACTTCCACTGTCTCCTC (SEQ ID NO: 15)			20
ID11	TTGCGCCTTCTCCAATACAC (SEQ ID NO: 16)			20
ID12	CTCTTCTTAGCATGCCACCT (SEQ ID NO: 17)			20

ID13	ACCACTTCTGCATTACCTCC (SEQ ID NO: 18)		20
ID14	ACAGCCTCCTCTTCTTCCTT (SEQ ID NO: 19)		20
ID15	AATCCTCTCCTTGCCAGTTC (SEQ ID NO: 20)		20
ID16	GATGCCTCCACCTGTAGAAC (SEQ ID NO: 21)		20
ID17	GAAGGAATGGAGGATATCGC (SEQ ID NO: 22)		20
ID18	GATCCAAGGACCATCGACTG (SEQ ID NO: 23)		20
ID19	CCACTGGAACCTGACAACCG (SEQ ID NO: 24)		20
ID20	CTGCTTCTTCCTGGAACTCA (SEQ ID NO: 25)		20

Sequences for free 5' surface probes and on-chip free 3' capture probe synthesis

Free 5' surface probe - A	GCGTTCAGAGTGGCAGTCGAGATCACGCGGCAATCATATCGGACAGATCGGAAGAGCGTAGTGTAG (SEQ ID NO: 26)	Amino C7	66
Free 5' surface probe - U	GCGTTCAGAGTGGCAGTCGAGATCACGCGGCAATCATATCGGACGGCTGCTGGTAAATAGAGATCA (SEQ ID NO: 27)	Amino C7	66

Nick	GCG		3
LP'	TTCAGAGTGGCAGTCGAGATCAC (SEQ ID NO: 28)		23
ID'	GCGGCAATCATATCGGAC (SEQ ID NO: 29)		18
A' 22bp MutY mismatch	AGATCGGAAGAGCGTAGTGTAG (SEQ ID NO: 30)		22
U' 22bp MutY mismatch	GGCTGCTGGTAAATAGAGATCA (SEQ ID NO: 31)		

Hybridized sequences for capture probe synthesis

Illumina amp handle A	ACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 32)		33
Universa ampl handle U	AAGTGTGGAAAGTTGATCGCTATTTACCAGCAGCC (SEQ ID NO: 33)		35
Capture_LP_Poly-dTVN	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 34)	Phosphorylated	45
Capture_LP_Poly-d24T	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 35)	Phosphorylated	47

Additional secondary universal amplification handles

Illumina amp handle B	AGACGTGTGCTCTTCCGATCT (SEQ ID NO: 36)	21
Universal amp handle X	ACGTCTGTGAATAGCCGCAT (SEQ ID NO: 37)	20
B_R6 handle (or X)	AGACGTGTGCTCTTCCGATCTNNNNNNNNN (SEQ ID NO: 38)	27 (26)
B_R8 handle (or X)	AGACGTGTGCTCTTCCGATCTNNNNNNNNNN (SEQ ID NO: 39)	29 (28)
B_polyTVN (or X)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 40)	43 (42)
B_poly24T (or X)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 41)	45 (44)

Amplification handle to incorporate A handle into P handle products

A_P handle	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATCTCGACTGCCACTCTGAA (SEQ ID NO: 42)	53

Preparation of in-house printed microarray with 3' to 5' oriented probes and synthesis of 5' to 3' oriented capture probes

Printing of surface probe oligonucleotides was performed as in the case with 5' to 3' oriented probes above, with an amino-C7 linker at the 3' end, as shown in Table 1.

To hybridize primers for capture probe synthesis, hybridization solution containing 4xSSC and 0.1% SDS, 2μM extension primer (the universal domain oligonucleotide) and 2μM thread joining primer (the capture domain oligonucleotide) was incubated for 4 min at 50°C. Meanwhile the in-house array was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50μL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake; 2) 0.2xSSC for 1 min at 300 rpm shake; and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry and placed back in the ChipClip.

For extension and ligation reaction (to generate the positional domain of the capture probe) 50μL of enzyme mix containing 10xAmpligase buffer, 2.5 U AmpliTaq DNA Polymerase Stoffel Fragment (Applied Biosystems), 10 U Ampligase (Epicentre Biotechnologies), dNTPs 2 mM each (Fermentas) and water, was pipetted to each well. The array was subsequently incubated at 55°C for 30 min. After incubation the array was washed according to the previously described array washing method but the first step has the duration of 10 min instead of 6 min.

The method is depicted in Figure 3.

Tissue Preparation

The following experiments demonstrate how tissue sample sections may be prepared for use in the methods of the invention.

Preparation of fresh frozen tissue and sectioning onto capture probe arrays

Fresh non-fixed mouse brain tissue was trimmed if necessary and frozen down in -40°C cold isopentane and subsequently mounted for sectioning with a cryostat at 10μm. A slice of tissue was applied onto each capture probe array to be used.

Preparation of formalin-fixed paraffin-embedded (FFPE) tissue

Mouse brain tissue was fixed in 4% formalin at 4°C for 24h. After that it was incubated as follows: 3x incubation in 70% ethanol for 1 hour; 1x incubation in 80% ethanol for 1 hour; 1x incubation in 96% ethanol for 1 hour; 3x incubation in 100% ethanol for 1 hour; and 2x incubation in xylene at room temperature for 1 h.

The dehydrated samples were then incubated in liquid low melting paraffin 52-54°C for up to 3 hours, during which the paraffin was changed once to wash out residual xylene. Finished tissue blocks were then stored at RT. Sections were then cut at 4µm in paraffin with a microtome onto each capture probe array to be used.

The sections were dried at 37°C on the array slides for 24 hours and stored at RT.

Deparaffinization of FFPE tissue

Formalin fixed paraffinized mouse brain 10µm sections attached to CodeLink slides were deparaffinised in xylene twice for: 10 min, 99.5% ethanol for 2 min; 96% ethanol for 2 min; 70% ethanol for 2 min; and were then air dried.

cDNA synthesis

The following experiments demonstrate that mRNA captured on the array from the tissue sample sections may be used as template for cDNA synthesis.

cDNA synthesis on chip

A 16 well mask and Chip Clip slide holder from Whatman was attached to a CodeLink slide. The SuperScript™III One-step RT-PCR System with Platinum®Taq DNA Polymerase from Invitrogen was used when performing the cDNA synthesis. For each reaction 25 µl 2x reaction mix (SuperScript™III One-step RT-PCR System with Platinum®Taq DNA Polymerase, Invitrogen), 22.5 µl H_2O and 0.5 µl 100xBSA were mixed and heated to 50°C. SuperScript III/Platinum Taq enzyme mix was added to the reaction mix, 2 µl per reaction, and 50 µl of the reaction mix was added to each well on the chip. The chip was incubated at 50°C for 30 min (Thermomixer Comfort, Eppendorf).

The reaction mix was removed from the wells and the slide was washed with: 2xSSC, 0.1% SDS at 50°C for 10 min; 0.2xSSC at room temperature for 1 min; and 0.1xSSC at room temperature for 1 min. The chip was then spin dried.

In the case of FFPE tissue sections, the sections could now be stained and visualized before removal of the tissue, see below section on visualization.

Visualization

Hybridization of fluorescent marker probes prior to staining

Prior to tissue application fluorescent marker probes were hybridized to features comprising marker oligonucleotides printed on the capture probe array. The fluorescent marker probes aid in the orientation of the resulting image after tissue visualization, making it possible to combine the image with the resulting expression profiles for individual capture probe "tag" (positional domain) sequences obtained after sequencing. To hybridize fluorescent probes a hybridization solution containing 4xSSC and 0.1% SDS, 2µM detection probe (P) was incubated for 4 min at 50°C. Meanwhile the in-house array was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50µL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry.

General histological staining of FFPE tissue sections prior to or post cDNA synthesis

FFPE tissue sections immobilized on capture probe arrays were washed and rehydrated after deparaffinization prior to cDNA synthesis as described previously, or washed after cDNA synthesis as described previously. They are then treated as follows: incubate for 3 minutes in Hematoxylin; rinse with deionized water; incubate 5 minutes in tap water; rapidly dip 8 to 12 times in acid ethanol; rinse 2x1 minute in tap water; rinse 2 minutes in deionized water; incubate 30 seconds in Eosin; wash 3x5 minutes in 95% ethanol; wash 3x5 minutes in 100% ethanol; wash 3x10 minutes in xylene (can be done overnight); place coverslip on slides using DPX; dry slides in the hood overnight.

General immunohistochemistry staining of a target protein in FFPE tissue sections prior to or post cDNA synthesis

FFPE tissue sections immobilized on capture probe arrays were washed and rehydrated after deparaffinization prior to cDNA synthesis as described previously, or washed after cDNA synthesis as described previously. They were

then treated as follows without being allowed to dry during the whole staining process: sections were incubated with primary antibody (dilute primary antibody in blocking solution comprising 1xTris Buffered Saline (50mM Tris, 150mM NaCl, pH 7.6), 4% donkey serum and 0.1% triton-x) in a wet chamber overnight at RT; rinse three times with 1xTBS; incubate section with matching secondary antibody conjugated to a fluorochrome (FITC, Cy3 or Cy5) in a wet chamber at RT for 1 hour. Rinse 3x with 1xTBS, remove as much as possible of TBS and mount section with ProLong Gold +DAPI (Invitrogen) and analyze with fluorescence microscope and matching filter sets.

Removal of residual tissue

Frozen tissue

For fresh frozen mouse brain tissue the washing step directly following cDNA synthesis was enough to remove the tissue completely.

FFPE tissue

The slides with attached formalin fixed paraffinized mouse brain tissue sections were attached to ChipClip slide holders and 16 well masks (Whatman). For each 150 µl Proteinase K Digest Buffer from the RNeasy FFPE kit (Qiagen), 10 µl Proteinase K Solution (Qiagen) was added. 50 µl of the final mixture was added to each well and the slide was incubated at 56°C for 30 min.

Capture probe (cDNA) release

Capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes)

A 16 well mask and CodeLink slide was attached to the ChipClip holder (Whatman). 50µl of a mixture containing 1x FastStart High Fidelity Reaction Buffer with 1.8 mM MgCl2 (Roche), 200 µM dNTPs (New England Biolabs) and 0.1U/1 µl USER Enzyme (New England Biolabs) was heated to 37°C and was added to each well and incubated at 37°C for 30 min with mixing (3 seconds at 300 rpm, 6 seconds at rest) (Thermomixer comfort; Eppendorf). The reaction mixture containing the released cDNA and probes was then recovered from the wells with a pipette.

Capture probe release with uracil cleaving USER enzyme mixture in TdT (terminal transferase) buffer (covalently attached probes)

50µl of a mixture containing: 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com); 0.1µg/µl BSA (New England Biolabs); and 0.1U/µl USER Enzyme (New England Biolabs) was heated to 37°C and was added to each well and incubated at 37°C for 30 min with mixing (3 seconds at 300 rpm, 6 seconds at rest) (Thermomixer comfort; Eppendorf). The reaction mixture containing the released cDNA and probes was then recovered from the wells with a pipette.

Capture probe release with boiling hot water (covalently attached probes)

A 16 well mask and CodeLink slide was attached to the ChipClip holder (Whatman). 50µl of 99°C water was pipetted into each well. The 99°C water was allowed to react for 30 minutes. The reaction mixture containing the released cDNA and probes was then recovered from the wells with a pipette.

Capture probe release with heated PCR buffer (hybridized *in situ* synthesized capture probes, i.e. capture probes hybridized to surface probes)

50µl of a mixture containing: 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com); 0.1µg/µl BSA (New England Biolabs); and 0.1U/µl USER Enzyme (New England Biolabs) was preheated to 95°C. The mixture was then added to each well and incubated for 5 minutes at 95°C with mixing (3 seconds at 300 rpm, 6 seconds at rest) (Thermomixer comfort; Eppendorf). The reaction mixture containing the released probes was then recovered from the wells.

Capture probe release with heated TdT (terminal transferase) buffer (hybridized *in situ* synthesized capture probes, i.e. capture probes hybridized to surface probes)

50µl of a mixture containing: 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com); 0.1µg/µl BSA (New England Biolabs); and 0.1U/µl USER Enzyme (New England Biolabs) was preheated to 95°C. The mixture was then added to each well and incubated for 5 minutes at 95°C with mixing (3 seconds at 300 rpm, 6

seconds at rest) (Thermomixer comfort; Eppendorf). The reaction mixture containing the released probes was then recovered from the wells.

The efficacy of treating the array with the USER enzyme and water heated to 99°C can be seen in Figure 3. Enzymatic cleavage using the USER enzyme and the Rsal enzyme was performed using the "in-house" arrays described above (Figure 4). Hot water mediated release of DNA surface probes was performed using commercial arrays manufactured by Agilent (see Figure 5).

Probe collection and linker introduction

The experiments demonstrate that first strand cDNA released from the array surface may be modified to produce double stranded DNA and subsequently amplified.

Whole Transcriptome Amplification by the Picoplex whole genome amplification kit (capture probe sequences including positional domain (tag) sequences not retained at the edge of the resulting dsDNA)

Capture probes were released with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached capture probes) or with heated PCR buffer (hybridized *in situ* synthesized capture probes, i.e. capture probes hybridized to surface probes).

The released cDNA was amplified using the Picoplex (Rubicon Genomics) random primer whole genome amplification method, which was carried out according to manufacturers instructions.

Whole Transcriptome Amplification by dA tailing with Terminal Transferase (TdT) (capture probe sequences including positional domain (tag) sequences retained at the end of the resulting dsDNA)

Capture probes were released with uracil cleaving USER enzyme mixture in TdT (terminal transferase) buffer (covalently attached capture probes) or with heated TdT (terminal transferase) buffer (hybridized *in situ* synthesized capture probes, i.e. capture probes hybridized to surface probes).

38µl of cleavage mixture was placed in a clean 0.2ml PCR tube. The mixture contained: 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com), 0.1µg/µl BSA (New England Biolabs); 0.1U/µl USER Enzyme (New England Biolabs) (not for heated release); released cDNA (extended from surface probes);

and released surface probes. To the PCR tube, 0.5µl RNase H (5U/µl, final concentration of 0.06U/µl), 1µl TdT (20U/µl, final concentration of 0.5U/µl), and 0.5µl dATPs (100mM, final concentration of 1.25mM), were added. For dA tailing, the tube was incubated in a thermocycler (Applied Biosystems) at 37°C for 15 min followed by an inactivation of TdT at 70°C for 10 min. After dA tailing, a PCR master mix was prepared. The mix contained: 1x Faststart HiFi PCR Buffer (pH 8.3) with 1.8mM $MgCl_2$ (Roche); 0.2mM of each dNTP (Fermentas); 0.2µM of each primer, A (complementary to the amplification domain of the capture probe) and B_(dT)24 (Eurofins MWG Operon) (complementary to the poly-A tail to be added to the 3' end of the first cDNA strand); and 0.1U/µl Faststart HiFi DNA polymerase (Roche). 23µl of PCR Master mix was placed into nine clean 0.2ml PCR tubes. 2µl of dA tailing mixture were added to eight of the tubes, while 2µl water (RNase/DNase free) was added to the last tube (negative control). PCR amplification was carried out with the following program: Hot start at 95°C for 2 minutes, second strand synthesis at 50°C for 2 minutes and 72°C for 3 minutes, amplification with 30 PCR cycles at 95°C for 30 seconds, 65°C for 1 minutes, 72°C for 3 minutes, and a final extension at 72°C for 10 minutes.

Post-reaction cleanup and analysis

Four amplification products were pooled together and were processed through a Qiaquick PCR purification column (Qiagen) and eluted into 30µl EB (10mM Tris-Cl, pH 8.5). The product was analyzed on a Bioanalyzer (Agilent). A DNA 1000 kit was used according to manufacturers instructions.

Sequencing

Illumina sequencing

dsDNA library for Illumina sequencing using sample indexing was carried out according to manufacturers instructions. Sequencing was carried out on an HiSeq2000 platform (Illumina).

Bioinformatics

Obtaining digital transcriptomic information from sequencing data from whole transcriptome libraries amplified using the dA tailing terminal transferase approach

The sequencing data was sorted through the FastX toolkit FASTQ Barcode splitter tool into individual files for the respective capture probe positional domain

(tag) sequences. Individually tagged sequencing data was then analyzed through mapping to the mouse genome with the Tophat mapping tool. The resulting SAM file was processed for transcript counts through the HTseq-count software.

Obtaining digital transcriptomic information from sequencing data from whole transcriptome libraries amplified using the Picoplex whole genome amplification kit approach

The sequencing data was converted from FASTQ format to FASTA format using the FastX toolkit FASTQ-to-FASTA converter. The sequencing reads was aligned to the capture probe positional domain (tag) sequences using Blastn and the reads with hits better than $1e^{-6}$ to one of tag sequences were sorted out to individual files for each tag sequence respectively. The file of tag sequence reads was then aligned using Blastn to the mouse transcriptome, and hits were collected.

Combining visualization data and expression profiles

The expression profiles for individual capture probe positional domain (tag) sequences are combined with the spatial information obtained from the tissue sections through staining. Thereby the transcriptomic data from the cellular compartments of the tissue section can be analyzed in a directly comparative fashion, with the availability to distinguish distinct expression features for different cellular subtypes in a given structural context

Example 2

Figures 8 to 12 show successful visualisation of stained FFPE mouse brain tissue (olfactory bulb) sections on top of a bar-coded transcriptome capture array, according to the general procedure described in Example 1. As compared with the experiment with fresh frozen tissue in Example 1, Figure 8 shows better morphology with the FFPE tissue. Figures 9 and 10 show how tissue may be positioned on different types of probe density arrays.

Example 3

Whole Transcriptome Amplification by Random primer second strand synthesis followed by universal handle amplification (capture probe sequences including tag sequences retained at the end of the resulting dsDNA)

Following capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes)

OR

Following capture probe release with heated PCR buffer (hybridized *in situ* synthesized capture probes)

1μl RNase H (5U/μl) was added to each of two tubes, final concentration of 0.12U/μl, containing 40μl 1x Faststart HiFi PCR Buffer (pH 8.3) with 1.8mM $MgCl_2$ (Roche, www.roche-applied-science.com), 0.2mM of each dNTP (Fermentas, www.fermentas.com), 0.1μg/μl BSA (New England Biolabs, www.neb.com), 0.1U/μl USER Enzyme (New England Biolabs), released cDNA (extended from surface probes) and released surface probes. The tubes were incubated at 37°C for 30 min followed by 70°C for 20 min in a thermo cycler (Applied Biosystems, www.appliedbiosystems.com). 1μl Klenow Fragment (3' to 5' exo minus) (Illumina, www.illumina.com) and 1μl handle coupled random primer (10μM) (Eurofins MWG Operon, www.eurofinsdna.com) was added to the two tubes (B_R8 (octamer) to one of the tubes and B_R6 (hexamer) to the other tube), final concentration of 0.23μM. The two tubes were incubated at 15°C for 15 min, 25°C for 15 min, 37°C for 15 min and finally 75°C for 20 min in a thermo cycler (Applied Biosystems). After the incubation, 1μl of each primer, A_P and B (10μM) (Eurofins MWG Operon), was added to both tubes, final concentration of 0.22μM each. 1μl Faststart HiFi DNA polymerase (5U/μl) (Roche) was also added to both tubes, final concentration of 0.11U/μl. PCR amplification were carried out in a thermo cycler (Applied Biosystems) with the following program: Hot start at 94°C for 2 min, followed by 50 cycles at 94°C for 15 seconds, 55°C for 30 seconds, 68°C for 1 minute, and a final extension at 68°C for 5 minutes. After the amplification, 40μl from each of the two tubes were purified with Qiaquick PCR purification columns (Qiagen, www.qiagen.com) and eluted into 30μl EB (10mM Tris-Cl, pH 8.5). The Purified products were analyzed with a Bioanalyzer (Agilent, www.home.agilent.com), DNA 7500 kit were used. The results are shown in Figures 13 and 14.

This Example demonstrates the use of random hexamer and random octamer second strand synthesis, followed by amplification to generate the population from the released cDNA molecules.

Example 4

Amplification of ID-specific and gene specific products after cDNA synthesis and probe collection

Following capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes).

The cleaved cDNA was amplified in final reaction volumes of 10 µl. 7 µl cleaved template, 1 µl ID-specific forward primer (2 µM), 1 µl gene-specific reverse primer (2 µM) and 1 µl FastStart High Fidelity Enzyme Blend in 1.4x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ to give a final reaction of 10 µl with 1x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ and 1 U FastStart High Fidelity Enzyme Blend. PCR amplification were carried out in a thermo cycler (Applied Biosystems) with the following program: Hot start at 94°C for 2 min, followed by 50 cycles at 94°C for 15 seconds, 55°C for 30 seconds, 68°C for 1 minute, and a final extension at 68°C for 5 minutes.

Primer sequences, resulting in a product of approximately 250 bp,
Beta-2 microglobulin (B2M) primer
5'-TGGGGGTGAGAATTGCTAAG-3' (SEQ ID NO: 43)
ID-1 primer
5'-CCTTCTCCTTCTCCTTCACC-3' (SEQ ID NO: 44)
ID-5 primer
5'-GTCCTCTATTCCGTCACCAT-3' (SEQ ID NO: 45)
ID-20 primer
5'-CTGCTTCTTCCTGGAACTCA-3' (SEQ ID NO: 46)

The results are shown in Figure 15. This shows successful amplification of ID-specific and gene-specific products using two different ID primers (i.e. specific for ID tags positioned at different locations on the microarray and the same gene specific primer from a brain tissue covering all the probes. Accordingly this experiment establishes that products may be identified by an ID tag -specific or target nucleic acid specific amplification reaction. It is further established that different ID tags may be distinguished. A second experiment, with tissue covering only half of the ID probes (i.e. capture probes) on the array resulted in a positive result (PCR product) for spots that were covered with tissue.

Example 5

Spatial Genomics

Background. The method has as its purpose to capture DNA molecules from a tissue sample with retained spatial resolution, making it possible to determine from what part of the tissue a particular DNA fragment stems.

Method. The principle of the method is to use microarrays with immobilized DNA oligos (capture probes) carrying spatial labeling tag sequences (positional domains). Each feature of oligos of the microarray carries a 1) a unique labeling tag (positional domain) and 2) a capture sequence (capture domain). Keeping track of where which labeling tag is geographically placed on the array surface makes it possible to extract positional information in two dimensions from each labeling tag. Fragmented genomic DNA is added to the microarray, for instance through the addition of a thin section of FFPE treated tissue. The genomic DNA in this tissue section is pre-fragmented due to the fixation treatment.

Once the tissue slice has been placed on the array, a universal tailing reaction is carried out through the use of a terminal transferase enzyme. The tailing reaction adds polydA tails to the protruding 3' ends of the genomic DNA fragments in the tissue. The oligos on the surface are blocked from tailing by terminal transferase through a hybridized and 3' blocked polydA probe.

Following the terminal transferase tailing, the genomic DNA fragments are able to hybridize to the spatially tagged oligos in their vicinity through the polydA tail meeting the polydT capture sequence on the surface oligos. After hybridization is completed a strand displacing polymerase such as Klenow exo- can use the oligo on the surface as a primer for creation of a new DNA strand complementary to the hybridized genomic DNA fragment. The new DNA strand will now also contain the positional information of the surface oligo's labeling tag.

As a last step the newly generated labeled DNA strands are cleaved from the surface through either enzymatic means, denaturation or physical means. The strands are then collected and can be subjected to downstream amplification of the entire set of strands through introduction of universal handles, amplification of specific amplicons, and/or sequencing.

Figure 16 is a schematic illustration of this process.

Materials and methods

Preparation of in-house printed microarray with 5' to 3' oriented probes

20 DNA-capture oligos with individual tag sequences (Table 1) were spotted on glass slides to function as capture probes. The probes were synthesized with a 5'-terminus amino linker with a C6 spacer. All probes where synthesized by Sigma-Aldrich (St. Louis, MO, USA). The DNA-capture probes were suspended at a concentration of 20 μM in 150 mM sodium phosphate, pH 8.5 and were spotted using a Nanoplotter NP2.1/E (Gesim, Grosserkmannsdorf, Germany) onto CodeLink™ Activated microarray slides (7.5cm x 2.5cm; Surmodics, Eden Prairie, MN, USA). After printing, surface blocking was performed according to the manufacturer's instructions. The probes were printed in 16 identical arrays on the slide, and each array contained a pre-defined printing pattern. The 16 sub-arrays were separated during hybridization by a 16-pad mask (ChipClip™ Schleicher & Schuell BioScience, Keene, NH, USA).

Preparation of in-house printed microarray with 3' to 5' oriented probes and synthesis of 5' to 3' oriented capture probes

Printing of oligos was performed as in the case with 5' to 3' oriented probes above.

To hybridize primers for capture probe synthesis hybridization solution containing 4xSSC and 0.1% SDS, 2 μM extension primer (A_primer) and 2 μM thread joining primer (p_poly_dT) was incubated for 4 min at 50°C. Meanwhile the in-house array was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50 μL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry and placed back in the ChipClip.

For extension and ligation 50 μL of enzyme mix containing 10xAmpligase buffer, 2.5 U AmpliTaq DNA Polymerase Stoffel Fragment (Applied Biosystems), 10 U Ampligase (Epicentre Biotechnologies), dNTPs 2 mM each (Fermentas) and water, is pipetted to each well. The array is subsequently incubated at 55°C for 30 min. After incubation the array is washed according to previously described array washing method but the first step has the duration of 10 min instead of 6 min.

Hybridization of polydA probe for protection of surface oligo capture sequences from dA tailing

To hybridize a 3'-biotin blocked polydA probe for protection of the surface oligo capture sequences a hybridization solution containing 4xSSC and 0.1% SDS, 2 μM 3'bio-polydA was incubated for 4 min at 50°C. Meanwhile the in-house array was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50 μL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry and placed back in the ChipClip.

Preparation of formalin-fixed paraffin-embedded (FFPE) tissue

Mouse brain tissue was fixed in 4% formalin at 4°C for 24h. After that it was incubated as follows: 3x incubation in 70% ethanol for 1hour, 1x incubation in 80% ethanol for 1hour, 1x incubation in 96% ethanol for 1hour, 3x incubation in 100% ethanol for 1 hour, 2x incubation in xylene at room temperature for 1 h.

The dehydrated samples were then incubated in liquid low melting paraffin 52-54 °C for up to 3 hours, during which the paraffin in changed once to wash out residual xylene. Finished tissue blocks were then stored at RT. Sections were then cut at 4 μm in paraffin with a microtome onto each capture probe array to be used.

The sections are dried at 37°C on the array slides for 24 hours and store at RT.

Deparaffinization of FFPE tissue

Formalin fixed paraffinized mouse brain 10 μm sections attached to CodeLink slides were deparaffinised in xylene twice for 10 min, 99.5% ethanol for 2 min, 96% ethanol for 2 min, 70% ethanol for 2 min and were then air dried.

Universal tailing of genomic DNA

For dA tailing a 50 μl reaction mixture containing 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com), 0.1μg/μl BSA (New England Biolabs) , 1μl TdT (20U/μl) and 0.5μl dATPs (100mM) was prepared. The mixture was added to the

array surface and the array was incubated in a thermo cycler (Applied Biosystems) at 37°C for 15 min followed by an inactivation of TdT at 70°C for 10 min. After this the temperature was lowered to 50°C again to allow for hybridization of dA tailed genomic fragments to the surface oligo capture sequences.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry.

Extension of labeled DNA

A 50 µl reaction mixture containing 50 µl of a mixture containing 1x Klenow buffer, 200 µM dNTPs (New England Biolabs) and 1 µl Klenow Fragment (3' to 5' exo minus) and was heated to 37°C and was added to each well and incubated at 37°C for 30 min with mixing (3 s. 300 rpm, 6 s. rest) (Thermomixer comfort; Eppendorf).

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry.

Removal of residual tissue

The slides with attached formalin fixed paraffinized mouse brain tissue sections were attached to ChipClip slide holders and 16 well masks (Whatman). For each 150 µl Proteinase K Digest Buffer from the RNeasy FFPE kit (Qiagen) 10 µl Proteinase K Solution (Qiagen) was added. 50 µl of the final mixture was added to each well and the slide was incubated at 56°C for 30 min.

Capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes)

A 16 well mask and CodeLink slide was attached to the ChipClip holder (Whatman). 50 µl of a mixture containing 1x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ (Roche), 200 µM dNTPs (New England Biolabs) and 0.1U/1 µl USER Enzyme (New England Biolabs) was heated to 37°C and was added to each well and incubated at 37°C for 30 min with mixing (3 s. 300 rpm, 6 s. rest)

(Thermomixer comfort; Eppendorf). The reaction mixture containing the released cDNA and probes was then recovered from the wells with a pipette.

Amplification of ID-specific and gene specific products after synthesis of labelled DNA and probe collection

Following capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes).

The cleaved DNA was amplified in final reaction volumes of 10 µl. 7 µl cleaved template, 1 µl ID-specific forward primer (2 µM), 1 µl gene-specific reverse primer (2 µM) and 1 µl FastStart High Fidelity Enzyme Blend in 1.4x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ to give a final reaction of 10 µl with 1x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ and 1 U FastStart High Fidelity Enzyme Blend. PCR amplification were carried out in a thermo cycler (Applied Biosystems) with the following program: Hot start at 94°C for 2 min, followed by 50 cycles at 94°C for 15 seconds, 55°C for 30 seconds, 68°C for 1 minute, and a final extension at 68°C for 5 minutes.

Whole Genome Amplification by Random primer second strand synthesis followed by universal handle amplification (capture probe sequences including tag sequences retained at the end of the resulting dsDNA)

Following capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes).

A reaction mixture containing 40µl 1x Faststart HiFi PCR Buffer (pH 8.3) with 1.8mM $MgCl_2$ (Roche, www.roche-applied-science.com), 0.2mM of each dNTP (Fermentas, www.fermentas.com), 0.1µg/µl BSA (New England Biolabs, www.neb.com), 0.1U/µl USER Enzyme (New England Biolabs), released DNA (extended from surface probes) and released surface probes. The tubes were incubated at 37°C for 30 min followed by 70°C for 20 min in a thermo cycler (Applied Biosystems, www.appliedbiosystems.com). 1µl Klenow Fragment (3' to 5' exo minus) (Illumina, www.illumina.com) and 1µl handle coupled random primer (10µM) (Eurofins MWG Operon, www.eurofinsdna.com) was added to the tube. The tube was incubated at 15°C for 15 min, 25°C for 15 min, 37°C for 15 min and finally 75°C for 20 min in a thermo cycler (Applied Biosystems). After the incubation, 1µl of each primer, A_P and B (10µM) (Eurofins MWG Operon), was added to the tube.

1μl Faststart HiFi DNA polymerase (5U/μl) (Roche) was also added to the tube. PCR amplification were carried out in a thermo cycler (Applied Biosystems) with the following program: Hot start at 94°C for 2 min, followed by 50 cycles at 94°C for 15 seconds, 55°C for 30 seconds, 68°C for 1 minute, and a final extension at 68°C for 5 minutes. After the amplification, 40μl from the tube was purified with Qiaquick PCR purification columns (Qiagen, www.qiagen.com) and eluted into 30μl EB (10mM Tris-Cl, pH 8.5). The Purified product was analyzed with a Bioanalyzer (Agilent, www.home.agilent.com), DNA 7500 kit were used.

Visualization

Hybridization of fluorescent marker probes prior to staining

Prior to tissue application fluorescent marker probes are hybridized to designated marker sequences printed on the capture probe array. The fluorescent marker probes aid in the orientation of the resulting image after tissue visualization, making it possible to combine the image with the resulting expression profiles for individual capture probe tag sequences obtained after sequencing. To hybridize fluorescent probes a hybridization solution containing 4xSSC and 0.1% SDS, 2 μM detection probe (P) was incubated for 4 min at 50°C. Meanwhile the in-house array was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50 μL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry.

General histological staining of FFPE tissue sections prior to or post synthesis of labeled DNA

FFPE tissue sections immobilized on capture probe arrays are washed and rehydrated after deparaffinization prior to synthesis of labeled as described previously, or washed after synthesis of labeled DNA as described previously. They are then treated as follows: incubate for 3 minutes in Hematoxylin, rinse with deionized water, incubate 5 minutes in tap water, rapidly dip 8 to 12 times in acid ethanol, rinse 2x1 minute in tap water, rinse 2 minutes in deionized water, incubate 30 seconds in Eosin, wash 3x5 minutes in 95% ethanol, wash 3x5 minutes in 100%

ethanol, wash 3x10 minutes in xylene (can be done overnight), place coverslip on slides using DPX, dry slides in the hood overnight.

General immunohistochemistry staining of a target protein in FFPE tissue sections prior to or post synthesis of labeled DNA

FFPE tissue sections immobilized on capture probe arrays are washed and rehydrated after deparaffinization prior to synthesis of labeled DNA as described previously, or washed after synthesis of labeled DNA as described previously. They are then treated as follows without being let to dry during the whole staining process: Dilute primary antibody in blocking solution (1xTBS (Tris Buffered Saline (50mM Tris, 150mM NaCl, pH 7.6),4% donkey serum, 0.1% triton-x), incubate sections with primary antibody in a wet chamber overnight at RT, rinse 3x with 1xTBS, incubate section with matching secondary antibody conjugated to a fluorochrome (FITC, Cy3 or Cy5) in a wet chamber at RT for 1h, Rinse 3x with 1xTBS, remove as much as possible of TBS and mount section with ProLong Gold +DAPI (Invitrogen) and analyze with fluorescence microscope and matching filter sets.

Example 6

This experiment was conducted following the principles of Example 5, but using fragmented genomic DNA on the array rather than tissue. The genomic DNA was pre-fragmented to a mean size of 200bp and 700bp respectively. This experiment shows that the principle works. Fragmented genomic DNA is very similar to FFPE tissue.

Amplification of internal gene specific products after synthesis of labelled DNA and probe collection

Following capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes) containing 1x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ (Roche), 200 µM dNTPs (New England Biolabs) and 0.1U/1 µl USER Enzyme (New England Biolabs).

The cleaved DNA was amplified in a final reaction volume of 50 µl. To 47 µl cleaved template was added 1 µl ID-specific forward primer (10 µM), 1 µl gene-specific reverse primer (10 µM) and 1 µl FastStart High Fidelity Enzyme Blend. PCR amplification were carried out in a thermo cycler (Applied Biosystems) with the

following program: Hot start at 94°C for 2 min, followed by 50 cycles at 94°C for 15 seconds, 55°C for 30 seconds, 68°C for 1 minute, and a final extension at 68°C for 5 minutes.

Amplification of label-specific and gene specific products after synthesis of labelled DNA and probe collection

Following capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes) containing 1x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ (Roche), 200 µM dNTPs (New England Biolabs) and 0.1U/1 µl USER Enzyme (New England Biolabs).

The cleaved DNA was amplified in a final reaction volume of 50 µl. To 47 µl cleaved template was added 1 µl label-specific forward primer (10 µM), 1 µl gene-specific reverse primer (10 µM) and 1 µl FastStart High Fidelity Enzyme Blend. PCR amplification were carried out in a thermo cycler (Applied Biosystems) with the following program: Hot start at 94°C for 2 min, followed by 50 cycles at 94°C for 15 seconds, 55°C for 30 seconds, 68°C for 1 minute, and a final extension at 68°C for 5 minutes.

Forward - Genomic DNA Human Primer
5'- GACTGCTCTTTTCACCCATC-3' (SEQ ID NO: 47)
Reverse - Genomic DNA Human Primer
5'-GGAGCTGCTGGTGCAGGG-3' (SEQ ID NO: 48)
P - label specific primer
5'- ATCTCGACTGCCACTCTGAA-3' (SEQ ID NO: 49)

The results are shown in Figures 17 to 20. The Figures show internal products amplified on the array - the detected peaks in Figures 17 and 18 are of the expected size. This thus demonstrates that genomic DNA may be captured and amplified. In Figures 19 and 20, the expected product is a smear given that the random fragmentation and terminal transferase labeling of genomic DNA will generate a very diverse sample pool.

Example 7

Alternative synthesis of 5' to 3' oriented capture probes using polymerase extension and terminal transferase tailing

To hybridize primers for capture probe synthesis hybridization solution containing 4xSSC and 0.1% SDS and 2 μM extension primer (A_primer) was incubated for 4 min at 50°C. Meanwhile the in-house array (see Example 1) was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50 μL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry and placed back in the ChipClip.

1μl Klenow Fragment (3' to 5' exo minus) (Illumina, www.illumina.com) together with 10x Klenow buffer, dNTPs 2 mM each (Fermentas) and water, was mixed into a 50 μl reaction and was pipetted into each well.

The array was incubated at 15°C for 15 min, 25°C for 15 min, 37°C for 15 min and finally 75°C for 20 min in an Eppendorf Thermomixer.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry and placed back in the ChipClip.

For dT tailing a 50 μl reaction mixture containing 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com), 0.1μg/μl BSA (New England Biolabs), 0.5μl RNase H (5U/μl) , 1μl TdT (20U/μl) and 0.5μl dTTPs (100mM) was prepared. The mixture was added to the array surface and the array was incubated in a thermo cycler (Applied Biosystems) at 37°C for 15 min followed by an inactivation of TdT at 70°C for 10 min.

Example 8

Spatial transcriptomics using 5' to 3' high probe density arrays and formalin-fixed frozen (FF-frozen) tissue with USER system cleavage and amplification via terminal transferase

Array preparation

Pre-fabricated high-density microarrays chips were ordered from Roche-Nimblegen (Madison, WI, USA). Each capture probe array contained 135,000

features of which 132,640 features carried a capture probe comprising a unique ID-tag sequence (positional domain) and a capture region (capture domain). Each feature was 13x13 µm in size. The capture probes were composed 5' to 3' of a universal domain containing five dUTP bases (a cleavage domain) and a general amplification domain, an ID tag (positional domain) and a capture region (capture domain) (Figure 22 and Table 2). Each array was also fitted with a frame of marker probes (Figure 23) carrying a generic 30 bp sequence (Table 2) to enable hybridization of fluorescent probes to help with orientation during array visualization.

Tissue preparation – preparation of formalin-fixed frozen tissue

The animal (mouse) was perfused with 50ml PBS and 100ml 4% formalin solution. After excision of the olfactory bulb, the tissue was put into a 4% formalin bath for post-fixation for 24 hrs. The tissue was then sucrose treated in 30% sucrose dissolved in PBS for 24 hrs to stabilize morphology and to remove excess formalin. The tissue was frozen at a controlled rate down to -40°C and kept at -20°C between experiments. Similar preparation of tissue postfixed for 3 hrs or without post-fixation was carried out for a parallel specimen. Perfusion with 2% formalin without post-fixation was also used successfully. Similarly the sucrose treatment step could be omitted. The tissue was mounted into a cryostat for sectioning at 10µm. A slice of tissue was applied onto each capture probe array to be used. Optionally for better tissue adherence, the array chip was placed at 50°C for 15 minutes.

Optional control - Total RNA preparation from sectioned tissue

Total RNA was extracted from a single tissue section (10µm) using the RNeasy FFPE kit (Qiagen) according to manufacturers instructions. The total RNA obtained from the tissue section was used in control experiments for a comparison with experiments in which the RNA was captured on the array directly from the tissue section. Accordingly, in the case where totalRNA was applied to the array the staining, visualization and degradation of tissue steps were omitted.

On-chip reactions

The hybridization of marker probe to the frame probes, reverse transcription, nuclear staining, tissue digestion and probe cleavage reactions were all performed

in a 16 well silicone gasket (ArrayIt, Sunnyvale, CA, USA) with a reaction volume of 50 µl per well. To prevent evaporation, the cassettes were covered with plate sealers (In Vitro AB, Stockholm, Sweden).

Optional - tissue permeabilization prior to cDNA synthesis

For permeabilization using Proteinase K, proteinase K (Qiagen, Hilden, Germany) was diluted to 1 µg/ml in PBS. The solution was added to the wells and the slide incubated at room temperature for 5 minutes, followed by a gradual increase to 80°C over 10 minutes. The slide was washed briefly in PBS before the reverse transcription reaction.

Alternatively for permeabilization using microwaves, after tissue attachment, the slide was placed at the bottom of a glass jar containing 50ml 0.2xSSC (Sigma-Aldrich) and was heated in a microwave oven for 1 minute at 800W. Directly after microwave treatment the slide was placed onto a paper tissue and was dried for 30 minutes in a chamber protected from unnecessary air exposure. After drying, the slide was briefly dipped in water (RNase/DNase free) and finally spin-dried by a centrifuge before cDNA synthesis was initiated.

cDNA synthesis

For the reverse transcription reaction the SuperScript III One-Step RT-PCR System with Platinum Taq (Life Technologies/Invitrogen, Carlsbad, CA, USA) was used. Reverse transcription reactions contained 1x reaction mix, 1x BSA (New England Biolabs, Ipswich, MA, USA) and 2 µl SuperScript III RT/Platinum Taq mix in a final volume of 50 µl. This solution was heated to 50°C before application to the tissue sections and the reaction was performed at 50°C for 30 minutes. The reverse transcription solution was subsequently removed from the wells and the slide was allowed to air dry for 2 hours.

Tissue visualization

After cDNA synthesis, nuclear staining and hybridization of the marker probe to the frame probes (probes attached to the array substrate to enable orientation of the tissue sample on the array) was done simultaneously. A solution with DAPI at a concentration of 300nM and marker probe at a concentration of 170 nM in PBS was prepared. This solution was added to the wells and the slide was

incubated at room temperature for 5 minutes, followed by brief washing in PBS and spin drying.

Alternatively the marker probe was hybridized to the frame probes prior to placing the tissue on the array. The marker probe was then diluted to 170 nM in hybridization buffer (4xSSC, 0.1% SDS). This solution was heated to 50°C before application to the chip and the hybridization was performed at 50°C for 30 minutes at 300 rpm. After hybridization, the slide was washed in 2xSSC, 0.1% SDS at 50°C and 300 rpm for 10 minutes, 0.2xSSC at 300 rpm for 1 minute and 0.1xSSC at 300 rpm for 1 minute. In that case the staining solution after cDNA synthesis only contained the nuclear DAPI stain diluted to 300nM in PBS. The solution was applied to the wells and the slide was incubated at room temperature for 5 minutes, followed by brief washing in PBS and spin drying.

The sections were microscopically examined with a Zeiss Axio Imager Z2 and processed with MetaSystems software.

Tissue removal

The tissue sections were digested using Proteinase K diluted to 1.25 µg/µl in PKD buffer from the RNeasy FFPE Kit (both from Qiagen) at 56°C for 30 minutes with an interval mix at 300 rpm for 3 seconds, then 6 seconds rest. The slide was subsequently washed in 2xSSC, 0.1% SDS at 50°C and 300 rpm for 10 minutes, 0.2xSSC at 300 rpm for 1 minute and 0.1xSSC at 300 rpm for 1 minute.

Probe release

The 16-well Hybridization Cassette with silicone gasket (ArrayIt) was preheated to 37°C and attached to the Nimblegen slide. A volume of 50µl of cleavage mixture preheated to 37°C, consisting of Lysis buffer at an unknown concentration (Takara), 0.1U/µl USER Enzyme (NEB) and 0.1µg/µl BSA was added to each of wells containing surface immobilized cDNA. After removal of bubbles the slide was sealed and incubated at 37°C for 30 minutes in a Thermomixer comfort with cycled shaking at 300 rpm for 3 seconds with 6 seconds rest in between. After the incubation 45µl cleavage mixture was collected from each of the used wells and placed into 0.2ml PCR tubes (Figure 24).

Library preparation

Exonuclease treatment

After cooling the solutions on ice for 2 minutes, Exonuclease I (NEB) was added, to remove unextended cDNA probes, to a final volume of 46.2μl and a final concentration of 0.52U/μl. The tubes were incubated in a thermo cycler (Applied Biosystems) at 37°C for 30 minutes followed by inactivation of the exonuclease at 80°C for 25 minutes.

dA-tailing by terminal transferase

After the exonuclease step, 45μl polyA-tailing mixture, according to manufacturers instructions consisting of TdT Buffer (Takara), 3mM dATP (Takara) and manufacturers TdT Enzyme mix (TdT and RNase H) (Takara), was added to each of the samples. The mixtures were incubated in a thermocycler at 37°C for 15 minutes followed by inactivation of TdT at 70°C for 10 minutes.

Second-strand synthesis and PCR-amplification

After dA-tailing, 23μl PCR master mix was placed into four new 0.2ml PCR tubes per sample, to each tube 2μl sample was added as a template. The final PCRs consisted of 1x Ex Taq buffer (Takara), 200μM of each dNTP (Takara), 600nM A_primer (MWG), 600nM B_dT20VN_primer (MWG) and 0.025U/μl Ex Taq polymerase (Takara)(Table 2). A second cDNA strand was created by running one cycle in a thermocycler at 95°C for 3 minutes, 50°C for 2 minutes and 72°C for 3 minutes. Then the samples were amplified by running 20 cycles (for library preparation) or 30 cycles (to confirm the presence of cDNA) at 95°C for 30 seconds, 67°C for 1 minute and 72°C for 3 minutes, followed by a final extension at 72°C for 10 minutes.

Library cleanup

After amplification, the four PCRs (100μl) were mixed with 500μl binding buffer (Qiagen) and placed in a Qiaquick PCR purification column (Qiagen) and spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane. The membrane was then washed with wash buffer (Qiagen) containing ethanol and finally eluted into 50μl of 10mM Tris-Cl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification (purification by superparamagnetic beads conjugated to carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The amplified cDNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and were then eluted into 15μl of 10mM Tris-Cl, pH 8.5.

Library Quality analysis

Samples amplified for 30 cycles were analyzed with an Agilent Bioanalyzer (Agilent) in order to confirm the presence of an amplified cDNA library, the DNA High Sensitivity kit or DNA 1000 kit were used depending on the amount of material.

Sequencing library preparation

Library indexing

Samples amplified for 20 cycles were used further to prepare sequencing libraries. An index PCR master mix was prepared for each sample and 23μl was placed into six 0.2ml tubes. 2μl of the amplified and purified cDNA was added to each of the six PCRs as template making the PCRs containing 1x Phusion master mix (Fermentas), 500nM InPE1.0 (Illumina), 500nM Index 1-12 (Illumina), and 0.4nM InPE2.0 (Illumina). The samples were amplified in a thermocycler for 18 cycles at 98°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 minute, followed by a final extension at 72°C for 5 minutes.

Sequencing library cleanup

After amplification, the six PCRs (150μl) were mixed with 750μl binding buffer and placed in a Qiaquick PCR purification column and spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane (because of the large sample volume (900μl), the sample was split in two (each 450μl) and was bound in two separate steps). The membrane was then washed with wash buffer containing ethanol and finally eluted into 50μl of 10mM Tris-Cl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification with an MBS robot. A final PEG concentration of 7.8% was used in order to remove fragments below 300-350bp. The amplified cDNA was allowed to

bind to the CA-beads for 10 min and were then eluted into 15µl of 10mM Tris-Cl, pH 8.5. Samples were analyzed with an Agilent Bioanalyzer in order to confirm the presence and size of the finished libraries, the DNA High Sensitivity kit or DNA 1000 kit were used according to manufacturers instructions depending on the amount of material (Figure 25).

Sequencing

The libraries were sequenced on the Illumina Hiseq2000 or Miseq depending on desired data throughput according to manufacturers instructions. Optionally for read 2, a custom sequencing primer B_r2 was used to avoid sequencing through the homopolymeric stretch of 20T.

Data analysis

Read 1 was trimmed 42 bases at 5' end. Read 2 was trimmed 25 bases at 5' end (optionally no bases were trimmed from read 2 if the custom primer was used). The reads were then mapped with bowtie to the repeat masked *Mus musculus* 9 genome assembly and the output was formatted in the SAM file format. Mapped reads were extracted and annotated with UCSC refGene gene annotations. Indexes were retrieved with 'indexFinder' (an inhouse software for index retrieval). A mongo DB database was then created containing information about all caught transcripts and their respective index position on the chip.

A matlab implementation was connected to the database and allowed for spatial visualization and analysis of the data (Figure 26).

Optionally the data visualization was overlaid with the microscopic image using the fluorescently labelled frame probes for exact alignment and enabling spatial transcriptomic data extraction.

Example 9

Spatial transcriptomics using 3' to 5' high probe density arrays and FFPE tissue with MutY system cleavage and amplification via TdT

Array preparation

Pre-fabricated high-density microarrays chips were ordered from Roche-Nimblegen (Madison, WI, USA). Each used capture probe array contained 72k features out of which 66,022 contained a unique ID-tag complementary sequence.

Each feature was 16x16 µm in size. The capture probes were composed 3' to 5' in the same way as the probes used for the in-house printed 3' to 5' arrays with the exeception to 3 additional bases being added to the upper (P') general handle of the probe to make it a long version of P', LP' (Table 2). Each array was also fitted with a frame of probes carrying a generic 30 bp sequence to enable hybridization of fluorescent probes to help with orientation during array visualization.

Synthesis of 5' to 3' oriented capture probes

The synthesis of 5' to 3' oriented capture probes on the high-density arrays was carried out as in the case with in-house printed arrays, with the exception that the extension and ligation steps were carried out at 55°C for 15 mins followed by 72°C for 15 mins. The A-handle probe (Table 2) included an A/G mismatch to allow for subsequent release of probes through the MutY enzymatic system described below. The P-probe was replaced by a longer LP version to match the longer probes on the surface.

Preparation of formalin-fixed paraffin-embedded tissue and deparaffinization

This was carried out as described above in the in-house protocol.

cDNA synthesis and staining

cDNA synthesis and staining was carried out as in the protocol for 5' to 3' oriented high-density Nimblegen arrays with the exception that biotin labeled dCTPs and dATPs were added to the cDNA synthesis together with the four regular dNTPs (each was present at 25x times more than the biotin labeled ones).

Tissue removal

Tissue removal was carried out in the same way as in the protocol for 5' to 3' oriented high-density Nimblegen arrays described in Example 8.

Probe cleavage by MutY

A 16-well Incubation chamber with silicone gasket (ArrayIT) was preheated to 37°C and attached to the Codelink slide. A volume of 50µl of cleavage mixture preheated to 37°C, consisting of 1x Endonuclease VIII Buffer (NEB), 10U/µl MutY (Trevigen), 10U/µl Endonuclease VIII (NEB), 0.1µg/µl BSA was added to each of wells containing surface immobilized cDNA. After removal of bubbles the slide was

sealed and incubated at 37°C for 30 minutes in a Thermomixer comfort with cycled shaking at 300rpm for 3 seconds with 6 seconds rest in between. After the incubation, the plate sealer was removed and 40µl cleavage mixture was collected from each of the used wells and placed into a PCR plate.

Library preparation

Biotin-streptavidin mediated library cleanup

To remove unextended cDNA probes and to change buffer, the samples were purified by binding the biotin labeled cDNA to streptavidin coated C1-beads (Invitrogen) and washing the beads with 0.1M NaOH (made fresh). The purification was carried out with an MBS robot (Magnetic Biosolutions), the biotin labelled cDNA was allowed to bind to the C1-beads for 10 min and was then eluted into 20µl of water by heating the bead-water solution to 80°C to break the biotin-streptavidin binding.

dA-tailing by terminal transferase

After the purification step, 18µl of each sample was placed into new 0.2ml PCR tubes and mixed with 22µl of a polyA-tailing master mix leading to a 40µl reaction mixture according to manufacturers instructions consisting of lysis buffer (Takara, Cellamp Whole Transcriptome Amplification kit), TdT Buffer (Takara), 1.5mM dATP (Takara) and TdT Enzyme mix (TdT and RNase H) (Takara). The mixtures were incubated in a thermocycler at 37°C for 15 minutes followed by inactivation of TdT at 70°C for 10 minutes.

Second-strand synthesis and PCR-amplification

After dA-tailing, 23µl PCR master mix was placed into four new 0.2ml PCR tubes per sample, to each tube 2µl sample was added as a template. The final PCRs consisted of 1x Ex Taq buffer (Takara), 200µM of each dNTP (Takara), 600nM A_primer (MWG), 600nM B_dT20VN_primer (MWG) and 0.025U/µl Ex Taq polymerase (Takara). A second cDNA strand was created by running one cycle in a thermo cycler at 95°C for 3 minutes, 50°C for 2 minutes and 72°C for 3 minutes. Then the samples were amplified by running 20 cycles (for library preparation) or 30 cycles (to confirm the presence of cDNA) at 95°C for 30 seconds, 67°C for 1 minute and 72°C for 3 minutes, followed by a final extension at 72°C for 10 minutes.

Library cleanup

After amplification, the four PCRs (100µl) were mixed with 500µl binding buffer (Qiagen) and placed in a Qiaquick PCR purification column (Qiagen) and spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane. The membrane was then washed with wash buffer (Qiagen) containing ethanol and finally eluted into 50µl of 10mM Tris-HCl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification (purification by superparamagnetic beads conjugated to carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The amplified cDNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and were then eluted into 15µl of 10mM Tris-HCl, pH 8.5.

Second PCR-amplification

The final PCRs consisted of 1x Ex Taq buffer (Takara), 200µM of each dNTP (Takara), 600nM A_primer (MWG), 600nM B_ primer (MWG) and 0.025U/µl Ex Taq polymerase (Takara). The samples were heated to 95°C for 3 minutes, and then amplified by running 10 cycles at 95°C for 30 seconds, 65°C for 1 minute and 72°C for 3 minutes, followed by a final extension at 72°C for 10 minutes.

Second library cleanup

After amplification, the four PCRs (100µl) were mixed with 500µl binding buffer (Qiagen) and placed in a Qiaquick PCR purification column (Qiagen) and spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane. The membrane was then washed with wash buffer (Qiagen) containing ethanol and finally eluted into 50µl of 10mM Tris-Cl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification (purification by super-paramagnetic beads conjugated to carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The amplified cDNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and were then eluted into 15µl of 10mM Tris-HCl, pH 8.5.

Sequencing library preparation

Library indexing

Samples amplified for 20 cycles were used further to prepare sequencing libraries. An index PCR master mix was prepared for each sample and 23μl was placed into six 0.2ml tubes. 2μl of the amplified and purified cDNA was added to each of the six PCRs as template making the PCRs containing 1x Phusion master mix (Fermentas), 500nM InPE1.0 (Illumina), 500nM Index 1-12 (Illumina), and 0.4nM InPE2.0 (Illumina). The samples were amplified in a thermo cycler for 18 cycles at 98°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 minute, followed by a final extension at 72°C for 5 minutes.

Sequencing library cleanup

After amplification, the samples was purified and concentrated by CA-purification with an MBS robot. A final PEG concentration of 7.8% was used in order to remove fragments below 300-350bp. The amplified cDNA was allowed to bind to the CA-beads for 10 min and were then eluted into 15μl of 10mM Tris-HCl, pH 8.5.

10μl of the amplified and purified samples were placed on a Caliper XT chip and fragments between 480bp and 720bp were cut out with the Caliper XT (Caliper). Samples were analyzed with an Agilent Bioanalyzer in order to confirm the presence and size of the finished libraries, the DNA High Sensitivity kit was used.

Sequencing and Data analysis

Sequencing and Bioinformatic was carried out in the same way as in the protocol for 5' to 3' oriented high-density Nimblegen arrays described in Example 8. However, in the data analysis, read 1 was not used in the mapping of transcripts. Specific Olfr transcripts could be sorted out using the Matlab visualization tool (Figure 27).

Example 10

Spatial transcriptomics using in house printed 41-tag microarray with 5' to 3' oriented probes and formalin-fixed frozen (FF-frozen) tissue with permeabilization

through ProteinaseK or microwaving with USER system cleavage and amplification via TdT

Array preparation

In-house arrays were printed as previously described but with a pattern of 41 unique ID-tag probes with the same composition as the probes in the 5' to 3' oriented high-density array in Example 8 (Figure 28).

All other steps were carried out in the same way as in the protocol described in Example 8.

Example 11

Alternative method for performing the cDNA synthesis step

cDNA synthesis on chip as described above can also be combined with template switching to create a second strand by adding a template switching primer to the cDNA synthesis reaction (Table 2). The second amplification domain is introduced by coupling it to terminal bases added by the reverse transcriptase at the 3' end of the first cDNA strand, and primes the synthesis of the second strand. The library can be readily amplified directly after release of the double-stranded complex from the array surface.

Example 12

Spatial genomics using in house printed 41-tag microarray with 5' to 3' oriented probes and fragmented poly-A tailed gDNA with USER system cleavage and amplification via TdT –tailing or translocation specific primers

Array preparation

In-house arrays were printed using Codelink slides (Surmodics) as previously described but with a pattern of 41 unique ID-tag probes with the same composition as the probes in the 5' to 3' oriented high-density in Example 8.

Total DNA preparation from cells

DNA fragmentation

Genomic DNA (gDNA) was extracted by DNeasy kit (Qiagen) according to the manufacturer's instructions from A431 and U2OS cell lines. The DNA was

fragmented to 500 bp on a Covaris sonicator (Covaris) according to manufacturer's instructions.

The sample was purified and concentrated by CA-purification (purification by super-paramagnetic beads conjugated to carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The fragmented DNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and were then eluted into 15μl of 10mM Tris-HCl, pH 8.5.

Optional control – spiking of different cell lines

Through spiking of A431 DNA into U2OS DNA different levels of capture sensitivity can be measured, such as from spiking of 1%, 10% or 50% of A431 DNA.

dA-tailing by terminal transferase

A 45μl polyA-tailing mixture, according to manufacturer's instructions consisting of TdT Buffer (Takara), 3mM dATP (Takara) and TdT Enzyme mix (TdT and RNase H) (Takara), was added to 0.5μg of fragmented DNA. The mixtures were incubated in a thermocycler at 37°C for 30 minutes followed by inactivation of TdT at 80°C for 20 minutes. The dA-tailed fragments were then cleaned through a Qiaquick (Qiagen) column according to manufacturer's instructions and the concentration was measured using the Qubit system (Invitrogen) according to manufacturer's instructions.

On-chip experiments

The hybridization, second strand synthesis and cleavage reactions were performed on chip in a 16 well silicone gasket (ArrayIt, Sunnyvale, CA, USA). To prevent evaporation, the cassettes were covered with plate sealers (In Vitro AB, Stockholm, Sweden).

Hybridization

117 ng of DNA was deposited onto a well on a prewarmed array (50°C) in a total volume of 45 μl consisting of 1x NEB buffer (New England Biolabs) and 1x

BSA. The mixture was incubated for 30mins at 50°C in a Thermomixer Comfort (Eppendorf) fitted with an MTP block at 300 rpm shake.

Second strand synthesis

Without removing the hybridization mixture, 15µl of a Klenow extension reaction mixture consisting of 1x NEB buffer 1.5µl Klenow polymerase, and 3.75µl dNTPs (2mM each) was added to the well. The reaction mixture was incubated in a Thermomixer Comport (Eppendorf) 37°C for 30 mins without shaking.

The slide was subsequently washed in 2xSSC, 0.1% SDS at 50°C and 300 rpm for 10 minutes, 0.2xSSC at 300 rpm for 1 minute and 0.1xSSC at 300 rpm for 1 minute.

Probe release

A volume of 50µl of a mixture containing 1x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ (Roche), 200 µM dNTPs (New England Biolabs), 1x BSA and 0.1U/1 µl USER Enzyme (New England Biolabs) was heated to 37°C and was added to each well and incubated at 37°C for 30 min with mixing (3 seconds at 300 rpm, 6 seconds at rest) (Thermomixer comfort; Eppendorf). The reaction mixture containing the released DNA which was then recovered from the wells with a pipette.

Library preparation

Amplification reaction

Amplification was carried out in 10µl reactions consisting of 7.5 µl released sample, 1µl of each primer and 0.5µl enzyme (Roche, FastStart HiFi PCR system). The reaction was cycled as 94°C for 2 mins, one cycle of 94°C 15 sec, 55°C for 2mins, 72°C for 2mins, 30 cycles of 94°C for 15 secs, 65°C for 30 secs, 72°C for 90 secs, and a final elongation at 72°C for 5 mins.

In the preparation of a library for sequencing the two primers consisted of the surface probe A-handle and either of a specific translocation primer (for A431) or a specific SNP primer coupled to the B-handle (Table 2).

Library cleanup

The purified and concentrated sample was further purified and concentrated by CA-purification (purification by superparamagnetic beads conjugated to

carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The amplified DNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and was then eluted into 15μl of 10mM Tris-HCl, pH 8.5.

Library Quality analysis

Samples were analyzed with an Agilent Bioanalyzer (Agilent) in order to confirm the presence of an amplified DNA library, the DNA High Sensitivity kit or DNA 1000 kit were used depending on the amount of material.

Library indexing

Samples amplified for 20 cycles were used further to prepare sequencing libraries. An index PCR master mix was prepared for each sample and 23μl was placed into six 0.2ml tubes. 2μl of the amplified and purified cDNA was added to each of the six PCRs as template making the PCRs containing 1x Phusion master mix (Fermentas), 500nM InPE1.0 (Illumina), 500nM Index 1-12 (Illumina), and 0.4nM InPE2.0 (Illumina). The samples were amplified in a thermo cycler for 18 cycles at 98°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 minute, followed by a final extension at 72°C for 5 minutes.

Sequencing library cleanup

The purified and concentrated sample was further purified and concentrated by CA-purification with an MBS robot. A final PEG concentration of 7.8% was used in order to remove fragments below 300-350bp. The amplified DNA was allowed to bind to the CA-beads for 10 min and were then eluted into 15μl of 10mM Tris-Cl, pH 8.5. Samples were analyzed with an Agilent Bioanalyzer in order to confirm the presence and size of the finished libraries, the DNA High Sensitivity kit or DNA 1000 kit were used according to manufacturers instructions depending on the amount of material (Figure 29).

Sequencing

Sequencing was carried out in the same way as in the protocol for 5' to 3' oriented high-density Nimblegen arrays described in Example 8.

Data analysis

Data analysis was carried out to determine the sensitivity of capture of the arrayed ID-capture probes. Read 2 was sorted based on its content of either of the translocation or SNP primers. These reads were then sorted per their ID contained in Read 1.

Optional Control – Direct amplification of cell-line specific translocations

This was used to measure the capture sensitivity of spiked cell lines directly by PCR. The forward and reverse primers (Table 2) for the A431 translocations were used to try and detect the presence of the translocation in the second strand copied and released material (Figure 30).

Table 2. Oligos used for spatial transcriptomics and spatial genomics

Example 8

Nimblegen 5' to 3' arrays with free 3' end Array probes

	5' to 3'
Probe1 (SEQ ID NO: 50)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCCGATATGATTGCCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe2 (SEQ ID NO: 51)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGAGCCGGGTTCATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe3 (SEQ ID NO: 52)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTGAGGCACTCTGTTGGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe4 (SEQ ID NO: 53)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGATTAGTCGCCATTCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe5 (SEQ ID NO: 54)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTGAGGGTAGATGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe6 (SEQ ID NO: 55)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGGCCAATACTGTTATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe7 (SEQ ID NO: 56)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTACCCTGATTCGACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe8 (SEQ ID NO: 57)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCCACTTTCGCCGTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe9 (SEQ ID NO: 58)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCAACTTTGAGCAAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe10 (SEQ ID NO: 59)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCAATTCGGAATTCCGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe11 (SEQ ID NO: 60)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCCCAAGGTAATACATTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe12 (SEQ ID NO: 61)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCATTTCCTATTCGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe13 (SEQ ID NO: 62)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGCTAAATCTAACCGCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe14 (SEQ ID NO: 63)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAATTAAATTCTGATGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe15 (SEQ ID NO: 64)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCATTACATAGGTGCTAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe16 (SEQ ID NO: 65)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGACTTGCGCTCGCACTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe17 (SEQ ID NO: 66)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGTATCTCCCAAGTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe18 (SEQ ID NO: 67)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGCGCCTGTAATCCGCATTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe19 (SEQ ID NO: 68)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGCCACTCTTTAGGTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe20 (SEQ ID NO: 69)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTATGCAAGTGATTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe21 (SEQ ID NO: 70)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAAGCCACGTTTATACGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe22 (SEQ ID NO: 71)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTACCTGATTGCTGTATAACTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe23 (SEQ ID NO: 72)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGCGCATCTATCCTCTATTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe24 (SEQ ID NO: 73)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCACGCGTAGGACTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe25 (SEQ ID NO: 74)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCGACTAAGTATGTAGCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTVN

Frame probe
Layout1 (SEQ ID NO: 75) AAATTTCGTCTGCTATCGCGCTTCTGTACC

Fluorescent marker probe
PS_1 (SEQ ID NO: 76) GGTACAGAAGCGCGATAGCAG – Cy3

Second strand synthesis and first PCR Amplification handles
A_primer (SEQ ID NO: 77) ACACTCTTTCCCTACACGACGCTCTTCCGATCT
B_dt20VN_primer
(SEQ ID NO: 78) AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTVN

Custom sequencing primer
B_r2 (SEQ ID NO: 79) TCA GAC GTG TGC TCT TCC GAT CTT TTT TTT TTT TTT TTT T

Example 9

Nimblegen 3' to 5' arrays with free 5' end Array probes

	5' to 3'
Probe1 (SEQ ID NO: 80)	GCGTTCAGAGTGGCAGTCGAGATCACGCGGCAATCATATCGGACAGATCGGAAGAGCGTAGTGTAG
Probe2 (SEQ ID NO: 81)	GCGTTCAGAGTGGCAGTCGAGATCACAAGATGAACCCGGCTCATAGATCGGAAGAGCGTAGTGTAG
Probe3 (SEQ ID NO: 82)	GCGTTCAGAGTGGCAGTCGAGATCACTCCCAACAGAGTGCCTCAAGATCGGAAGAGCGTAGTGTAG
Probe4 (SEQ ID NO: 83)	GCGTTCAGAGTGGCAGTCGAGATCACCGAATGGCGACTAATCATAGATCGGAAGAGCGTAGTGTAG
Probe5 (SEQ ID NO: 84)	GCGTTCAGAGTGGCAGTCGAGATCACAAACATCTACCCTCAAGTAGATCGGAAGAGCGTAGTGTAG
Probe6 (SEQ ID NO: 85)	GCGTTCAGAGTGGCAGTCGAGATCACGATAACAGTATTGGCCATAGATCGGAAGAGCGTAGTGTAG
Probe7 (SEQ ID NO: 86)	GCGTTCAGAGTGGCAGTCGAGATCACGGTCGAATCAGGGTAGCGAGATCGGAAGAGCGTAGTGTAG
Probe8 (SEQ ID NO: 87)	GCGTTCAGAGTGGCAGTCGAGATCACACTACGGCGAAAGTGGGCAGATCGGAAGAGCGTAGTGTAG
Probe9 (SEQ ID NO: 88)	GCGTTCAGAGTGGCAGTCGAGATCACATCTTGCTCAAAGTTGCTAGATCGGAAGAGCGTAGTGTAG
Probe10 (SEQ ID NO: 89)	GCGTTCAGAGTGGCAGTCGAGATCACCCGGAATTCCGAATTGGCAGATCGGAAGAGCGTAGTGTAG
Probe11 (SEQ ID NO: 90)	GCGTTCAGAGTGGCAGTCGAGATCACATGTATTACCTTGGGCGAAGATCGGAAGAGCGTAGTGTAG
Probe12 (SEQ ID NO: 91)	GCGTTCAGAGTGGCAGTCGAGATCACCTCGAATAGGAAATGCGAAGATCGGAAGAGCGTAGTGTAG
Probe13 (SEQ ID NO: 92)	GCGTTCAGAGTGGCAGTCGAGATCACGGCGGTTAGATTTAGCAAAGATCGGAAGAGCGTAGTGTAG

Probe14 (SEQ ID NO: 93)	GCGTTCAGAGTGGCAGTCGAGATCACCCATCAGAATTTAATTCCAGATCGGAAGAGCGTAGTGTAG
Probe15 (SEQ ID NO: 94)	GCGTTCAGAGTGGCAGTCGAGATCACCTTAGCACCTATGTAATGAGATCGGAAGAGCGTAGTGTAG
Probe16 (SEQ ID NO: 95)	GCGTTCAGAGTGGCAGTCGAGATCACGTGCCGAGCGCAAGTCAATAGATCGGAAGAGCGTAGTGTAG
Probe17 (SEQ ID NO: 96)	GCGTTCAGAGTGGCAGTCGAGATCACGAACTTGGGAGATACTATAGATCGGAAGAGCGTAGTGTAG
Probe18 (SEQ ID NO: 97)	GCGTTCAGAGTGGCAGTCGAGATCACTGCGGATTACAGGCGCACACAGATCGGAAGAGCGTAGTGTAG
Probe19 (SEQ ID NO: 98)	GCGTTCAGAGTGGCAGTCGAGATCACCTACCTAAAGAGTGGCGCAGATCGGAAGAGCGTAGTGTAG
Probe20 (SEQ ID NO: 99)	GCGTTCAGAGTGGCAGTCGAGATCACAAGCCAATCACTTGCATAAGATCGGAAGAGCGTAGTGTAG
Probe21 (SEQ ID NO: 100)	GCGTTCAGAGTGGCAGTCGAGATCACCGTATAAACGTGGCTTGGAGATCGGAAGAGCGTAGTGTAG
Probe22 (SEQ ID NO: 101)	GCGTTCAGAGTGGCAGTCGAGATCACGTTATACAGCAATCAGGTAGATCGGAAGAGCGTAGTGTAG
Probe23 (SEQ ID NO: 102)	GCGTTCAGAGTGGCAGTCGAGATCACTAGAGGATAGATGCGCTGAGATCGGAAGAGCGTAGTGTAG
Probe24 (SEQ ID NO: 103)	GCGTTCAGAGTGGCAGTCGAGATCACACTAGTCCTACGCGTGGAAGATCGGAAGAGCGTAGTGTAG
Probe25 (SEQ ID NO: 104)	GCGTTCAGAGTGGCAGTCGAGATCACGCGCTACATACTTAGTCGAGATCGGAAGAGCGTAGTGTAG
Frame probe	
Layout1 (SEQ ID NO: 105)	AAATTTCGTCTGCTATCGCGCTTCTGTACC
Capture probe	
LP_Poly-dTVN (SEQ ID NO: 106)	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Amplification handle probe	
A-handle (SEQ ID NO: 107)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Second strand synthesis and first PCR amplification handles	
A_primer (SEQ ID NO: 108)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
B_dt20VN_primer (SEQ ID NO: 109)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Second PCR	
A_primer (SEQ ID NO: 110)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
B_primer (SEQ ID NO: 111)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Example 11

Template switching

Templateswitch_longB (SEQ ID NO: 112)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATrGrGrG

Example 12

Spatial genomics

A_primer (SEQ ID NO: 113)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
B_A431_Chr2+2_FW_A (SEQ ID NO: 114)	AGACGTGTGCTCTTCCGATCTTGGCTGCCTGAGGCAATG
B_A431_Chr2+2_RE_A (SEQ ID NO: 115)	AGACGTGTGCTCTTCCGATCTCTCGCTAACAAGCAGAGAGAAC
B_A431_Chr3+7_FW_B (SEQ ID NO: 116)	AGACGTGTGCTCTTCCGATCTTGAGAACAAGGGGGAAGAG
B_A431_Chr3+7_RE_B (SEQ ID NO: 117)	AGACGTGTGCTCTTCCGATCTCGGTGAAACAAGCAGGTAAC
B_NT_1_FW (SEQ ID NO: 118)	AGACGTGTGCTCTTCCGATCTCATTCCCACACTCATCACAC
B_NT_1_RE (SEQ ID NO: 119)	AGACGTGTGCTCTTCCGATCTTCACACTGGAGAAAGACCC
B_NT_2_FW (SEQ ID NO: 120)	AGACGTGTGCTCTTCCGATCTGGGGTTCAGAGTGATTTTTCAG
B_NT_2_RE (SEQ ID NO: 121)	AGACGTGTGCTCTTCCGATCTTCCGTTTTCTTTCAGTGCC

Claims

1. A method for localised detection of nucleic acid in a tissue sample comprising:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample and allowing nucleic acid of the tissue sample to hybridise to the capture domain in said capture probes;

(c) generating DNA molecules from the captured nucleic acid molecules using said capture probes as extension or ligation primers, wherein said extended or ligated DNA molecules are tagged by virtue of the positional domain;

(d) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying said tagged DNA;

(e) releasing at least part of the tagged DNA molecules and/or their complements or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof; and

(f) directly or indirectly analysing the sequence of the released DNA molecules.

2. The method of claim 1, further comprising step (g) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (c).

3. The method of claim 1 being a method for determining and/or analysing a transcriptome of a tissue sample comprising:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array; and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample and allowing RNA of the tissue sample to hybridise to the capture domain in said capture probes;

(c) generating cDNA molecules from the captured RNA molecules using said capture probes as RT primers and optionally amplifying said cDNA molecules;

(d) releasing at least part of the cDNA molecules and optionally their amplicons from the surface of the array, wherein said released molecule may be a first strand and/or second strand cDNA molecule or an amplicon thereof and wherein said part includes the positional domain or a complement thereof;

(e) directly or indirectly analysing the sequence of the released DNA molecules; and optionally

(f) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (c).

4. The method of any one of claims 1 to 3, wherein the capture probes are DNA molecules.

5. The method of any one of claims 1 to 4, wherein the capture probes further comprise a universal domain which is 5' relative to the positional domain, wherein said universal domain comprises:

(i) an amplification domain, for amplifying the generated DNA molecules; and/or

(ii) a cleavage domain for releasing the generated DNA molecules from the surface of the array.

6. The method of any one of claims 1 to 5, wherein the positional domain of each species of capture probe comprises a unique barcode sequence.

7. The method of any one of claims 1 to 6, wherein the capture domain comprises a poly-T DNA oligonucleotide comprising at least 10 deoxythymidine residues and/or a random or degenerate oligonucleotide sequence.

8. The method of any one of claims 1 to 7, wherein the capture probes are directly immobilized on the array surface by their 5' end.

9. The method of any one of claims 1 to 7, wherein the capture probes are indirectly immobilized on the array surface by hybridization to a surface probe, wherein the capture domain of the capture probes comprises an upstream sequence that is capable of hybridizing to 5' end of surface probes that are immobilized on the array.

10. The method of claim 9, wherein the surface probes are immobilized to the array surface by their 3' end.

11. The method of claim 9 or 10, wherein the surface probes comprise a sequence that is complementary to:

(i) at least part of the capture domain;

(ii) the positional domain; and

(iii) at least part of the universal amplification domain.

12. The method of any one of claims 1 to 11, wherein said complementary strand or said second strand cDNA molecule is generated before or after said tagged DNA molecules or said cDNA molecules are released from the surface of the array.

13. The method of any one of claims 1 to 12, wherein a strand displacement polymerase is used for the synthesis of the complementary strand of the second strand of cDNA.

14. The method of any one of claims 1 to 13, wherein random primers are used for complementary strand or second strand cDNA synthesis.

15. The method of claim 14, wherein the random primers comprise an amplification domain.

16. The method of claim 15, wherein the amplification domain of the random primers consists of a nucleotide sequence that is different to the amplification domain of the capture probe.

17. The method of any one of claims 1 to 16, wherein adaptors are ligated to one or both ends of the tagged molecules or their complements or the cDNA molecules.

18. The method of any one of claims 1 to 17, further comprising the step of amplifying the tagged molecules or cDNA molecules prior to sequence analysis.

19. The method of claim 18, wherein the amplification step is performed after the release of the tagged DNA or cDNA molecules from the substrate of the array, or wherein the amplification step is performed *in situ* on the array.

20. The method of claim 18 or 19, wherein the step of amplifying comprises a PCR.

21. The method of any one of claims 1 to 20, wherein the molecules are released from the surface of the array by:

(i) nucleic acid cleavage;

(ii) denaturation; and/or

(iii) physical means.

22. The method of claim 21(i), wherein the molecules are released by enzymatic cleavage of a cleavage domain, which is located in the universal domain or positional domain of the capture probe.

23. The method of claim 21(ii) or (iii), wherein the molecules are released by applying hot water or buffer to the array.

24. The method of any one of claims 1 to 23, further comprising a step of purifying the released molecules prior to sequencing.

25. The method of any one of claims 1 to 24, wherein the substrate is selected from the group consisting of glass, silicon, poly-L-lysine coated material, nitrocellulose, polystyrene, cyclic olefin copolymers (COCs), cyclic olefin polymers (COPs), polypropylene, polyethylene and polycarbonate.

26. The method of any one of claims 1 to 25, wherein the tissue sample is a tissue section.

27. The method of claim 26, wherein the tissue section is prepared using a fixed tissue, e.g. a formalin-fixed paraffin-embedded (FFPE) tissue, or deep-frozen tissue.

28. The method of any one of claims 1 to 27, further comprising a step of rehydrating the tissue sample after contacting the sample with the array and prior to step (c).

29. The method of any one of claims 1 to 28, wherein the tissue sample is from a plant, animal or fungus.

30. The method of any one of claims 1 to 29, further comprising a step of washing the array to remove residual tissue, wherein said step is prior to the release of the molecules from the array.

31. The method of any one of claims 1 to 30, wherein the array comprises at least one positional marker to enable orientation of the tissue sample on the array.

32. The method of claim 31, wherein the array comprises at least 10, 50, 100, 200, 500, 1000 or 2000 positional markers in distinct positions on the surface of the array.

33. The method of claim 31 or 32, wherein the positional markers are capable of hybridizing to a labelled, preferably fluorescently labelled, marker nucleic acid molecule.

34. The method of any one of claims 2 to 33, wherein the tissue sample is imaged using light, bright field, dark field, phase contrast, fluorescence, reflection, interference or confocal microscopy or a combination thereof.

35. The method of claim 34, wherein the tissue sample is imaged using fluorescence microscopy.

36. The method of any one of claims 1 to 35, wherein the sequence analysis step includes a step of sequencing, and preferably wherein the sequencing step comprises a sequencing reaction based on reversible dye-terminators.

37. A method for making or producing an array (i) for use in capturing nucleic acid from a tissue sample that is contacted with said array; or (ii) for use in localised detection of nucleic acid in a tissue sample, said method comprising immobilizing, directly or indirectly, multiple species of capture probe to an array substrate, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array; and

(ii) a capture domain.

38. The method of claim 37, wherein each species of capture probe occupies a distinct position on the array.

39. The method of claims 37 or 38, further comprising any one or more of the following steps:

(a) immobilizing directly or indirectly multiple surface probes to an array substrate, wherein the surface probes comprise:

(i) a domain capable of hybridizing to part of the capture domain oligonucleotide (a part not involved in capturing the nucleic acid);

(ii) a complementary positional domain; and

(iii) a complementary universal domain;

(b) hybridizing to the surface probes immobilized on the array capture domain oligonucleotides and universal domain oligonucleotides;

(c) extending the universal domain oligonucleotides, by templated polymerization, to generate the positional domain of the capture probe; and

(d) ligating the positional domain to the capture domain oligonucleotide to produce the capture oligonucleotide.

40. An array for use in (i) capturing nucleic acid from a tissue sample that is contacted with said array; or (ii) for use in the localised detection of nucleic acid in a tissue sample, comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as an extension or ligation primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain to capture nucleic acid of a tissue sample that is contacted with said array.

41. The method of any one of claims 37 to 39 or the array of claim 40, wherein the capture probes are as defined in any one of claims 4 to 9, the surface probes are as defined in claims 10 or 11, the substrate is as defined in claim 25, the tissue sample is as defined in any one of claims 26, 27 or 28, the array is as defined in any one of claims 31 to 33 and the positional markers are as defined in claim 33.

42. The method of claim 1 or claim 2 being a method for localised detection of DNA in a tissue sample comprising:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction,

wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample and allowing DNA of the tissue sample to hybridise to the capture domain in said capture probes;

(c) fragmenting DNA in said tissue sample, wherein said fragmentation is carried out before, during or after contacting the array with the tissue sample in step (b);

(d) extending said capture probes in a primer extension reaction using the captured DNA fragments as templates to generate extended DNA molecules, or ligating the captured DNA fragments to the capture probes in a ligation reaction to generate ligated DNA molecules, wherein said extended or ligated DNA molecules are tagged by virtue of the positional domain;

(e) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying said tagged DNA;

(f) releasing at least part of the tagged DNA molecules and/or their complements and/or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof ;

(g) directly or indirectly analysing the sequence of (e.g. sequencing) the released DNA molecules; and optionally;

(h) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (d).

43. The method of claim 42 comprising:

(a) providing an array comprising a substrate on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the array and is oriented to have a free 3' end to enable said probe to function as a primer for a primer extension or ligation reaction, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the array, and

(ii) a capture domain;

(b) contacting said array with a tissue sample such that the position of a capture probe on the array may be correlated with a position in the tissue sample;

(c) fragmenting DNA in said tissue sample, wherein said fragmentation is carried out before, during or after contacting the array with the tissue sample in step (b);

(d) providing said DNA fragments with a binding domain which is capable of hybridising to said capture domain;

(e) allowing said DNA fragments to hybridise to the capture domain in said capture probes;

(f) extending said capture probes in a primer extension reaction using the captured DNA fragments as templates to generate extended DNA molecules, or ligating the captured DNA fragments to the capture probes in a ligation reaction to generate ligated DNA molecules, wherein said extended or ligated DNA molecules are tagged by virtue of the positional domain;

(g) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying the tagged DNA

(h) releasing at least part of the tagged DNA molecules and/or their complements and/or amplicons from the surface of the array, wherein said part includes the positional domain or a complement thereof;

(i) directly or indirectly analysing the sequence of (e.g. sequencing)the released DNA molecules; and optionally

(j) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (f).

Figure 1

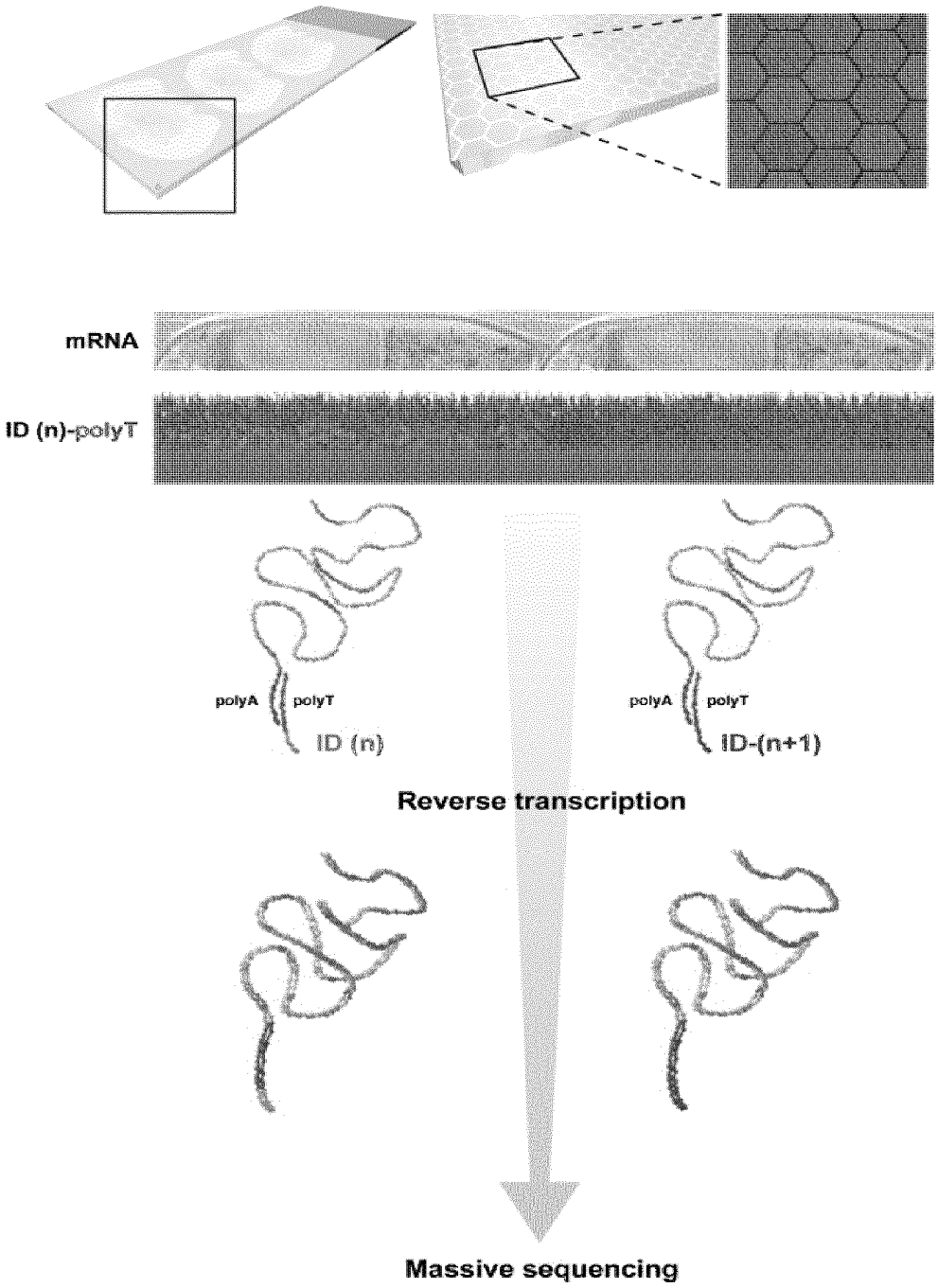

mRNA
ID (n)-polyT
polyA
polyT
ID (n)
polyA
polyT
ID-(n+1)
Reverse transcription
Massive sequencing

Figure 2

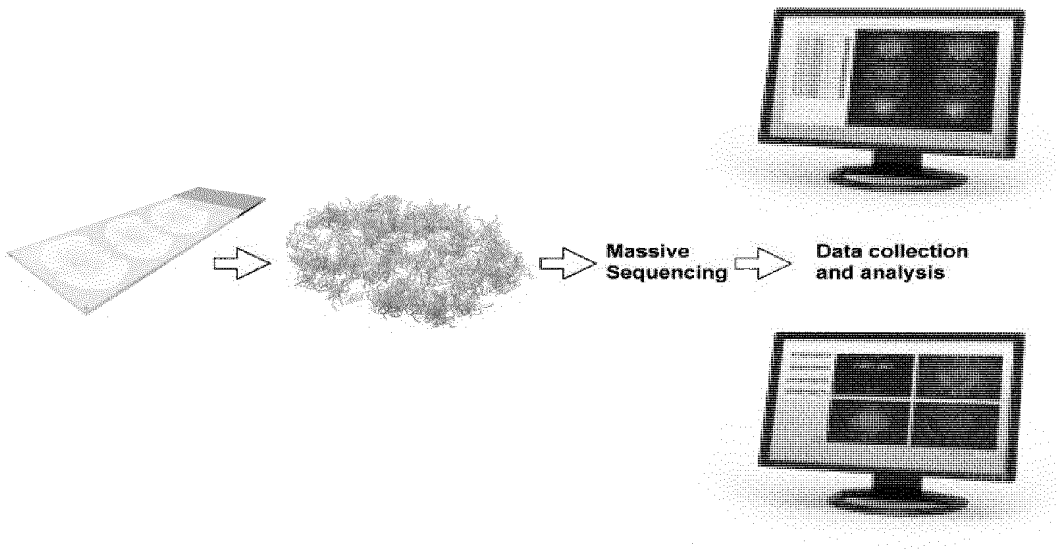


Figure 3

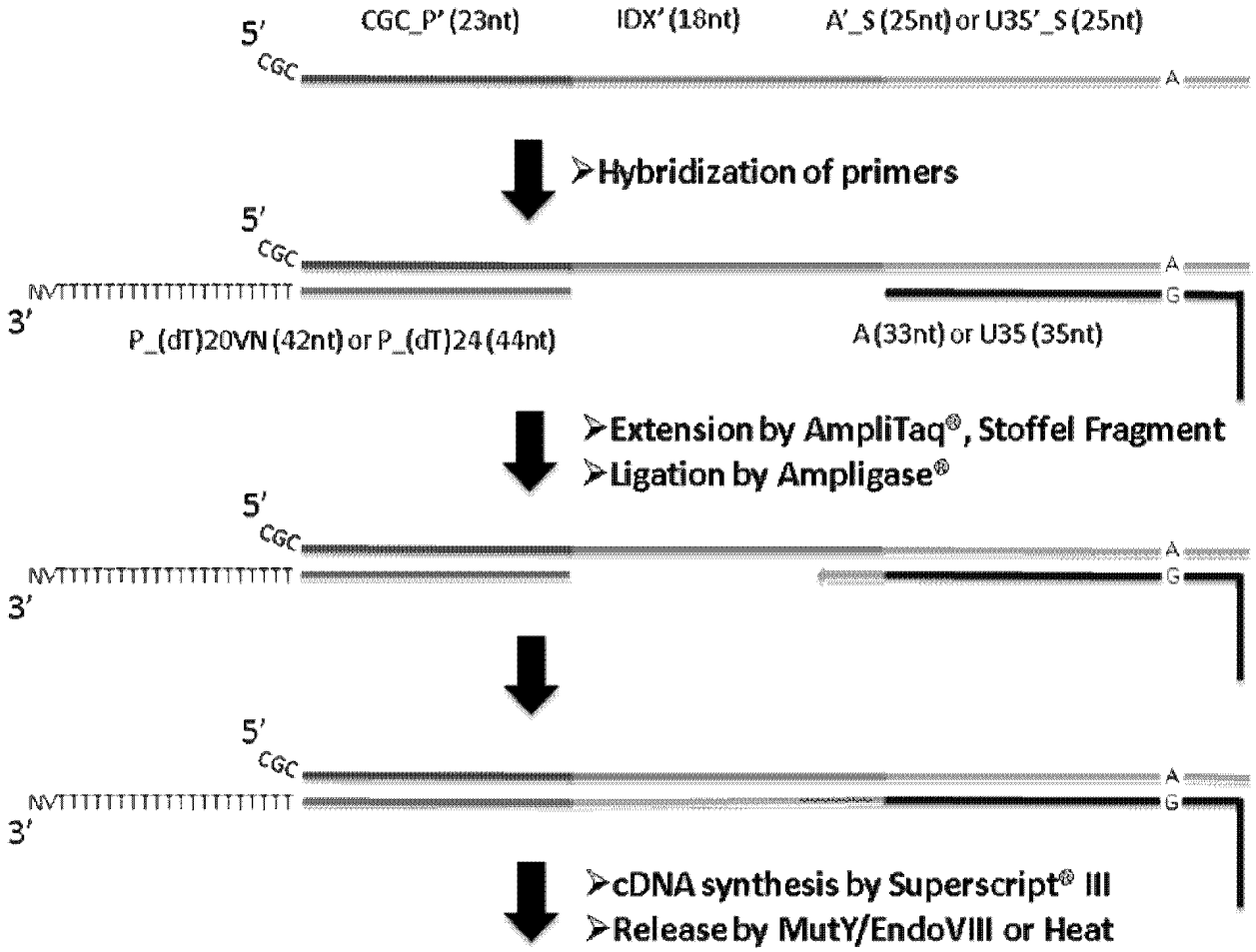

5'
CGC_P' (23nt)
IDX' (18nt)
A'_S (25nt) or U35'_S (25nt)
CGC
A
Hybridization of primers
5'
CGC
A
NVTTTTTTTTTTTTTTTTTTTT
G
3'
P_(dT)20VN (42nt) or P_(dT)24 (44nt)
A (33nt) or U35 (35nt)
Extension by AmpliTaq®, Stoffel Fragment
Ligation by Ampligase®
5'
CGC
A
NVTTTTTTTTTTTTTTTTTTTT
G
3'
5'
CGC
A
NVTTTTTTTTTTTTTTTTTTTT
G
3'
cDNA synthesis by Superscript® III
Release by MutY/EndoVIII or Heat

Figure 4

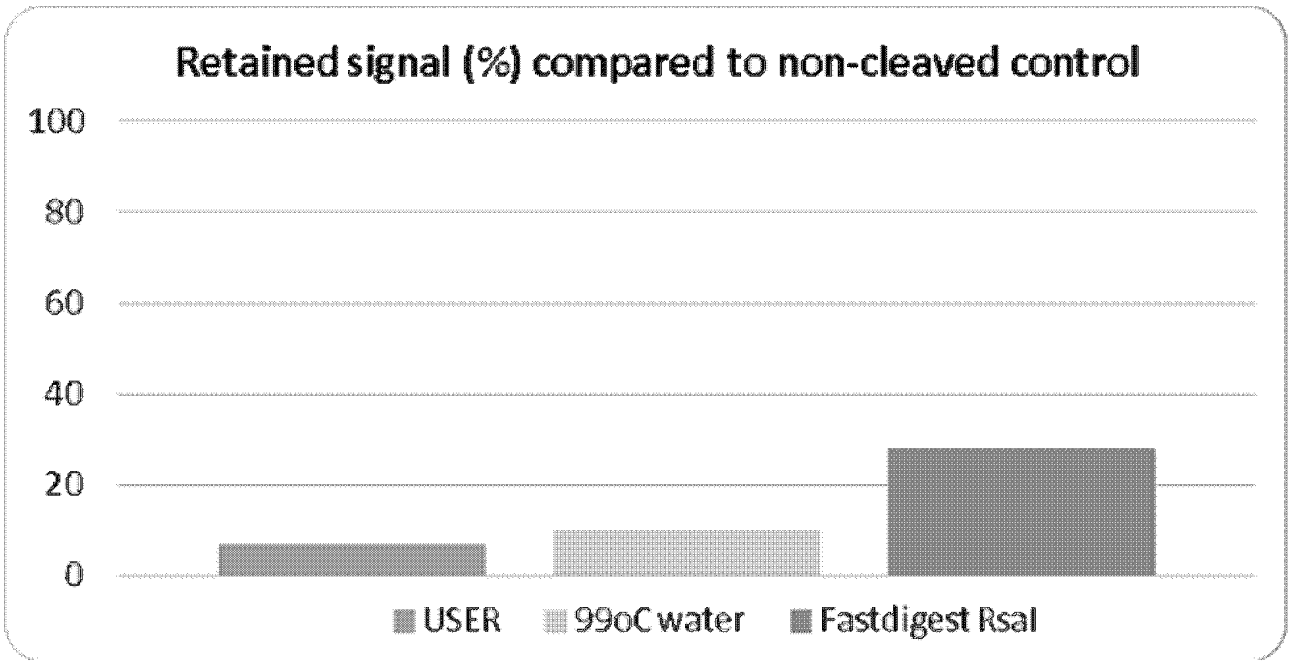

Retained signal (%) compared to non-cleaved control
100
80
60
40
20
0
USER
99oC water
Fastdigest Rsal

Figure 5

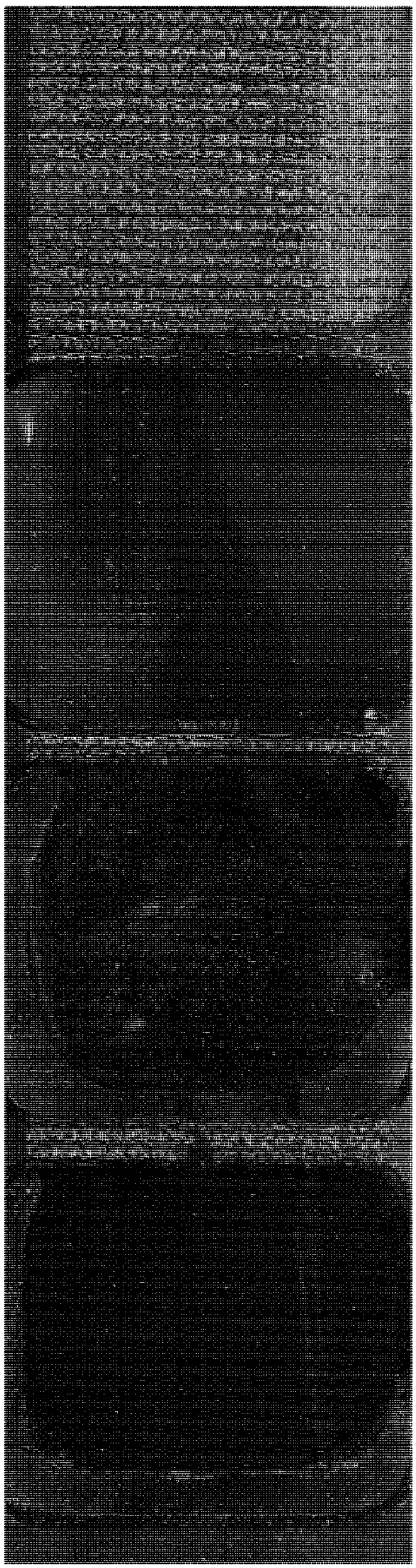


Figure 6

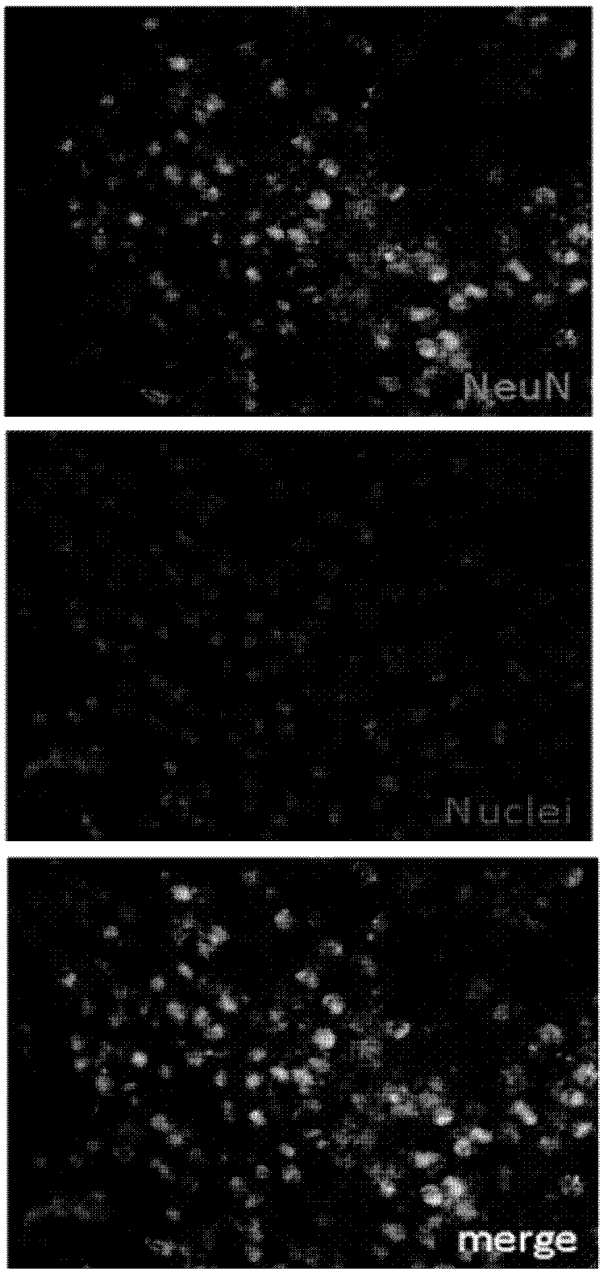


Figure 7

Lane 7 blast results	
E-value	1,00E-06
Barcode	Hits
Tag1	2165
Tag2	1223
Tag3	1338
Tag4	3605
Tag5	763
Tag6	1083
Tag7	1175
Tag8	1350
Tag9	616
Tag10	710
Tag11	1602
Tag12	1359
Tag13	1404
Tag14	1124
Tag15	1077
Tag16	1238
Tag17	1122
Tag18	1547
Tag19	1401
Tag20	1354

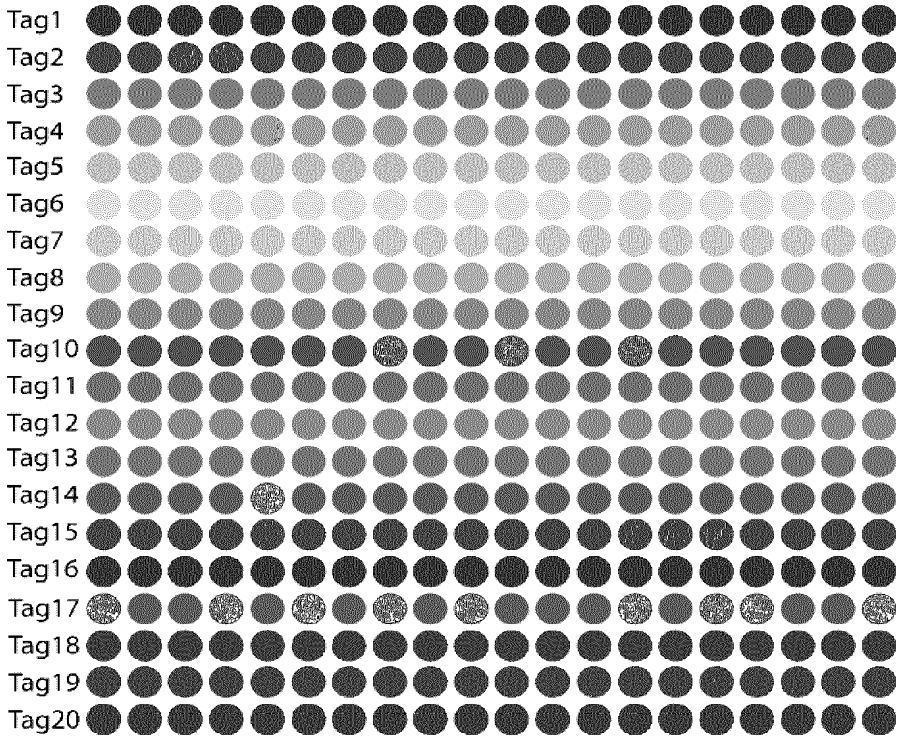


Figure 8

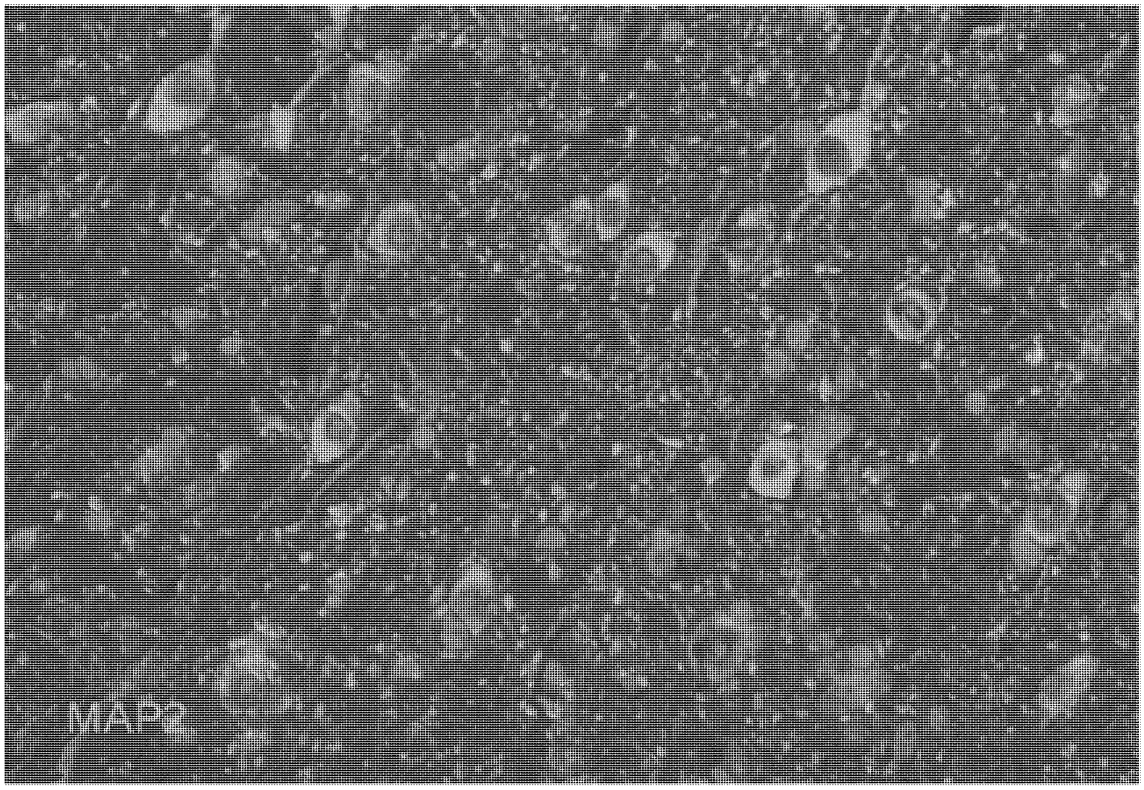


Figure 9

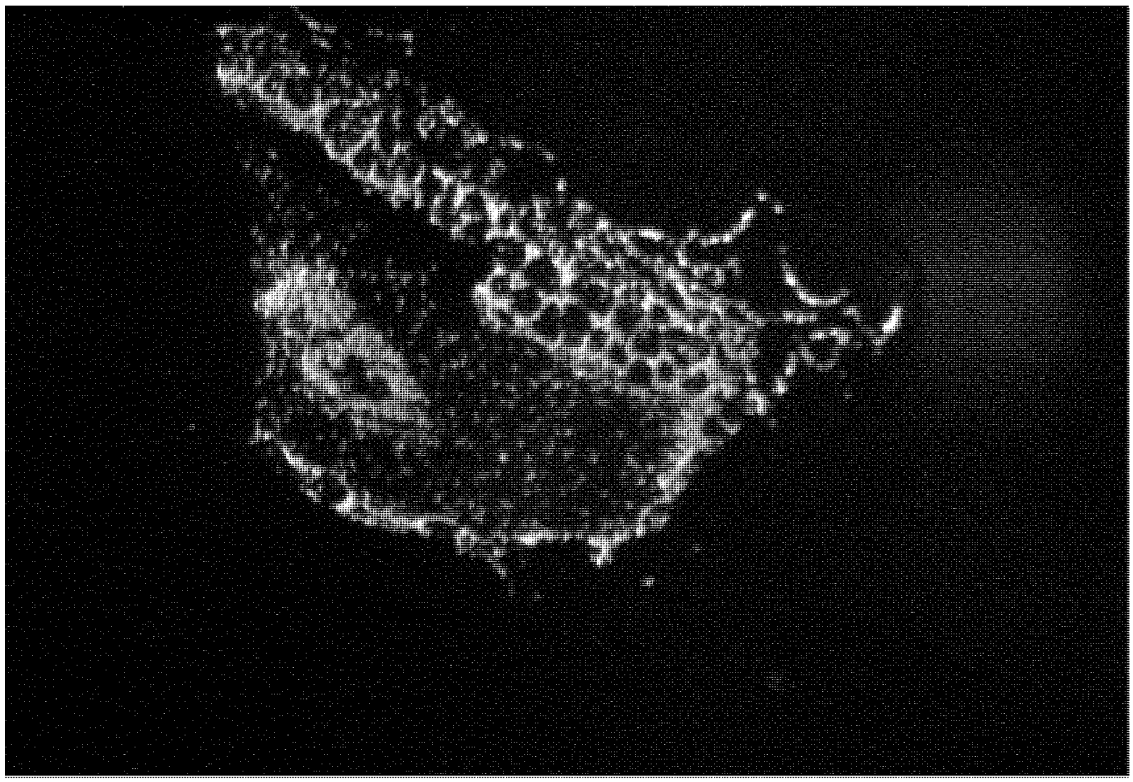


Figure 10

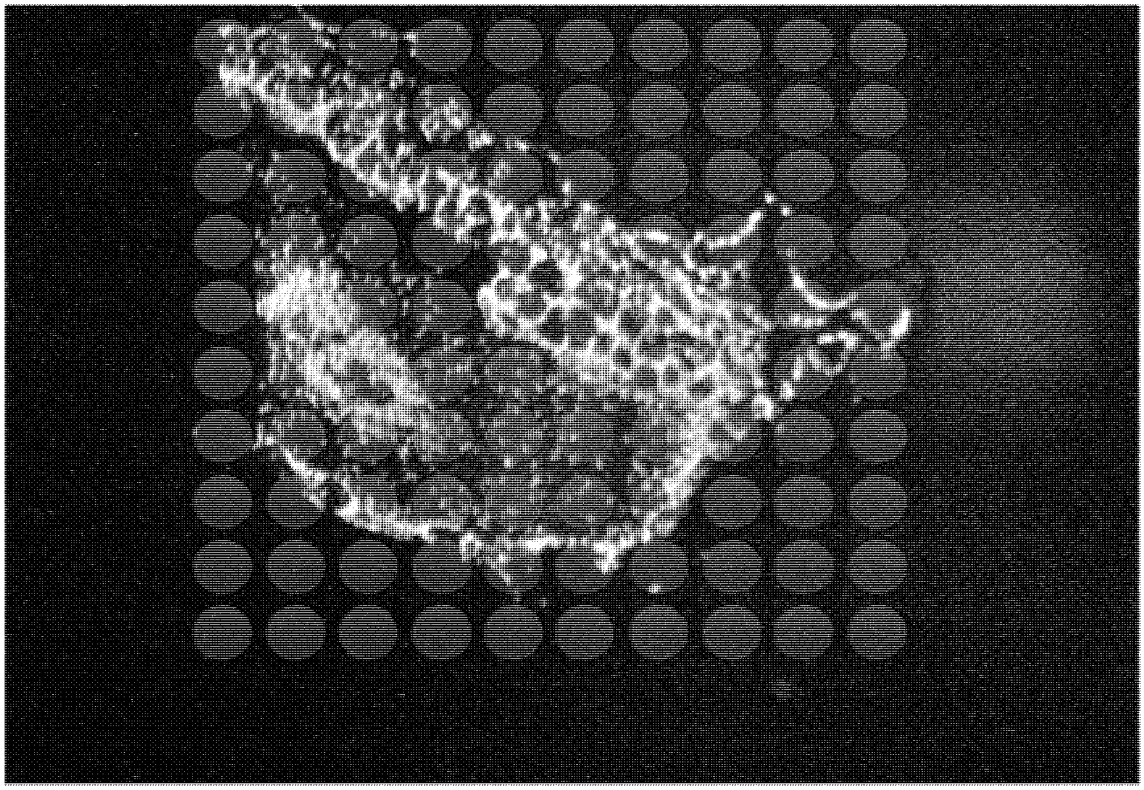


Figure 11

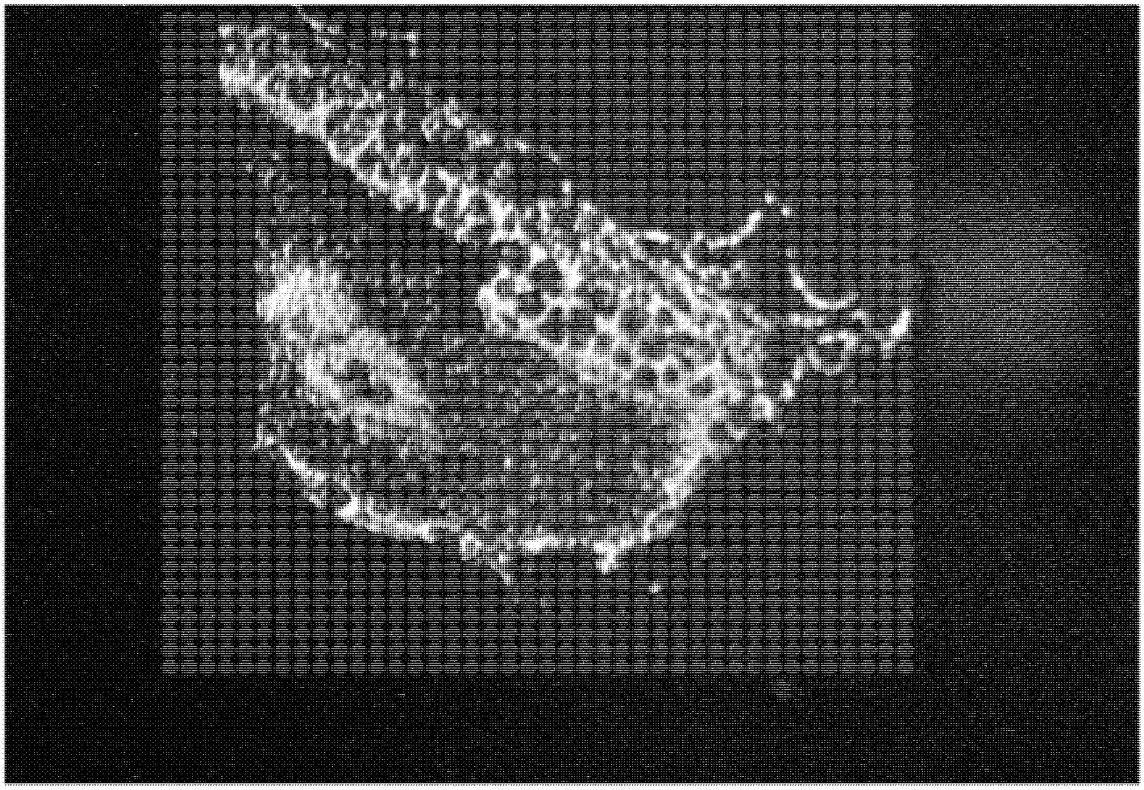


Figure 12

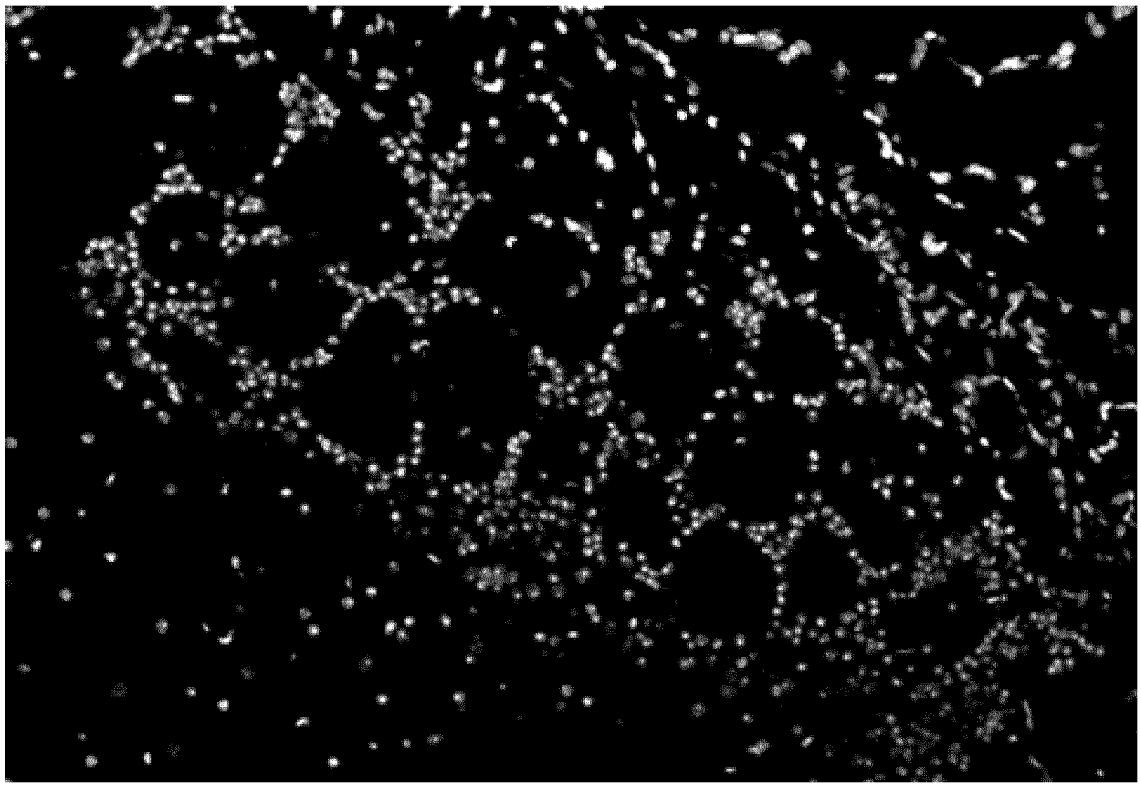
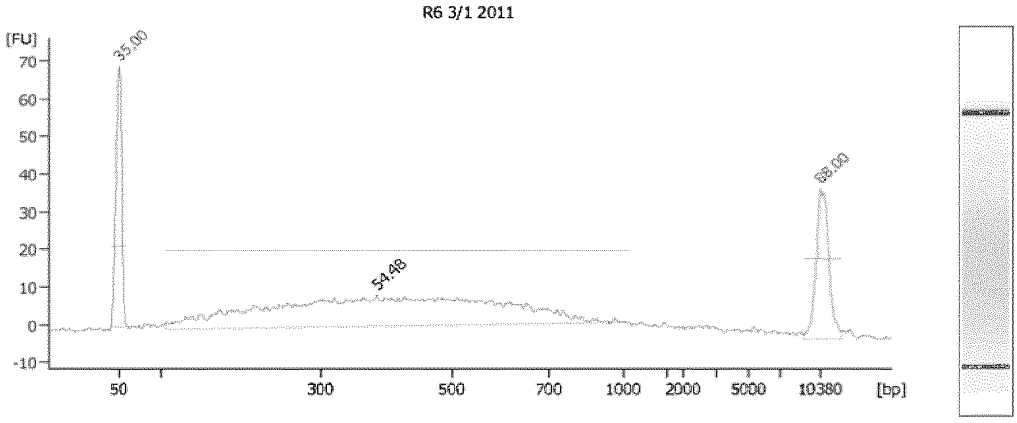


Figure 13



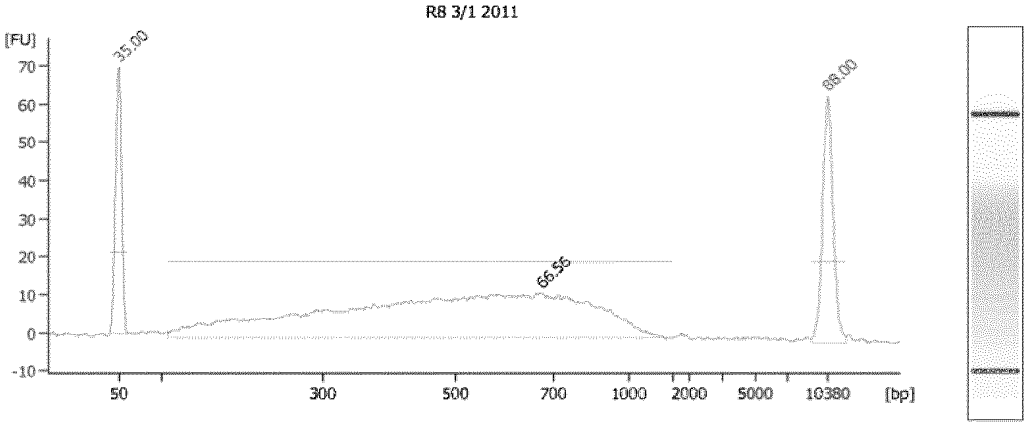
Overall Results for sample 3 : R6 3/1 2011

Number of peaks found: 1

Peak table for sample 3 : R6 3/1 2011

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	50	8.30	251.5	Lower Marker
2		386	22.87	89.8	
3	▶	10,380	4.20	0.6	Upper Marker

Figure 14



Overall Results for sample 4 : R8 3/1 2011

Number of peaks found: 1

Peak table for sample 4 : R8 3/1 2011

Peak		Size [bp]	Conc. [ng/µl]	Molarity [nmol/l]	Observations
1	◀	50	8.30	251.5	Lower Marker
2		676	21.53	48.2	
3	▶	10,380	4.20	0.6	Upper Marker

Figure 15

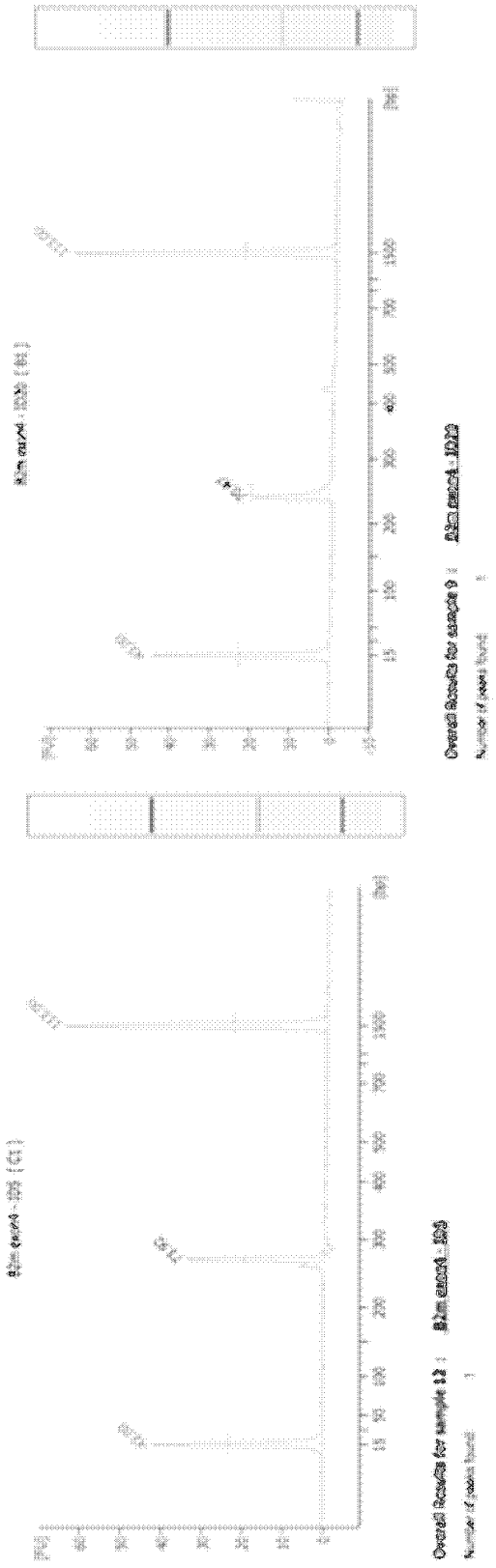


Figure 16

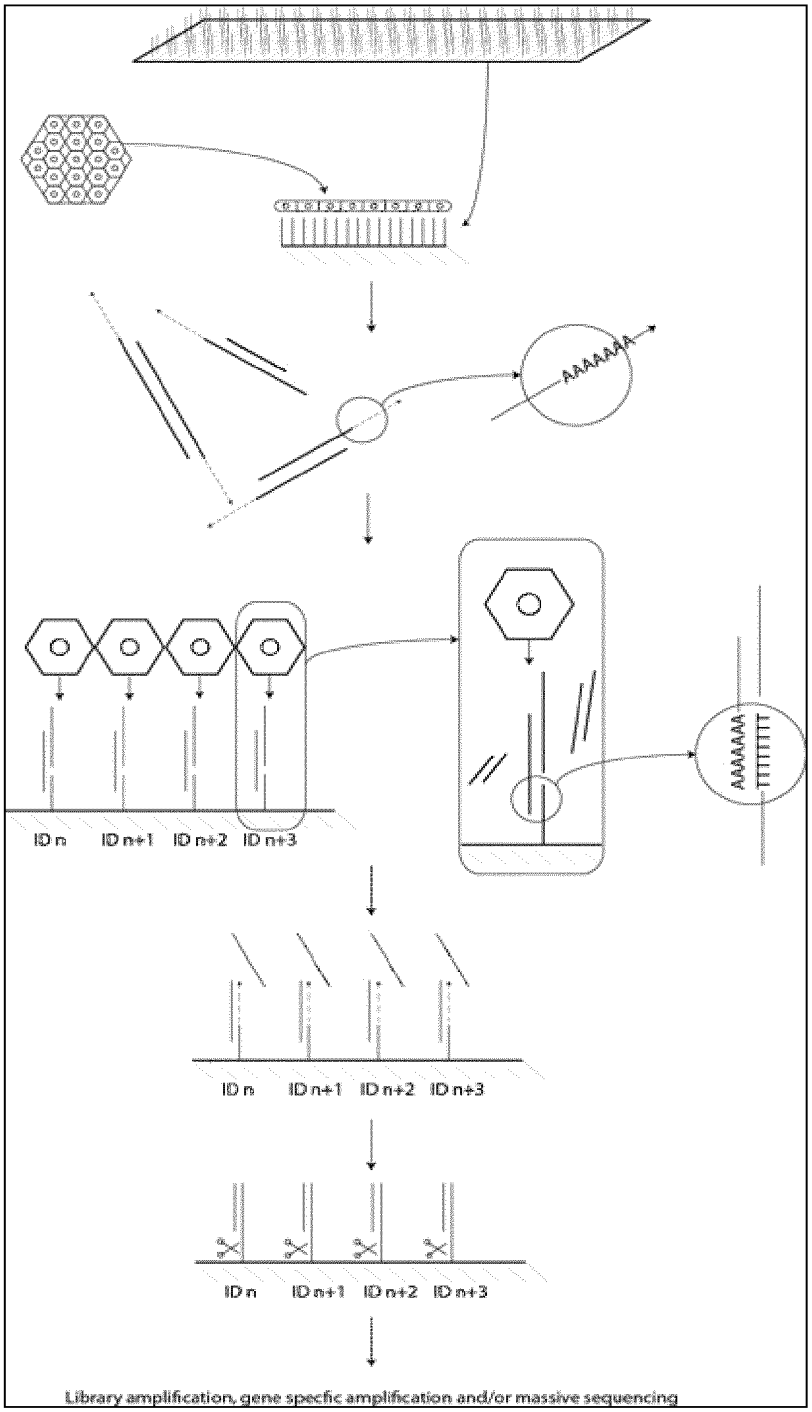
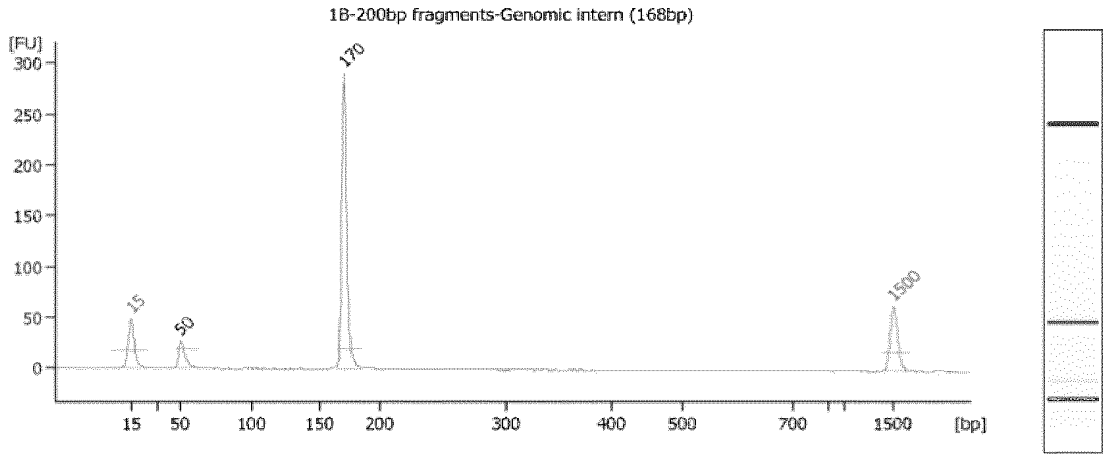

AAAAAAA
ID n
ID n+1
ID n+2
ID n+3
AAAAAAA
TTTTTTTT
ID n
ID n+1
ID n+2
ID n+3
ID n
ID n+1
ID n+2
ID n+3
Library amplification, gene specific amplification and/or massive sequencing

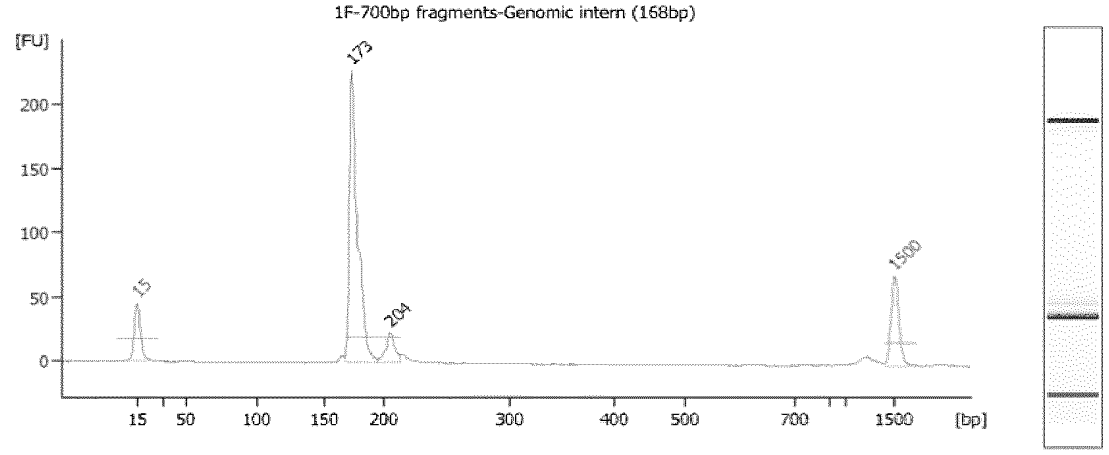
Figure 17



Overall Results for sample 3 : **1B-200bp fragments-Genomic intern (168bp)**

Number of peaks found: 2

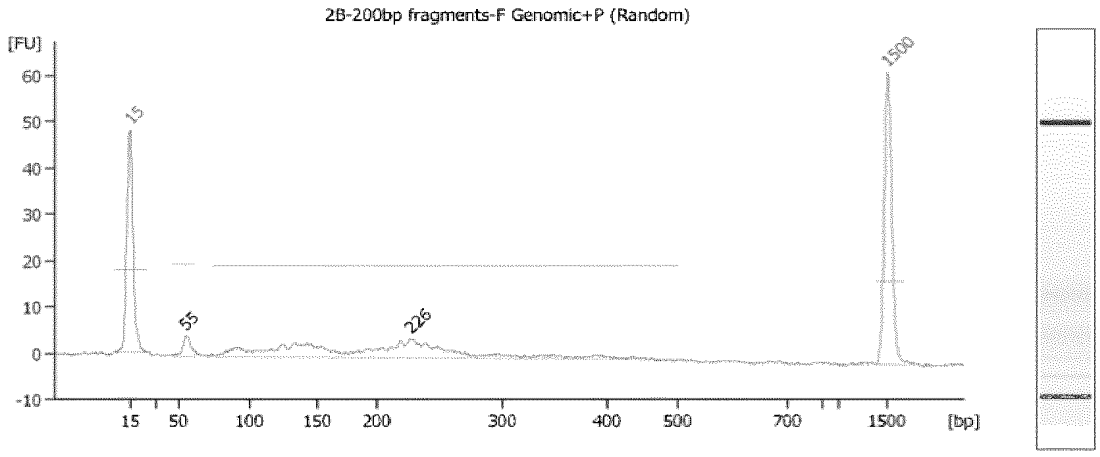
Figure 18



Overall Results for sample 7 : **1F-700bp fragments-Genomic intern (168bp)**

Number of peaks found: 2

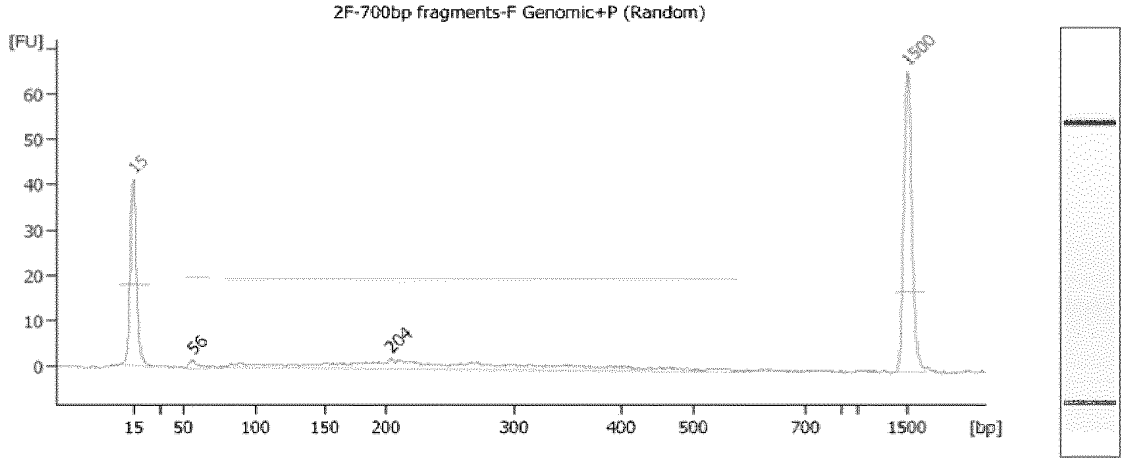
Figure 19



Overall Results for sample 4 : **2B-200bp fragments-F Genomic+P (Random)**

Number of peaks found: 2

Figure 20



Overall Results for sample 8 : **2F-700bp fragments-F Genomic+P (Random)**

Number of peaks found: 2

Figure 21

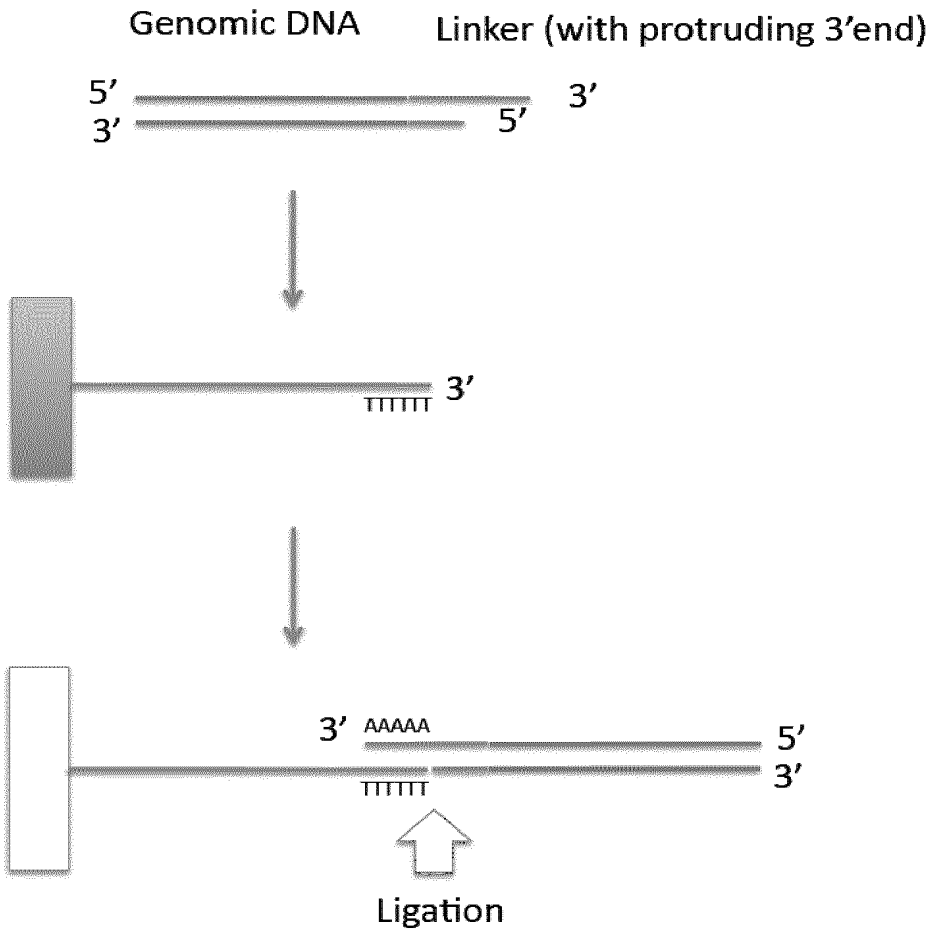

Genomic DNA
Linker (with protruding 3'end)
5'
3'
3'
5'
3'
3'
AAAAA
5'
3'
Ligation

Figure 22

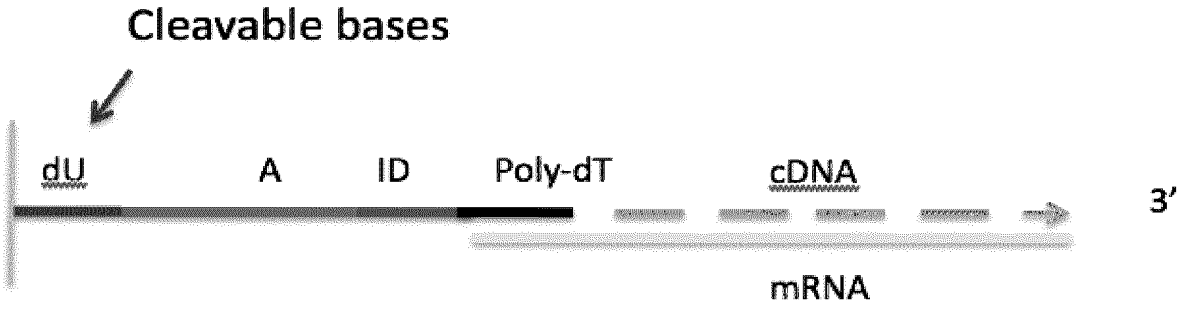


Figure 23

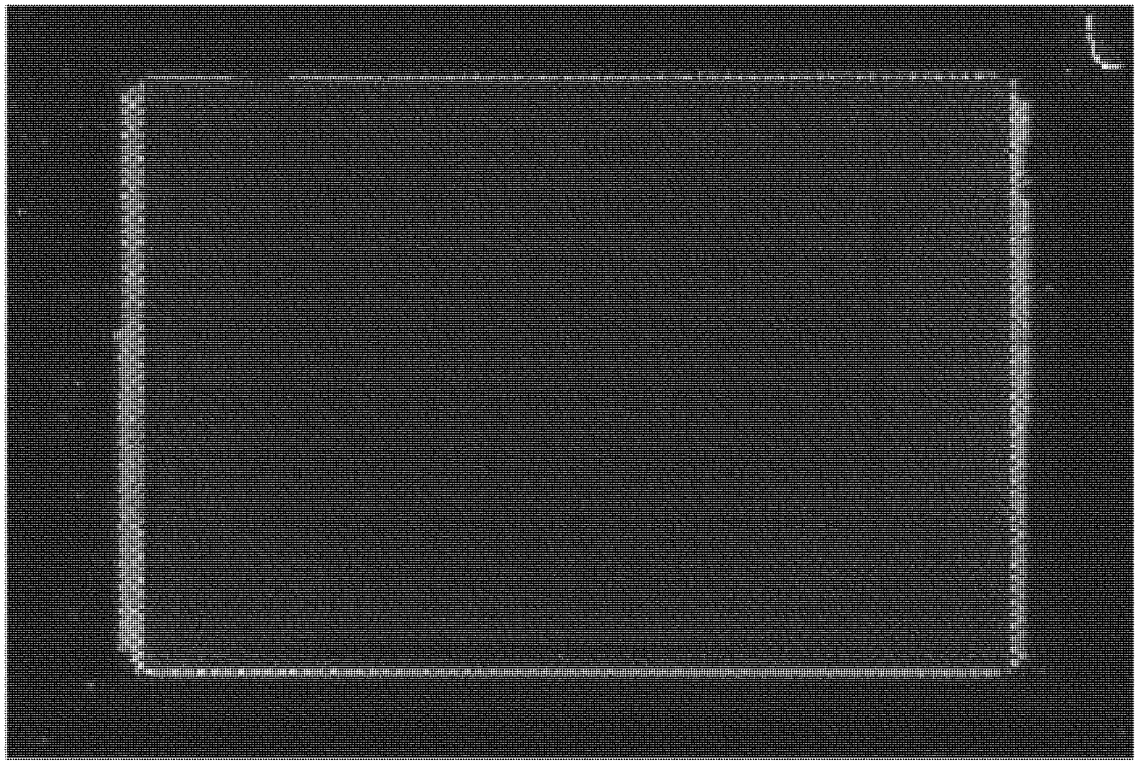


Figure 24

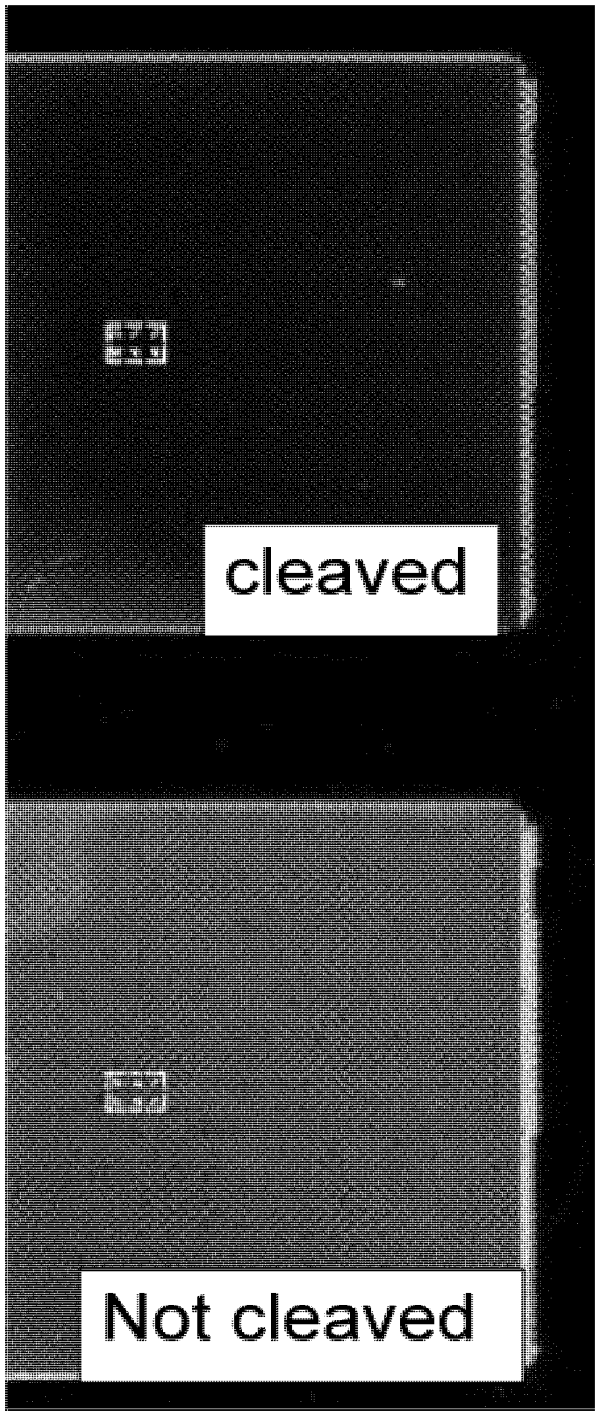


Figure 25

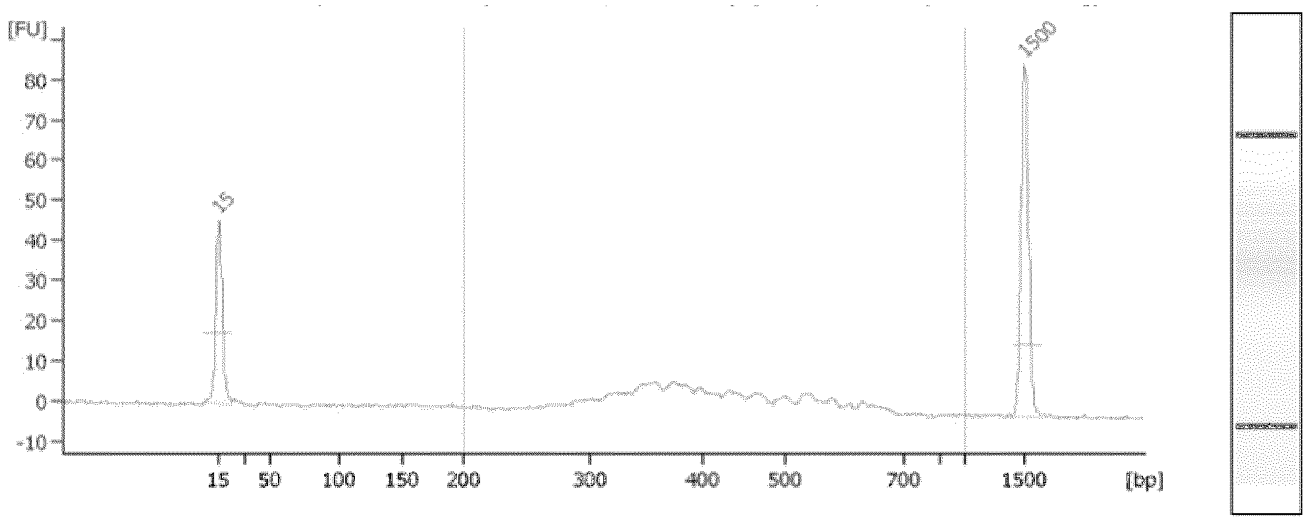

[FU]
80
70
60
50
40
30
20
10
0
-10
15
1500
15
50
100
150
200
300
400
500
700
1500
[bp]

Figure 26

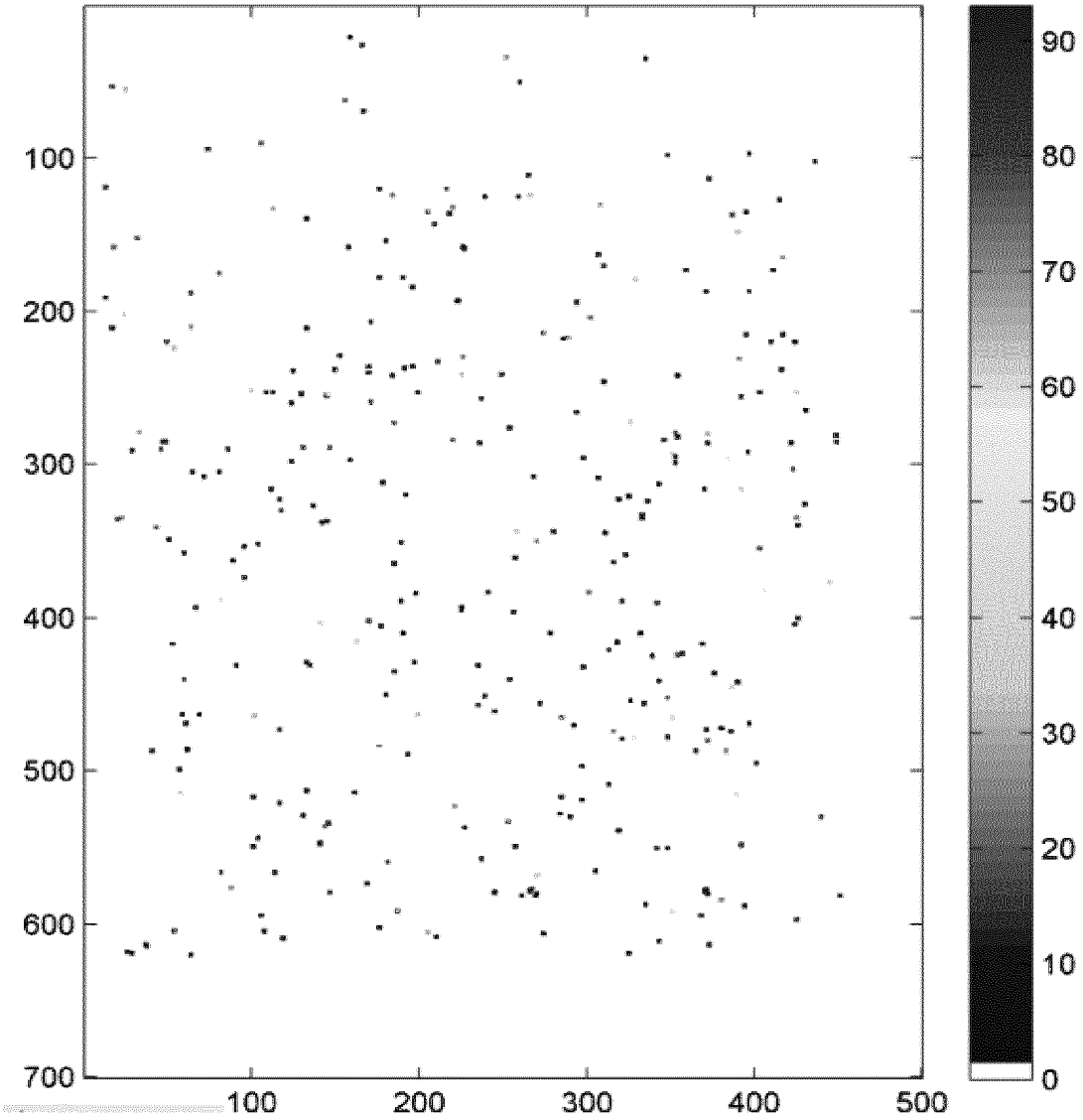

100
200
300
400
500
600
700
100
200
300
400
500
90
80
70
60
50
40
30
20
10
0

Figure 27

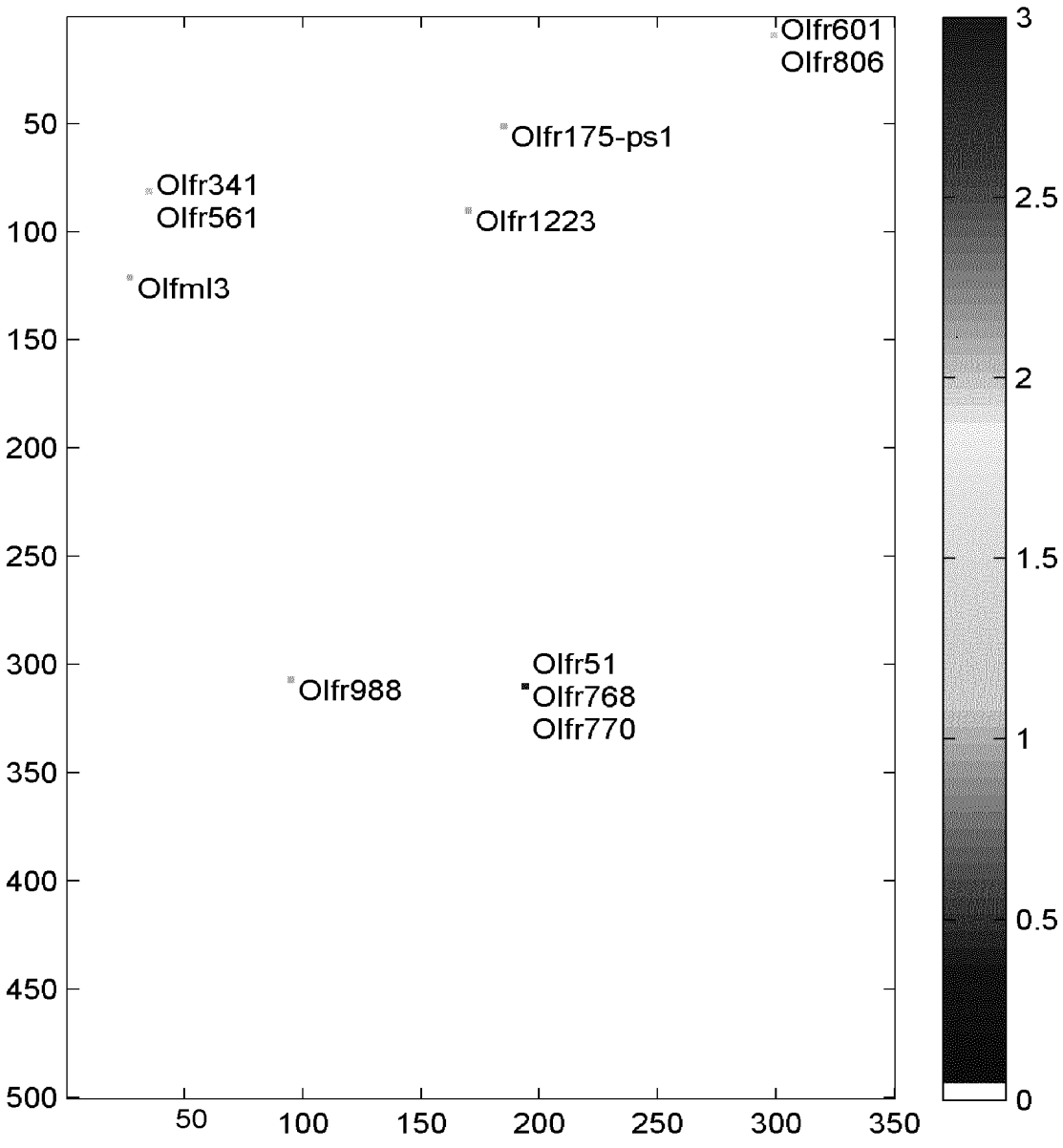

Olfr601
Olfr806
Olfr175-ps1
Olfr341
Olfr561
Olfr1223
Olfml3
Olfr988
Olfr51
Olfr768
Olfr770
50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
3
2.5
2
1.5
1
0.5
0

Figure 28

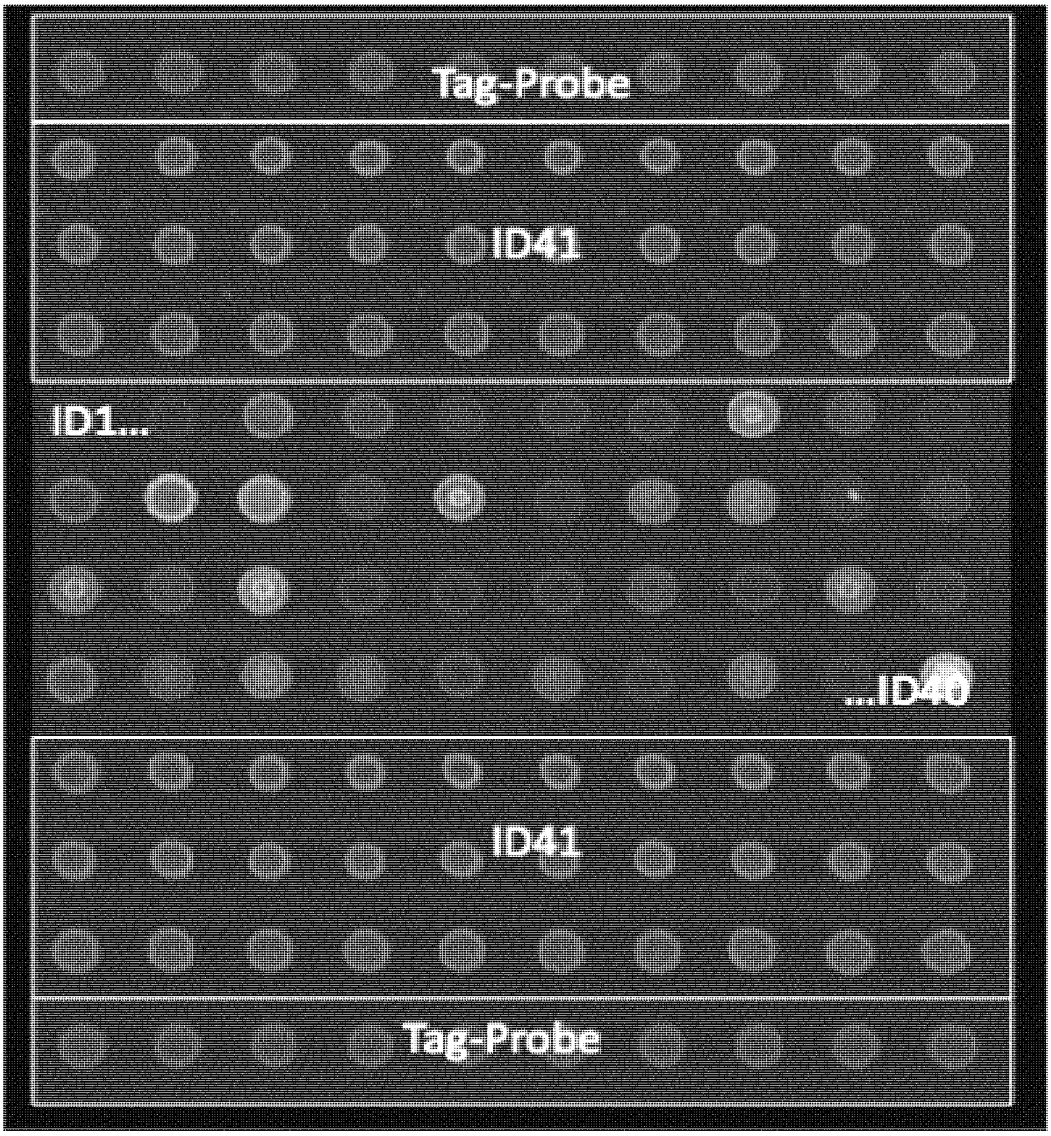


Figure 29

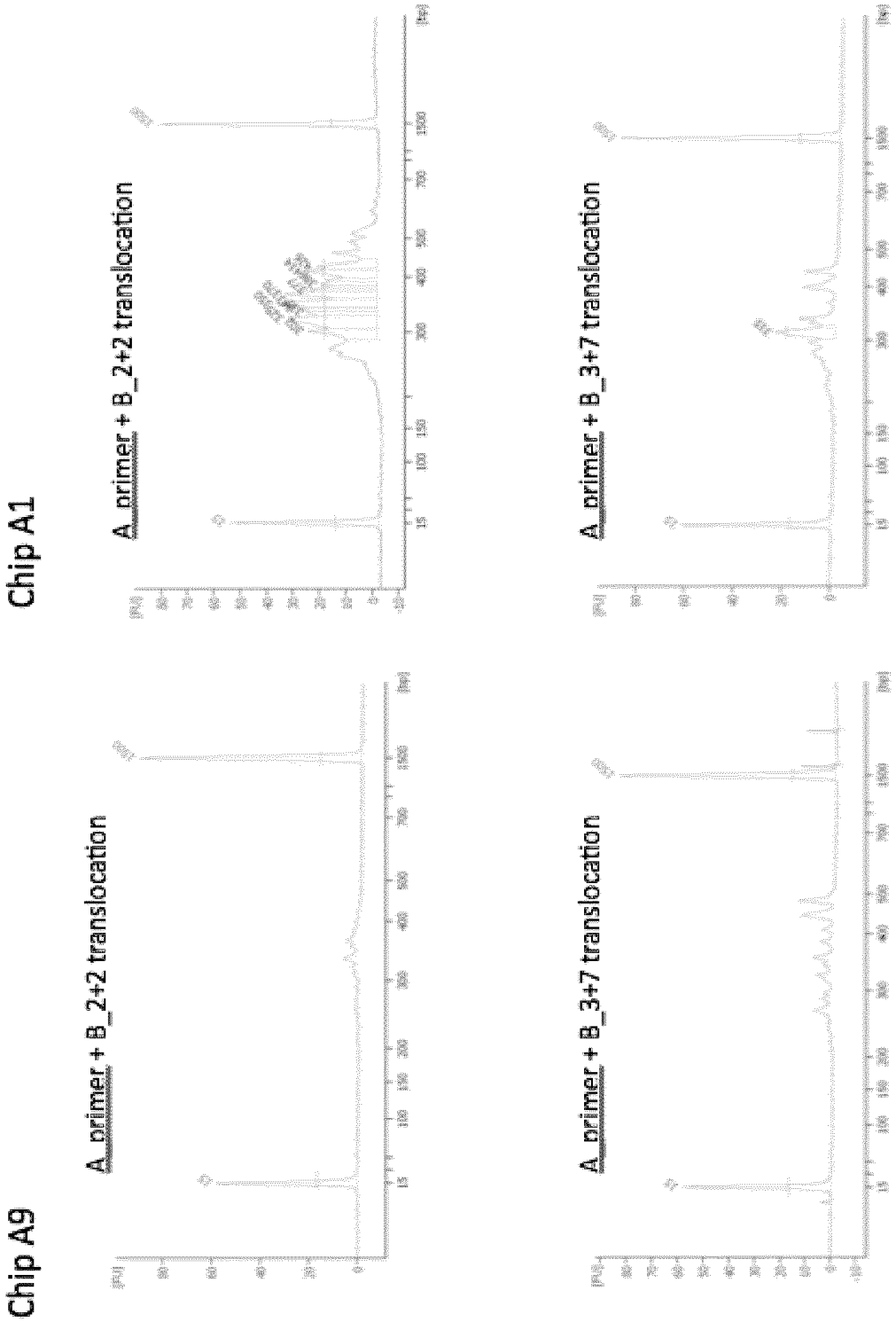

Chip A1
A_primer + B_2+2 translocation
A_primer + B_3+7 translocation
Chip A9
A_primer + B_2+2 translocation
A_primer + B_3+7 translocation

Figure 30

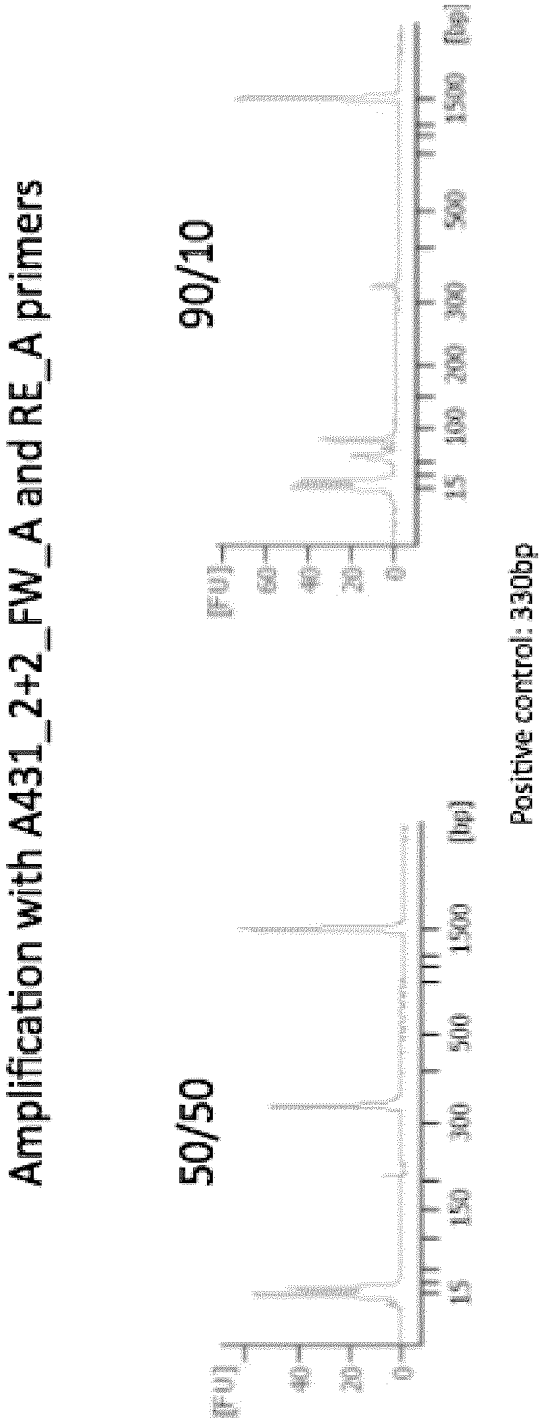
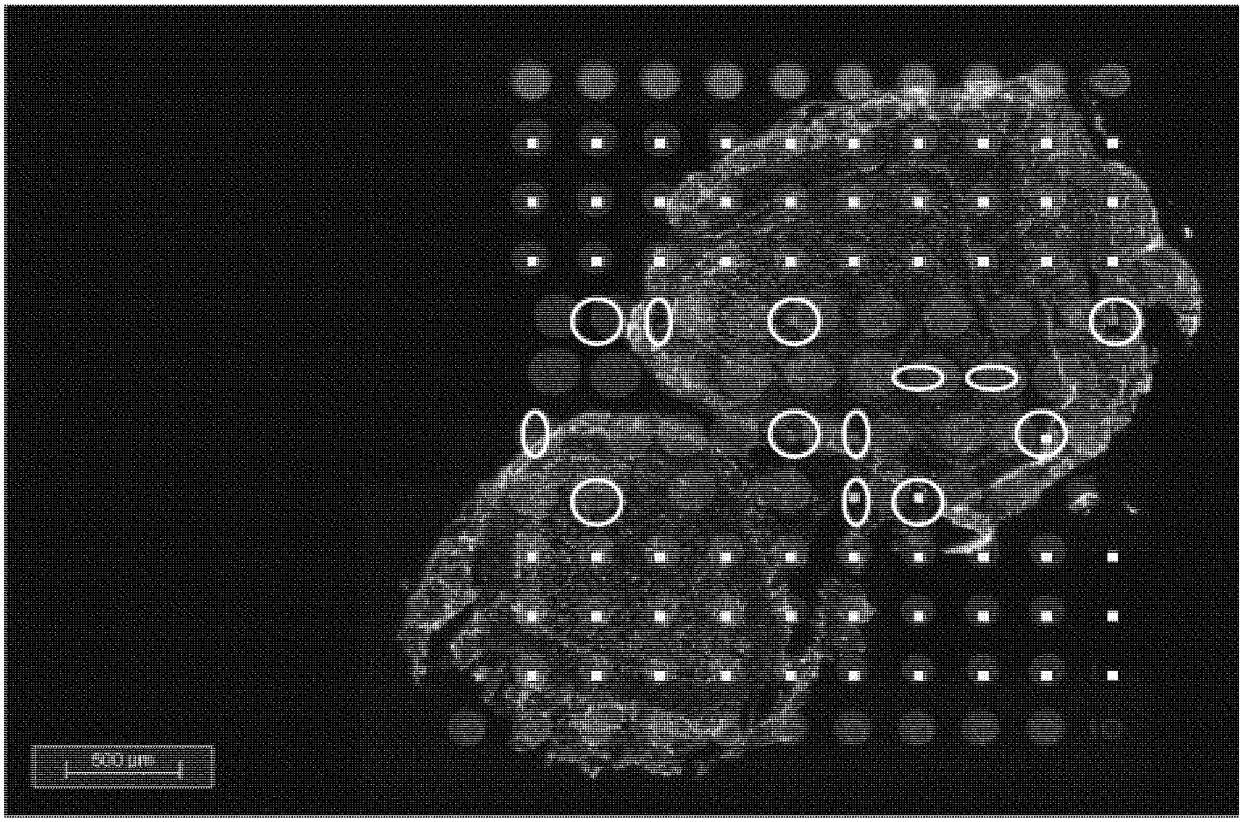

Amplification with A431_2+2_FW_A and RE_A primers
50/50
90/10
Positive control: 330bp
[FU]
[bp]
15
150
300
500
1500
100
200
40
20
0

Figure 31



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/056823

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	"Human Genome U95Av2", INTERNET CITATION, 2 October 2002 (2002-10-02), XP002215481, Retrieved from the Internet: URL:http://www.affymetrix.com [retrieved on 2002-10-02] abstract ----- "GeneChip Human Genome U133 Set", INTERNET CITATION, 26 February 2003 (2003-02-26), XP002232760, Retrieved from the Internet: URL:http://www.affymetrix.com/support/technical/datasheets/hgu133_datasheet.pdf [retrieved on 2003-02-26] abstract ----- -/--	37-41 37-41

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 June 2012	12/07/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Barz, Wolfgang

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CONSTANTINE L ET AL: "Use of genechip high-density oligonucleotide arrays for gene expression monitoring", LIFE SCIENCE NEWS, AMERSHAM LIFE SCIENCE, US, 1 January 1998 (1998-01-01), pages 11-14, XP002964122, ISSN: 0969-0190 abstract	37-41
A	US 2006/211001 A1 (YU JINGWEI [US] ET AL) 21 September 2006 (2006-09-21) abstract; examples 1-3 claims 1,5-6,13-14	1-43
A	US 2003/040035 A1 (SLAMON DENNIS J [US] ET AL) 27 February 2003 (2003-02-27) abstract; figures 1-4 claims 1-19	1-43
A	WO 2007/030373 A2 (ST JUDE CHILDRENS RES HOSPITAL [US]; CURRAN THOMAS [US]; RICE DENNIS S) 15 March 2007 (2007-03-15) abstract; claims 1-7	1-43
A	EGUILUZ C ET AL: "Multitissue array review: A chronological description of tissue array techniques, applications and procedures", PATHOLOGY RESEARCH AND PRACTICE, GUSTAV FISCHER, STUTTGART, DE, vol. 202, no. 8, 10 August 2006 (2006-08-10), pages 561-568, XP028031005, ISSN: 0344-0338, DOI: 10.1016/J.PRP.2006.04.003 [retrieved on 2006-08-10] the whole document	1-43
A	SIEVERTZON MARIA ET AL: "Transcriptome analysis in primary neural stem cells using a tag cDNA amplification method", BMC NEUROSCIENCE, BIOMED CENTRAL, LONDON, GB, vol. 6, no. 1, 15 April 2005 (2005-04-15), page 28, XP021002918, ISSN: 1471-2202, DOI: 10.1186/1471-2202-6-28 the whole document	1-43

-/--

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WANG D ET AL: "Single cell analysis: the new frontier in 'omics'", TRENDS IN BIOTECHNOLOGY, ELSEVIER PUBLICATIONS, CAMBRIDGE, GB, vol. 28, no. 6, 1 June 2010 (2010-06-01), pages 281-290, XP027059906, ISSN: 0167-7799 [retrieved on 2010-05-21] cited in the application the whole document	1-43
A	SCHENA M ET AL: "QUANTITATIVE MONITORING OF GENE EXPRESSION PATTERNS WITH A COMPLEMENTARY DNAMICROARRAY", SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, WASHINGTON, DC; US, vol. 270, no. 5235, 20 October 1995 (1995-10-20), pages 467-470, XP000644675, ISSN: 0036-8075, DOI: 10.1126/SCIENCE.270.5235.467 the whole document	1-43
A	WILLI-MONNERAT ET AL: "Comprehensive spatiotemporal transcriptomic analyses of the ganglionic eminences demonstrate the uniqueness of its caudal subdivision", MOLECULAR AND CELLULAR NEUROSCIENCES, SAN DIEGO, US, vol. 37, no. 4, 26 January 2008 (2008-01-26), pages 845-856, XP022576911, ISSN: 1044-7431 abstract; figures 1-3	1-43

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2012/056823

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006211001 A1	21-09-2006	NONE	
US 2003040035 A1	27-02-2003	US 2003040035 A1	27-02-2003
		US 2004171101 A1	02-09-2004
WO 2007030373 A2	15-03-2007	NONE	

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 April 2014 (24.04.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/060483 A1

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

(21) International Application Number:
PCT/EP2013/071645

(22) International Filing Date:
16 October 2013 (16.10.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1218654.0 17 October 2012 (17.10.2012) GB
1304585.1 14 March 2013 (14.03.2013) GB

(71) Applicant: **SPATIAL TRANSCRIPTOMICS AB** [SE/SE]; Bragevägen 5, SE-114 26 Stockholm (SE).

(72) Inventors: **FRISEN, Jonas**; Bragevägen 5, SE-114 26 Stockholm (SE). **STÅHL, Patrik**; Fatburstrappan 18, T3:1, SE-118 26 Stockholm (SE). **LUNDEBERG, Joakim**; Torsvikssvän 20, SE-181 34 Lidingö (SE). **SALMÉN, Fredrik**; Birger Jarlsgatan 104 F (4 tr), SE-114 20 Stockholm (SE).

(74) Agent: **DZIEGLEWSKA, Hanna**; Dehns, St Bride's House, 10 Salisbury Square, London Greater London EC4Y 8JD (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, 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:

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

WO 2014/060483 A1

(54) Title: METHODS AND PRODUCT FOR OPTIMISING LOCALISED OR SPATIAL DETECTION OF GENE EXPRESSION IN A TISSUE SAMPLE

(57) Abstract: The present invention relates to methods and products for localized or spatial detection and/or analysis of RNA in a tissue sample or a portion thereof, comprising: (a) providing an object substrate on which at least one species of capture probe, comprising a capture domain, is directly or indirectly immobilized such that the probes are oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer; (b) contacting said substrate with a tissue sample and allowing RNA of the tissue sample to hybridise to the capture probes; (c) generating cDNA molecules from the captured RNA molecules using said capture probes as RT primers; (d) labelling the cDNA molecules generated in step (c), wherein said labelling step may be contemporaneous with, or subsequent to, said generating step; (e) detecting a signal from the labelled cDNA molecules; and optionally (f) imaging the tissue sample, wherein the tissue sample is imaged before or after step (c).

Methods and product for optimising localised or spatial detection of gene expression in a tissue sample

The present invention relates generally to the localised or spatial detection of nucleic acid in a tissue sample or a portion thereof. Particularly, the present invention provides methods for detecting and/or analysing RNA, e.g. RNA transcripts, so as to obtain spatial information about the localisation, distribution or expression of genes in a tissue sample, for example in an individual cell. The present invention thus enables spatial transcriptomics.

More particularly, the present invention relates to a method for localised or spatial detection of transcripts in a tissue sample or a portion thereof, e.g. for determining and/or analysing a transcriptome, and especially the global transcriptome, of a tissue sample or a portion thereof. In particular the method relates to a quantitative and/or qualitative method for analysing the distribution, location or expression of nucleic acid molecules in a tissue sample wherein the spatial expression or distribution or location pattern within the tissue sample is retained. Thus, the method provides a process for performing "spatial transcriptomics", which enables the user to determine simultaneously the expression pattern, or the location/distribution pattern of the genes expressed, present in a tissue sample or a portion thereof. The present invention also provides methods for determining the optimum conditions for localized or spatial detection of (e.g. for capturing) nucleic acids from a tissue sample on a substrate, thereby allowing the maximum amount of nucleic acid molecules to be captured, whilst retaining the distribution or location pattern that originated within the tissue sample.

The invention is particularly based on array technology and may be coupled with high throughput DNA sequencing technologies. The methods of the invention allow the nucleic acid molecules (e.g. RNA molecules) in the tissue sample, particularly mRNA, to be captured on an object substrate (e.g. a slide or chip, which may be an array) and labelled, which may include the incorporation of a positional tag. The labelled molecules may be visualised to determine or assess the efficacy of the conditions used to capture the nucleic acid molecules. Alternatively or additionally, the captured nucleic acid molecules (or a subset thereof, e.g. a portion of the nucleic acid molecules captured from the tissue sample) may be analysed further, e.g. by sequence analysis. For instance, the captured nucleic acid molecules may be used to template the synthesis of DNA molecules which are sequenced and analysed to determine which genes are expressed in all or one or

more parts of the tissue sample. Advantageously, the individual, separate and specific transcriptome of each cell in the tissue sample may be obtained at the same time. Hence, the methods of the invention may be said to provide highly parallel comprehensive transcriptome signatures from individual cells (or groups of cells) within a tissue sample, or a portion thereof, without losing spatial information within said investigated tissue sample. The invention also provides an object substrate, such as a chip or slide (e.g. an array) for performing the method of the invention and methods for making the object substrate e.g. chip or slide, of the invention.

The human body comprises over 100 trillion cells and is organized into more than 250 different organs and tissues. The development and organization of complex organs, such as the brain, are far from understood and there is a need to dissect the expression of genes expressed in such tissues using quantitative methods to investigate and determine the genes that control the development and function of such tissues. The organs are in themselves a mixture of differentiated cells that enable all bodily functions, such as nutrient transport, defence etc. to be coordinated and maintained. Consequently, cell function is dependent on the position of the cell within a particular tissue structure and the interactions it shares with other cells within that tissue, both directly and indirectly. Hence, there is a need to disentangle how these interactions influence each cell within a tissue at the transcriptional level.

Recent findings by deep RNA sequencing have demonstrated that a majority of the transcripts can be detected in a human cell line and that a large fraction (75%) of the human protein-coding genes are expressed in most tissues. Similarly, a detailed study of 1% of the human genome showed that chromosomes are ubiquitously transcribed and that the majority of all bases are included in primary transcripts. The transcription machinery can therefore be described as promiscuous at a global level.

It is well-known that transcripts are merely a proxy for protein abundance, because the rates of RNA translation, degradation etc will influence the amount of protein produced from any one transcript. In this respect, a recent antibody-based analysis of human organs and tissues suggests that tissue specificity is achieved by precise regulation of protein levels in space and time, and that different tissues in the body acquire their unique characteristics by controlling not which proteins are expressed but how much of each is produced.

However, in subsequent global studies transcriptome and proteome correlations have been compared demonstrating that the majority of all genes were shown to be expressed. Interestingly, there was shown to be a high correlation between changes in RNA and protein levels for individual gene products which is indicative of the biological usefulness of studying the transcriptome in individual cells in the context of the functional role of proteins.

Indeed, analysis of the histology and expression pattern in tissues is a cornerstone in biomedical research and diagnostics. Histology, utilizing different staining techniques, first established the basic structural organization of healthy organs and the changes that take place in common pathologies more than a century ago. Developments in this field resulted in the possibility of studying protein distribution by immunohistochemistry and gene expression by *in situ* hybridization.

However, the parallel development of increasingly advanced histological and gene expression techniques has resulted in the separation of imaging and transcriptome analysis and, until the methods of the present invention, there has not been any feasible method available for global transcriptome analysis with spatial resolution.

As an alternative, or in addition, to *in situ* techniques, methods have developed for the *in vitro* analysis of proteins and nucleic acids, i.e. by extracting molecules from whole tissue samples, single cell types, or even single cells, and quantifying specific molecules in said extracts, e.g. by ELISA, qPCR etc.

Recent developments in the analysis of gene expression have resulted in the possibility of assessing the complete transcriptome of tissues using microarrays or RNA sequencing, and such developments have been instrumental in our understanding of biological processes and for diagnostics. However, transcriptome analysis typically is performed on mRNA extracted from whole tissues (or even whole organisms), and methods for collecting smaller tissue areas or individual cells for transcriptome analysis are typically labour intensive, costly and have low precision.

Hence, the majority of gene expression studies based on microarrays or next generation sequencing of RNA use a representative sample containing many cells. Thus the results represent the average expression levels of the investigated genes. The separation of cells that are phenotypically different has been used in some cases together with the global gene expression platforms (Tang F *et al*, Nat Protoc. 2010; 5: 516-35; Wang D & Bodovitz S, Trends Biotechnol. 2010; 28:281-

90) and resulted in very precise information about cell-to-cell variations. However, high throughput methods to study transcriptional activity with high resolution in intact tissues have not, until now, been available.

Thus, existing techniques for the analysis of gene expression patterns provide spatial transcriptional information only for one or a handful of genes at a time or offer transcriptional information for all of the genes in a sample at the cost of losing positional information. Hence, it is evident that methods to determine simultaneously, separately and specifically the transcriptome of each cell in a sample are required, i.e. to enable global gene expression analysis in tissue samples that yields transcriptomic information with spatial resolution, and the present invention addresses this need. The present invention may also be seen to provide alternative methods for the analysis of gene expression patterns that provide spatial transcriptional information for one or a handful of genes.

The novel approach of the methods and products of the present invention utilizes now conventional array technology and may utilise well established sequencing technology, which may yield transcriptional information for all of the genes in a sample, whilst retaining positional information for the transcripts. It will be evident to the person of skill in the art that this represents a milestone in the life sciences. The new technology opens a new field of so-called "spatial transcriptomics", which is likely to have profound consequences for our understanding of tissue development and tissue and cellular function in all multicellular organisms. It will be apparent that such techniques will be particularly useful in our understanding of the cause and progress of disease states and in developing effective treatments for such diseases, e.g. cancer. The methods of the invention will also find uses in the diagnosis of numerous medical conditions.

Array technology, particularly microarrays, arose from research at Stanford University where small amounts of DNA oligonucleotides were successfully attached to a glass surface in an ordered arrangement, a so-called "array", and used it to monitor the transcription of 45 genes (Schena M *et al*, Science. 1995; 270: 368-9, 371).

Since then, researchers around the world have published more than 30,000 papers using microarray technology. Multiple types of microarray have been developed for various applications, e.g. to detect single nucleotide polymorphisms (SNPs) or to genotype or re-sequence mutant genomes, and an important use of microarray technology has been for the investigation of gene expression. Indeed,

the gene expression microarray was created as a means to analyze the level of expressed genetic material in a particular sample, with the real gain being the possibility to compare expression levels of many genes simultaneously. Several commercial microarray platforms are available for these types of experiments but it has also been possible to create custom made gene expression arrays.

Whilst the use of microarrays in gene expression studies is now commonplace, it is evident that new and more comprehensive so-called "next-generation DNA sequencing" (NGS) technologies are starting to replace DNA microarrays for many applications, e.g. in-depth transcriptome analysis.

The development of NGS technologies for ultra-fast genome sequencing represents a milestone in the life sciences (Petterson E *et al*, *Genomics.* 2009; 93: 105-11). These new technologies have dramatically decreased the cost of DNA sequencing and enabled the determination of the genome of higher organisms at an unprecedented rate, including those of specific individuals (Wade CM *et al* Science. 2009; 326: 865-7; Rubin J *et al, Nature* 2010; 464: 587-91). The new advances in high-throughput genomics have reshaped the biological research landscape and in addition to complete characterization of genomes it is possible also to study the full transcriptome in a digital and quantitative fashion. The bioinformatics tools to visualize and integrate these comprehensive sets of data have also been significantly improved during recent years.

However, it has surprisingly been found that a unique combination of histological and microarray techniques, which may also be coupled with NGS techniques, can yield comprehensive transcriptional information from multiple cells in a tissue sample which information is characterised by a two-dimensional spatial resolution. Thus, at one extreme the methods of the present invention can be used to analyse the expression of a single gene in a single cell in a sample, whilst retaining the cell within its context in the tissue sample. At the other extreme, and in a preferred aspect of the invention, the methods can be used to determine the expression of every gene in each and every cell, or substantially all cells, in a tissue sample (or portion thereof) simultaneously, i.e. the global spatial expression pattern of a tissue sample or portion thereof. It will be apparent that the methods of the invention also enable intermediate analyses to be performed. For instance, the methods may be used to determine or quantify the transcriptional activity of cells in a tissue sample (e.g. the relative abundance of transcripts in different cell or tissue types), which would allow transcriptome analysis to focus on specific regions of a

tissue sample, e.g. regions or portions of tissues samples with high (or low) transcriptional activity.

It will be evident that the efficacy of the step of capturing nucleic acid molecules from tissue samples may be dependent on the source of the tissue and/or the methods used to prepare the tissue sample. Accordingly, the methods and arrays of the invention may be used to determine the optimum conditions to capture the nucleic acid molecules from a tissue sample on an object substrate, e.g. an array.

In its simplest form, the invention may be illustrated by the following summary. The invention requires reverse transcription (RT) primers to be immobilised on an object substrate, e.g. a glass slide or chip. Thin tissue sections are placed onto the substrate and a reverse transcription reaction is performed in the tissue section on the substrate. The RT primers, to which the RNA in the tissue sample binds (or hybridizes), are extended using the bound RNA as a template to obtain cDNA, which is therefore bound to the surface of the substrate. The synthesized part of the cDNA is labelled, e.g. with a visibly detectable label, such as a fluorescently labelled nucleotide which may be incorporated into the synthesized cDNA molecules. A consequence of labelling the synthesized cDNA is that each cDNA strand provides a detectable signal that corresponds to the presence of a RNA molecule in the tissue section, and the location of the cDNA on the surface of the substrate corresponds to its original location in the tissue section. The signal from the labelled cDNA is detected, e.g. the surface of the substrate is imaged, and accordingly, the relative abundance of transcript present in each part of the tissue section, e.g. each cell, can be detected and quantified. Optionally, the tissue section may be visualised or imaged, e.g. stained and photographed, before or after the cDNA synthesis step to enable the labelled cDNA molecule to be correlated with a position within the tissue sample.

Thus, in one aspect the method may be viewed as a method for capturing and/or labelling the transcriptome of a tissue sample on an object substrate and it will be evident from the discussion below that this method may find a variety of utilities, particularly in methods for detecting and/or analysing the transcriptome of a tissue sample or a portion thereof.

For instance, quantifying the relative abundance of the labelled cDNA may be used to determine the transcriptional activity of different regions of a tissue sample, e.g. to identify cells with high or no transcriptional activity. The method can

be used to determine the optimum reaction conditions for detecting, e.g. capturing, the transcriptome of a tissue sample on an object substrate, e.g. by repeating the method using different conditions (e.g. conditions for permeabilizing the tissue to allow the nucleic acids in the tissue sample to interact with the immobilized primers on the substrate), comparing the intensity and/or resolution of the signal obtained from labelled cDNA molecules on the imaged substrates and optionally selecting the conditions that provide the optimum image intensity and/or resolution.

In some embodiments, the method may be viewed as a method for localized or spatial detection, e.g. quantification, of the relative abundance of one or more transcripts in a tissue sample, e.g. the RT primer may be specific for one or more transcripts thereby enabling a specific transcript (or set of transcripts) to be captured and labelled on the object surface, wherein subsequent detection, e.g. imaging, of the intensity and distribution of the signal from labelled transcript is representative of the amount of transcript (or set of transcripts) present in the tissue sample in specific locations.

In a related embodiment the method may be viewed as a method for localized or spatial detection and/or analysis of the transcriptome of one or more portions of a tissue sample, e.g. the RT primer may be capable of capturing all of the transcripts on the object surface, which are then labelled as described above. Subsequent detection, e.g. imaging, of the intensity and distribution of the signal from labelled transcriptome can be used to select one or more portions of the substrate for further analysis, e.g. sequence analysis. The labelled transcripts on portions of the substrate that are not selected for further analysis may be removed from the surface of the substrate, e.g. by laser ablation, and discarded. The remaining subset of labelled transcripts may be analysed, e.g. sequenced, and the sequence information may be correlated with the portion(s) of the substrate from which the labelled transcripts were not removed and that correspond(s) to a position in the tissue sample.

In yet another aspect of the invention the reverse transcription (RT) primers comprise also unique positional tags (domains) and the RT primers may be arrayed on the object substrate, e.g. a glass slide or chip, to generate an "array". The unique positional tags correspond to the location of the RT primers on the array (the features of the array). Thin tissue sections are placed onto the array and a reverse transcription reaction is performed in the tissue section on the substrate. The RT primers, to which the RNA in the tissue sample binds (or hybridizes), are extended

using the bound RNA as a template to obtain cDNA, which is therefore bound to the surface of the array. The synthesized part of the cDNA is labelled, e.g. a detectable label, such as a fluorescently labelled nucleotide, may be incorporated into the synthesized cDNA. As consequence of the unique positional tags in the RT primers, each cDNA strand carries information about the position of the template RNA in the tissue section. The labelled cDNA is imaged, which enables the efficacy of the transcriptome capture step to be assessed. In some embodiments, the visualisation of the cDNA molecules also allows areas of the substrate to be targeted for removal of surface bound material, e.g. by laser ablation, such that only transcripts from portions of tissue sample that are of interest may be analysed further, as described above. The tissue section may be visualised or imaged, e.g. stained and photographed, before or after the cDNA synthesis step to enable the positional tag in the cDNA molecule to be correlated with a position within the tissue sample. The cDNA is sequenced, which results in a transcriptome with exact positional information. The sequence data can then be matched to a position in the tissue sample, which enables the visualization, e.g. using a computer, of the sequence data together with the tissue section, for instance to display the expression pattern of any gene of interest across the tissue. Similarly, it would be possible to mark different areas of the tissue section on the computer screen and obtain information on differentially expressed genes between any selected areas of interest. It will be evident that the methods of the invention may result in data that is in stark contrast to the data obtained using current methods to study mRNA populations. For example, methods based on *in situ* hybridization provide only relative information of single mRNA transcripts. Thus, the methods of the present invention have clear advantages over current *in situ* technologies. The global gene expression information obtainable from the methods of the invention also allows co-expression information and quantitative estimates of transcript abundance. It will be evident that this is a generally applicable strategy available for the analysis of any tissue in any species, e.g. animal, plant, fungus.

It will be seen from the above explanation that there is an immense value in coupling positional information to transcriptome information. For instance, it enables global gene expression mapping at high resolution, which will find utility in numerous applications, including e.g. cancer research and diagnostics.

Furthermore, it is evident that the methods described herein differ significantly from the previously described methods for analysis of the global

transcriptome of a tissue sample and these differences result in numerous advantages. The present invention is predicated on the surprising discovery that the use of tissue sections does not interfere with synthesis of DNA (e.g. cDNA) primed by primers (e.g. reverse transcription primers) that are coupled to the surface of an object substrate, e.g. an array. Moreover, labelling the cDNA synthesized on the object substrate allows specific portions of a tissue sample to be selected for further analysis, which enables resources, e.g. sequencing resources, to be focussed on the analysis of specific cell or tissue types within a tissue sample. This may result in reduced costs and a less complex data set which may be analysed more efficiently and robustly than a data set from the analysis of the transcriptome of the whole tissue sample.

Thus, in its first and broadest aspect, the present invention provides a method for localized or spatial detection and/or analysis of RNA, and particularly of transcripts, in a tissue sample or a portion thereof, comprising:

(a) providing an object substrate on which at least one species of capture probe is directly or indirectly immobilized such that the probes are oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer;

(b) contacting said substrate with a tissue sample and allowing RNA of the tissue sample to hybridise to the capture probes;

(c) generating cDNA molecules from the captured RNA molecules using said capture probes as RT primers;

(d) labelling the cDNA molecules generated in step (c), wherein said labelling step may be contemporaneous with, or subsequent to, said generating step, preferably wherein the label is incorporated into the synthesized part of the cDNA molecules;

(e) detecting a signal from the labelled cDNA molecules, e.g. imaging the substrate such that the signal from the labelled cDNA molecules is detected; and optionally

(f) imaging the tissue sample, wherein the tissue sample is imaged before or after step (c).

The method may alternatively or additionally be viewed as a method for determining and/or analysing a transcriptome of a tissue sample or a portion thereof or a method for capturing and/or labelling the transcriptome of a tissue sample on an object substrate, e.g. an array.

Thus, in a second more particular aspect of the invention, the present invention can be seen to provide a method for determining the optimum conditions for localised or spatial detection of RNA (e.g. transcripts) in a tissue sample on an object substrate, comprising steps (a)-(e), and optionally step (f), described above on a first object substrate and further steps:

(g) repeating steps (a)-(e), and optionally step (f), with a second object substrate, using different conditions in step (b);

(h) comparing the intensity and/or resolution of the signal from the labelled cDNA molecules immobilized on said first and second object substrate; and optionally

(i) selecting the conditions that provide the optimum signal intensity and/or resolution of the labelled cDNA molecules.

In a third aspect, the present invention can be seen to provide a method for determining and/or analysing RNA or a transcriptome of a tissue sample or a portion thereof comprising steps (a)-(e), and optionally step (f), described above and further steps:

(g') removing the labelled cDNA from at least one portion of the surface of the object substrate;

(h') optionally amplifying the remaining cDNA molecules immobilized on the surface of the object substrate;

(i') releasing at least part of the remaining cDNA molecules and/or optionally their amplicons from the surface of the object substrate, wherein said released molecules may be a first strand and/or second strand cDNA molecule or an amplicon thereof;

(j') directly or indirectly analysing the sequence of the released molecules.

It will be understood that this third aspect allows a part of the RNA or transcriptome to be determined and/or analysed, and in particular to be selectively analysed; by removing the labelled cDNA from at least a portion of the object substrate surface, it may be selected which of the cDNA (and hence which of the captured RNA) to be analysed. The removal step is discussed further below, but it will be understood that this could be achieved by removing a portion of the tissue sample from the substrate (which will concomitantly remove the labelled cDNA)

In a particularly preferred embodiment of the third aspect of the invention, the object substrate is an array and the capture probes each comprise a positional

domain that corresponds to the position of each capture probe on the array. Accordingly, the method may be viewed as comprising:

(a") providing an object substrate (e.g. an array) on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the object substrate and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the object substrate, and

(ii) a capture domain;

(b") contacting said object substrate with a tissue sample such that the position of a capture probe on the object substrate may be correlated with a position in the tissue sample and allowing RNA of the tissue sample to hybridise to the capture domain in said capture probes;

(c") generating cDNA molecules from the captured RNA molecules using said capture probes as RT primers,

(d") labelling the cDNA molecules generated in step (c'), wherein said labelling step may be contemporaneous with, or subsequent to, said generating step, preferably wherein the label is incorporated into the synthesized part of the cDNA molecules;

(e") detecting a signal from the labelled cDNA molecules;

(f") optionally imaging the tissue sample, wherein the tissue sample is imaged before or after step (c").

(g") optionally removing the labelled cDNA from at least one portion of the surface of the object substrate;

(h") optionally amplifying the cDNA molecules immobilized on the surface of the object substrate;

(i") releasing at least part of the cDNA molecules and/or optionally their amplicons from the surface of the object substrate, wherein said released molecules may be a first strand and/or second strand cDNA molecule or an amplicon thereof and wherein said part includes the positional domain or a complement thereof;

(j") directly or indirectly analysing the sequence of the released molecules.

The methods allow the abundance of the transcripts from a tissue sample to visualised directly, e.g. by fluorescence, akin to a standard microarray. However, unlike a standard microarray, the abundance of the transcripts can be correlated directly with their position in the tissue sample. Advantageously, the detection of the labelled cDNA molecules *in situ* on the surface of the object substrate (i.e. such that their distribution the object substrate corresponds directly their distribution in the tissue sample) also allows for a portion or portions of the tissue sample of interest to be selected for analysis, thereby minimising the amount of data to be evaluated. Furthermore, the detection of the labelled cDNA molecules *in situ* on the surface of the object substrate enables the provision of a method for determining the optimum conditions for capturing the transcriptome of a tissue sample on an object substrate, wherein the spatial distribution of the transcripts in the tissue sample is transferred directly to the surface of the object substrate. The object substrates used for optimising the conditions for capturing a transcriptome (e.g. wherein the capture probes do not contain a positional domain and/or the probes are immobilized uniformly on the surface of the object substrate) are relatively inexpensive in comparison to an object substrate on which capture probes comprising a positional domain are arrayed, such that each feature comprises a species of capture probe with a unique positional domain. However, the optimum conditions determined using the inexpensive object substrate can be used to perform analyses using the expensive arrays. Hence the methods may also be seen to reduce costs over other spatial transcriptomics methods. Thus, this aspect of the method may be viewed as providing or enabling a so-called "quality control" (QC) step to determine the optimum or most appropriate, e.g. cost-effective, conditions in which to perform a more detailed analysis using more expensive arrays.

The methods of the invention also represent a significant advance over other methods for spatial transcriptomics known in the art. For example the methods described herein may result in a global and spatial profile of all transcripts in the tissue sample or a portion thereof. Moreover, the methods may enable the expression of every gene to be quantified for each position or feature on an array, which enables a multiplicity of analyses to be performed based on data from a single assay. Thus, the methods of the present invention make it possible to detect and/or quantify the spatial expression of all genes in single tissue sample or a portion thereof. In some aspects of the invention the abundance of the transcripts also may be visualised both directly and indirectly. When the methods include

methods of indirect detection, e.g. sequence analysis, it is possible to measure the expression of genes in a single sample simultaneously even wherein said transcripts are present at vastly different concentrations in the same sample.

As described in more detail below, any method of nucleic acid analysis may be used in the analysis step (j). Typically this may involve sequencing, but it is not necessary to perform an actual sequence determination. For example sequence-specific methods of analysis may be used. For example a sequence-specific amplification reaction may be performed, for example using primers which are specific for the positional domain and/or for a specific target sequence, e.g. a particular target DNA to be detected (i.e. corresponding to a particular cDNA/RNA etc). An exemplary analysis method is a sequence-specific PCR reaction.

The sequence analysis information obtained in step (j) may be used to obtain spatial information as to the RNA in the sample or a portion of the sample. In other words the sequence analysis information may provide information as to the location of the RNA in the tissue sample. This spatial information may be derived from the nature of the sequence analysis information determined, for example it may reveal the presence of a particular RNA which may itself be spatially informative in the context of the tissue sample used, and/or the spatial information (e.g. spatial localisation) may be derived from the position of the tissue sample on the array, coupled with the sequencing information. Thus, the method may involve simply correlating the sequence analysis information to a position in the tissue sample e.g. by virtue of the positional tag and its correlation to a position in the tissue sample. However, as described above, spatial information may conveniently be obtained by correlating the expression data, e.g. the intensity of the signal from the labelled cDNA detected, e.g. imaged, in step (e) or the sequence analysis data obtained from step (j), to an image of the tissue sample and this represents one preferred embodiment of the invention. Accordingly, in a preferred embodiment the method also includes a step of:

(f) imaging the tissue sample wherein the tissue sample is imaged before or after step (c), preferably wherein the image of the labelled cDNA is correlated with an image of said tissue sample.

Hence, the method described in the third aspect of the invention may comprise a step of:

(k) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (c).

In its broadest sense, the method of the invention may be used for localized or spatial detection of a nucleic acid, specifically RNA, in a tissue sample. Thus, in one embodiment, the method of the invention may be used for determining and/or analysing all of the transcriptome of a tissue sample e.g. the global transcriptome of a tissue sample. However, the method is not limited to this and encompasses determining and/or analysing all or part of the transcriptome of a tissue sample or a portion thereof. Thus, the method may involve determining and/or analysing a part or subset of the transcriptome, e.g. a transcriptome corresponding to one gene or a subset of genes, e.g. a set of particular genes, for example related to a particular disease or condition, tissue type etc. Alternatively or additionally, the method may involve determining and/or analysing all of the transcriptome of a portion of a tissue sample.

Viewed from another aspect, the method steps set out above can be seen as providing a method of obtaining a spatially defined transcriptome, and in particular the spatially defined global transcriptome, of a tissue sample or portion thereof.

Alternatively viewed, the method of the invention may be seen as a method for localised or spatial detection of nucleic acid, e.g. RNA, in a tissue sample or a portion thereof, or for localised or spatial determination and/or analysis of nucleic acid (e.g. RNA) in a tissue sample or a portion thereof. In particular, the method may be used for the localised or spatial detection or determination and/or analysis of gene expression in a tissue sample or a portion thereof. The localised/spatial detection/determination/analysis means that the RNA may be localised to its native position or location within a cell or tissue in the tissue sample. Thus for example, the RNA may be localised to a cell or group of cells, or type of cells in the sample, or to particular regions of areas within a tissue sample. The native location or position of the RNA (or in other words, the location or position of the RNA in the tissue sample), e.g. an expressed gene, may be determined.

The invention may also be viewed as providing methods for determining the optimum conditions for localised detection and/or analysis of nucleic acids in a tissue sample on an object substrate, i.e. for capturing and/or labelling nucleic acids from a tissue sample on an object substrate, e.g. a slide or chip.

The invention can also be seen to provide an object substrate, e.g. a slide or chip, for use in the methods of the invention comprising a substrate on which one or more species of capture probe is directly or indirectly immobilized such that each

probe is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer.

Optionally the probes are immobilised uniformly on the object substrate, i.e. the probes are not arrayed as distinct features. Hence, in some embodiments the object substrate of the invention may be viewed as a featureless array, i.e. a microarray without distinct features, which are defined below. In a particular embodiment of the invention, one species of capture probe is immobilized on the object surface, i.e. the capture probes are identical.

In some embodiments of the invention the probes are capable of hybridizing to (i.e. capturing) all mRNA, i.e. RNA molecules with a polyA tail. Hence, in particularly preferred embodiments of the invention the probes may comprise sequences of consecutive dTTP or dUTP nucleotides, e.g. oligoT and/or oligoU, as described in more detail below. In a preferred embodiment of the invention, the object substrate of the invention is for use in methods for determining the optimum conditions for the localised or spatial detection of transcripts from a tissue sample on an object substrate, e.g. array.

In some embodiments of the invention the probes may be capable of hybridizing to (i.e. capturing) specific types of mRNA, i.e. RNA expressed from a specific gene of set of genes. Hence, in some embodiments of the invention the probes may comprise gene specific sequences or sequences that are degenerate for a family of genes. In a preferred embodiment of the invention, the object substrate of the invention is for use in methods for determining and/or analysing one or more transcripts from a tissue sample or a portion thereof.

It will be seen therefore that the object substrate of the present invention may be used to capture RNA, e.g. mRNA, of a tissue sample that is contacted with said array. The array may also be used for determining and/or analysing a partial or global transcriptome of a tissue sample or for obtaining a spatially defined partial or global transcriptome of a tissue sample. The object substrate may be for use in methods of the invention that may be considered as methods of determining, e.g. quantifying, the localised or spatial expression of one or more genes in a tissue sample or portion thereof. Expressed another way, the object substrate may be for use in methods used to detect the spatial expression of one or more genes in a tissue sample or portion thereof. In yet another way, the object substrate may be for use in methods used to determine simultaneously the expression of one or more genes at one or more positions within a tissue sample or a portion thereof. Still

further, the object substrate may be for use in methods for partial or global transcriptome analysis of a tissue sample or portion thereof with two-dimensional spatial resolution.

The RNA may be any RNA molecule which may occur in a cell. Thus it may be mRNA, tRNA, rRNA, viral RNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), microRNA (miRNA), small interfering RNA (siRNA), piwi-interacting RNA (piRNA), ribozymal RNA, antisense RNA or non-coding RNA. Preferably however it is mRNA.

Step (c) in the methods above of generating cDNA from the captured RNA will be seen as relating to the synthesis of the cDNA. This will involve a step of reverse transcription of the captured RNA, extending the capture probe, which functions as the RT primer, using the captured RNA as template. Such a step generates so-called first strand cDNA. As will be described in more detail below, second strand cDNA synthesis may optionally take place on the array, or it may take place in a separate step, after release of first strand cDNA from the array. As also described in more detail below, in certain embodiments second strand synthesis may occur in the first step of amplification of a released first strand cDNA molecule.

Step (d) in the methods above of labelling the cDNA molecules generated in step (c) will be seen as relating to any suitable method of labelling the cDNA molecules. In particular, the label will be provided to, e.g. as part of, or on, the generated cDNA molecule. In preferred embodiments of the invention, the labelling step is performed contemporaneously with, i.e. at the same time as, the generating step. Thus, labelling the may involve the incorporation of labelled nucleotides into the synthesized cDNA molecule directly. As discussed in more detail below, the nucleotides may be labelled with directly or indirectly signal giving molecules. For instance, directly signal giving labels may be fluorescent molecules, i.e. the labelled nucleotides may be fluorescently labelled nucleotides. Indirectly signal giving labels may be, for example, biotin molecules, i.e. the labelled nucleotides may be biotin labelled nucleotides, which require additional steps to provide a signal, e.g. the addition of streptavidin conjugated to an enzyme which may act on a chemical substrate to provide a detectable signal, e.g. a visibly detectable colour change.

In some embodiments the cDNA generated in step (c) may be labelled after its synthesis, e.g. stained. Various methods for labelling nucleic acid molecules are known in the art and could be employed in the methods of the invention. For

instance, the cDNA may be stained with a nucleic acid stain. If the cDNA generating step only creates a first strand of cDNA, it may be advantageous to use a strain capable of detecting single stranded nucleic acid such as SYBR Gold® or GelStar®, as the RNA template may degrade. However, if a second strand of cDNA is generated a stain capable of detecting double stranded nucleic acid, such as ethidium bromide or SYBR Green® may be used. In some embodiments it may be advantageous to remove the tissue sample from the object substrate before labelling the cDNA, e.g. to avoid background signals from nucleic acid material remaining in the tissue sample, e.g. genomic DNA etc. In embodiments where it is desirable to image the tissue sample to be able to correlate the detection, e.g. image, of the labelled cDNA with an image of the tissue sample, it may be desirable to image the tissue sample before the step of generating the cDNA and particularly before the step of labelling the cDNA.

The "object substrate" or "substrate" of the invention may be any solid substrate on which nucleic acid molecules can be immobilized directly or indirectly, e.g. a slide or chip. In preferred embodiments the object substrate may be viewed as being an array substrate, i.e. any substrate that could be used to generate a nucleic acid array, e.g. a microarray substrate. In many embodiments the capture probes may be immobilized on the object substrate in the form of an array, i.e. in some embodiments the object substrate is an array, e.g. a microarray. In other embodiments the capture probes are not immobilized on the object substrate in an array format, i.e. the substrate may have no distinct features and the capture probes may be immobilized on the object substrate uniformly. Hence, in some embodiments the object substrate is not an array. Alternatively, the object subject may be viewed as a featureless array or alternatively as an array comprising a single large feature. Array substrates, i.e. object substrates, for use in the context of nucleic acid analysis are discussed and described below.

As used herein the term "multiple" means two or more, or at least two, e.g. 3, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 400, 500, 1000, 2000, 5000, 10,000, or more etc. Thus for example, the number of distinct capture probes (i.e. capture probes with different sequences, e.g. different positional domains) may be any integer in any range between any two of the aforementioned numbers. It will be appreciated however that it is envisaged that conventional-type arrays with many hundreds, thousands, tens of thousands, hundreds of thousands or even millions of capture probes may be used.

Thus, the methods outlined herein may utilise, but are not limited to, high density nucleic acid arrays comprising "capture probes" for capturing and labelling transcripts from all of the single cells within a tissue sample e.g. a thin tissue sample slice, or "section". The tissue samples or sections for analysis are produced in a highly parallelized fashion, such that the spatial information in the section is retained. The captured RNA (preferably mRNA) molecules for each cell, or "transcriptomes", are transcribed into cDNA and the resultant cDNA molecules are labelled and detected and/or analyzed. In the first instance the labelled cDNA molecules are detected and/or analysed by detecting the signal from the labelled molecules, e.g. by imaging the object substrate such that the signal from the label can be detected, e.g. quantified. The labelled cDNA molecules, or a portion thereof, may be subject to analysis, e.g. by high throughput sequencing. The resultant data from the first and/or subsequent detection and/or analysis may be correlated to images of the original tissue samples. For instance, an overlay of the two images may be used to determine areas of high, low or no expression, which may enable portions of the tissue sample to be selected for further analysis. The data from the further analyses may be correlated to images of the tissue sample by, e.g. so-called barcode sequences (or ID tags, defined herein as positional domains) incorporated into the arrayed nucleic acid probes.

High density nucleic acid arrays or microarrays form a core component of some of the spatial transcriptome labelling methods described herein. A microarray is a multiplex technology used in molecular biology. A typical microarray consists of an arrayed series of microscopic spots of oligonucleotides (hundreds of thousands of spots, generally tens of thousands, can be incorporated on a single array). The distinct position of each nucleic acid (oligonucleotide) spot (each species of oligonucleotide/nucleic acid molecule) is known as a "feature" (and hence in some of the methods set out above each species of capture probe may be viewed as a specific feature of the array; each feature occupies a distinct position on the array), and typically each separate feature contains in the region of picomoles (10^{-12} moles) of a specific DNA sequence (a "species"), which are known as "probes" (or "reporters"). Typically, these can be a short section of a gene or other nucleic acid element to which a cDNA or cRNA sample (or "target") can hybridize under high-stringency hybridization conditions. However, as described below, the probes of the present invention and/or their distribution on the array may differ from the probes of standard microarrays.

In gene expression microarrays, probe-target hybridization is usually detected and quantified by detection of visual signal, e.g. a fluorophore, silver ion, or chemiluminescence-label, which has been incorporated into all of the targets before the targets are contacted with the array. The intensity of the visual signal correlates to the relative abundance of each target nucleic acid in the sample, but does not provide any spatial information about the origin of the target nucleic acid in the sample. Since an array can contain tens of thousands of probes, a microarray experiment can accomplish many genetic tests in parallel.

In standard microarrays, the probes are attached to a solid surface or substrate by a covalent bond to a chemical matrix, e.g. epoxy-silane, amino-silane, lysine, polyacrylamide etc. The substrate typically is a glass, plastic or silicon chip or slide, although other microarray platforms are known, e.g. microscopic beads.

The probes may be attached to the object substrate, e.g. array, of the invention by any suitable means. In a preferred embodiment the probes are immobilized to the substrate by chemical immobilization. This may be an interaction between the substrate (support material) and the probe based on a chemical reaction. Such a chemical reaction typically does not rely on the input of energy via heat or light, but can be enhanced by either applying heat, e.g. a certain optimal temperature for a chemical reaction, or light of certain wavelength. For example, a chemical immobilization may take place between functional groups on the substrate and corresponding functional elements on the probes. Such corresponding functional elements in the probes may either be an inherent chemical group of the probe, e.g. a hydroxyl group or be additionally introduced. An example of such a functional group is an amine group. Typically, the probe to be immobilized comprises a functional amine group or is chemically modified in order to comprise a functional amine group. Means and methods for such a chemical modification are well known.

The localization of said functional group within the probe to be immobilized may be used in order to control and shape the binding behaviour and/or orientation of the probe, e.g. the functional group may be placed at the 5' or 3' end of the probe or within sequence of the probe. A typical substrate for a probe to be immobilized comprises moieties which are capable of binding to such probes, e.g. to amine-functionalized nucleic acids. Examples of such substrates are carboxy, aldehyde or epoxy substrates. Such materials are known to the person skilled in the art. Functional groups, which impart a connecting reaction between probes that are

chemically reactive by the introduction of an amine group, and array substrates are known to the person skilled in the art.

Alternative substrates on which probes may be immobilized may have to be chemically activated, e.g. by the activation of functional groups, available on the object substrate, e.g. array substrate. The term "activated substrate" relates to a material in which interacting or reactive chemical functional groups were established or enabled by chemical modification procedures as known to the person skilled in the art. For example, a substrate comprising carboxyl groups has to be activated before use. Furthermore, there are substrates available that contain functional groups that can react with specific moieties already present in the nucleic acid probes.

In some embodiments the probes may be immobilized on beads, e.g. plastic microbeads, which can be used to immobilize the probes on the substrate. Suitable techniques for immobilizing the nucleic acid molecules on beads may be selected from the techniques discussed above or selected from methods known in the art. The beads may be contacted with, and immobilized on, an object substrate, thereby indirectly immobilizing the probes on the surface of the substrate. For example, after contacting the beads with the substrate, the substrate may be treated crosslink the beads to each other and/or the surface of the substrate, e.g. the substrate may heated to partially melt the beads which are allowed to solidify, to generate an object substrate on which probes are indirectly immobilized.

Alternatively, the probes may be synthesized directly on the substrate. Suitable methods for such an approach are known to the person skilled in the art. Examples are manufacture techniques developed by Agilent Inc., Affymetrix Inc., Roche Nimblegen Inc. or Flexgen BV. Typically, lasers and a set of mirrors that specifically activate the spots where nucleotide additions are to take place are used. Such an approach may provide, for example, spot sizes (i.e. features) of around 30 µm or larger. However, in some embodiments the probes may be immobilized uniformly on the substrate, i.e. a uniform, consistent or homogeneous distribution of probes across the surface of the substrate. Hence, it may be necessary simply to activate a portion or area of the substrate on which the probes will be immobilized. A "portion" of the substrate is described below.

The object substrate therefore may be any suitable substrate known to the person skilled in the art. The substrate may have any suitable form or format, e.g. it may be flat, curved, e.g. convexly or concavely curved towards the area where the

interaction between the tissue sample and the substrate takes place. Particularly preferred is the where the substrate is a flat, i.e. planar, such as a chip or slide.

Typically, the substrate is a solid support and thereby allows for an accurate and traceable positioning of the probes on the substrate. An example of a substrate is a solid material or a substrate comprising functional chemical groups, e.g. amine groups or amine-functionalized groups. A substrate envisaged by the present invention is a non-porous substrate. Preferred non-porous substrates are glass, silicon, poly-L-lysine coated material, nitrocellulose, polystyrene, cyclic olefin copolymers (COCs), cyclic olefin polymers (COPs), polypropylene, polyethylene and polycarbonate.

Any suitable material known to the person skilled in the art may be used. Typically, glass or polystyrene is used. Polystyrene is a hydrophobic material suitable for binding negatively charged macromolecules because it normally contains few hydrophilic groups. For nucleic acids immobilized on glass slides, it is furthermore known that by increasing the hydrophobicity of the glass surface the nucleic acid immobilization may be increased. Such an enhancement may permit a relatively more densely packed formation, which is advantageous when the probes are arranged in an array format. In addition to a coating or surface treatment with poly-L-lysine, the substrate, in particular glass, may be treated by silanation, e.g. with epoxy-silane or amino-silane or by silynation or by a treatment with polyacrylamide.

A number of standard arrays and array substrates are commercially available and both the number and size of the features may be varied. In the present invention, when the probes are distributed uniformly on the surface of the substrate, the concentration of the probes immobilized may be altered to correspond to the size and/or density of the cells present in different tissues or organisms. Similarly, when the probes are in an array format the arrangement of the features may be altered to correspond to the size and/or density of the cells present in different tissues or organisms. For instance, animal cells typically have a cross-section in the region of 1-100µm, whereas the cross-section of plant cells typically may range from 1-10000µm. Hence, in embodiments where the probes are arrayed on the substrate, Nimblegen® arrays, which are available with up to 2.1 million features, or 4.2 million features, and feature sizes of 13 micrometers, may be preferred for tissue samples from an animal or fungus, whereas other formats, e.g. with 8x130k features, may be sufficient for plant tissue samples. Commercial

arrays are also available or known for use in the context of sequence analysis and in particular in the context of NGS technologies. Such arrays may also be used as the substrate, e.g. array substrate in the context of the present invention, e.g. an Illumina bead array. In addition to commercially available arrays, which can themselves be customized, it is possible to make custom or non-standard "in-house" arrays and methods for generating arrays are well-established. The methods of the invention may utilise both standard and non-standard arrays that comprise probes as defined below.

The probes on a substrate may be immobilized, i.e. attached or bound, to the substrate, e.g. array, via the 5' or 3' end, depending on the chemical matrix of the array. Typically, for commercially available arrays, the probes are attached via a 3' linkage, thereby leaving a free 5' end. However, substrates, e.g. arrays, comprising probes attached to the substrate via a 5' linkage, thereby leaving a free 3' end, are available and may be synthesized using standard techniques that are well known in the art and are described elsewhere herein.

The covalent linkage used to couple a nucleic acid probe to a substrate may be viewed as both a direct and indirect linkage, in that the although the probe is attached by a "direct" covalent bond, there may be a chemical moiety or linker separating the "first" nucleotide of the nucleic acid probe from the, e.g. glass or silicon, substrate i.e. an indirect linkage. For the purposes of the present invention probes that are immobilized to the substrate by a covalent bond and/or chemical linker are generally seen to be immobilized or attached directly to the substrate.

As will be described in more detail below, the capture probes of the invention may be immobilized on, or interact with, the substrate, e.g. array, directly or indirectly. Thus the capture probes need not bind directly to the substrate, but may interact indirectly, for example by binding to a molecule which itself binds directly or indirectly to the array (e.g. the capture probe may interact with (e.g. bind or hybridize to) a binding partner for the capture probe, i.e. a surface probe, which is itself bound to the substrate directly or indirectly). Generally speaking, however, the capture probe will be, directly or indirectly (by one or more intermediaries), bound to, or immobilized on, the substrate.

The method and object substrate, e.g. slide or chip, of the invention may comprise probes that are immobilized via their 5' or 3' end. However, when the capture probe is immobilized directly to the array substrate, it may be immobilized only such that the 3' end of the capture probe is free to be extended, e.g. it is

immobilized by its 5' end. The capture probe may be immobilized indirectly, such that it has a free, i.e. extendible, 3' end.

By extended or extendible 3' end, it is meant that further nucleotides may be added to the most 3' nucleotide of the nucleic acid molecule, e.g. capture probe, to extend the length of the nucleic acid molecule, i.e. the standard polymerization reaction utilized to extend nucleic acid molecules, e.g. templated polymerization catalyzed by a polymerase.

Thus, in one embodiment, the substrate, e.g. array, comprises probes that are immobilized directly via their 3' end, so-called surface probes, which are defined below. Each species of surface probe comprises a region of complementarity to each species of capture probe, such that the capture probe may hybridize to the surface probe, resulting in the capture probe comprising a free extendible 3' end. In a preferred aspect of the invention, when the substrate comprises surface probes, the capture probes are synthesized *in situ* on the substrate.

The probes may be made up of ribonucleotides and/or deoxyribonucleotides as well as synthetic nucleotide residues that are capable of participating in Watson-Crick type or analogous base pair interactions. Thus, the nucleic acid domain may be DNA or RNA or any modification thereof, e.g. PNA or other derivatives containing non-nucleotide backbones. However, the capture probe, e.g. the capture domain of the capture probe, must capable of priming a reverse transcription reaction to generate cDNA that is complementary to the captured RNA molecules.

In a preferred embodiment of the invention at least the capture domain of the capture probe comprises or consists of deoxyribonucleotides (dNTPs). In a particularly preferred embodiment the whole of the capture probe comprises or consists of deoxyribonucleotides.

In a preferred embodiment of the invention the capture probes are immobilized on the substrate directly, i.e. by their 5' end, resulting in a free extendible 3' end.

The capture probes of the invention comprise at least one domain, a capture domain, which is capable of interacting with (i.e. binding or hybridizing to) the RNA from the tissue, i.e. to capture the RNA. In some embodiments in which the capture probes are arrayed on the substrate, the probes preferably comprise at least two domains, a capture domain and a positional domain (or a feature identification tag or domain; the positional domain may alternatively be defined as an identification (ID) domain or tag, or as a positional tag). The capture probe may further comprise

a universal domain as defined further below. Where the capture probe is indirectly attached to the array surface via hybridization to a surface probe, the capture probe requires a sequence (e.g. a portion or domain) which is complementary to the surface probe. Such a complementary sequence may be complementary to a positional/identification domain (if present in the capture probe) and/or a universal domain on the surface probe. In other words the positional domain and/or universal domain may constitute the region or portion of the probe which is complementary to the surface probe. However, the capture probe may also comprise an additional domain (or region, portion or sequence) which is complementary to the surface probe. For ease of synthesis, as described in more detail below, such a surface probe-complementary region may be provided as part of, or as an extension of, the capture domain (such a part or extension not itself being used for, or capable of, binding to the target nucleic acid, e.g. RNA).

Thus, in their simplest form the capture probes for use in the invention may comprise or consist of a capture domain. However, in some embodiments of the invention the capture probes may comprise or consist of: (i) a capture domain and a positional domain; (ii) a capture domain and a universal domain; (iii) a capture domain and a domain that is complementary to a surface probe; (iv) a capture domain, positional domain and a universal domain, and so forth.

In some embodiments a single species of capture probe is immobilized to the substrate, preferably wherein the capture probe is immobilized on the substrate uniformly. However, a single species of capture probe may be arrayed on the substrate, such that each feature on the array comprises the same probe. In some embodiments multiple species of capture probe are immobilized to the substrate, preferably wherein each species of capture probe is immobilized at a different position on the substrate (i.e. each species forms a feature in an array), although a single capture probe may in some embodiments be immobilized at more than one position (a single species of capture probe may be used to form more than one feature). In some embodiments multiple species of capture probe may be combined to form a mixture which is immobilized on the substrate uniformly, i.e. such that there is an even distribution of each species of capture probe on the surface of the substrate.

The capture domain is typically located at the 3' end of the capture probe and comprises a free 3' end that can be extended, e.g. by template dependent polymerization. The capture domain comprises a nucleotide sequence that is

capable of hybridizing to nucleic acid, e.g. RNA (preferably mRNA) present in the cells of the tissue sample contacted with the array.

Advantageously, the capture domain may be selected or designed to bind (or put more generally may be capable of binding) selectively or specifically to the particular nucleic acid, e.g. RNA, it is desired to detect or analyse. For example the capture domain may be selected or designed for the selective capture of mRNA. As is well known in the art, this may be on the basis of hybridisation to the poly-A tail of mRNA. Thus, in a preferred embodiment the capture domain comprises a poly-T DNA oligonucleotide, i.e. a series of consecutive deoxythymidine residues linked by phosphodiester bonds, which is capable of hybridizing to the poly-A tail of mRNA. Alternatively, the capture domain may comprise nucleotides which are functionally or structurally analogous to poly-T, i.e. are capable of binding selectively to poly-A, for example a poly-U oligonucleotide or an oligonucleotide comprised of deoxythymidine analogues, wherein said oligonucleotide retains the functional property of binding specifically to poly-A. In a particularly preferred embodiment the capture domain, or more particularly the poly-T element of the capture domain, comprises at least 10 nucleotides, preferably at least 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides. In such embodiments, the poly-T element of the capture domain may comprise up to 30 or up to 35 nucleotides. In a further embodiment, the capture domain, or more particularly the poly-T element of the capture domain comprises at least 25, 30 or 35 nucleotides. For instance, the poly-T element of the capture domain may comprise 10-14 nucleotides, 15-25 nucleotides or 25-35 nucleotides.

Random sequences may also be used in the capture of nucleic acid, as is known in the art, e.g. random hexamers or similar sequences, and hence such random sequences may be used to form all or a part of the capture domain. For example, random sequences may be used in conjunction with poly-T (or poly-T analogue etc.) sequences. Thus where a capture domain comprises a poly-T (or a "poly-T-like") oligonucleotide, it may also comprise a random oligonucleotide sequence. This may for example be located 5' or 3' of the poly-T sequence, e.g. at the 3' end of the capture probe, but the positioning of such a random sequence is not critical. Such a construct may facilitate the capturing of the initial part of the poly-A of mRNA. Alternatively, the capture domain may be an entirely random sequence. Degenerate capture domains may also be used, according to principles known in the art.

The capture domain may be capable of binding selectively to a desired sub-type or subset of nucleic acid, e.g. RNA, for example a particular type of RNA such mRNA or rRNA etc. as listed above, or to a particular subset of a given type of RNA, for example, a particular mRNA species e.g. corresponding to a particular gene or group of genes. Such a capture probe may be selected or designed based on sequence of the RNA it is desired to capture. Thus it may be a sequence-specific capture probe, specific for a particular RNA target or group of targets (target group etc). Thus, it may be based on a particular gene sequence or particular motif sequence or common/conserved sequence etc., according to principles well known in the art.

In embodiments where the capture probe is immobilized on the substrate indirectly, e.g. via hybridization to a surface probe, the capture domain may further comprise an upstream sequence (5' to the sequence that hybridizes to the nucleic acid, e.g. RNA of the tissue sample) that is capable of hybridizing to 5' end of the surface probe. Alone, the capture domain of the capture probe may be seen as a capture domain oligonucleotide, which may be used in the synthesis of the capture probe in embodiments where the capture probe is immobilized on the array indirectly.

The positional domain (feature identification domain or tag) of the capture probe, if present, is located directly or indirectly upstream, i.e. closer to the 5' end of the capture probe nucleic acid molecule, of the capture domain. Preferably the positional domain is directly adjacent to the capture domain, i.e. there is no intermediate sequence between the capture domain and the positional domain. In some embodiments the positional domain forms the 5' end of the capture probe, which may be immobilized directly or indirectly on the substrate of the array.

As discussed above, when the capture probes are arrayed on the substrate, each feature (distinct position) of the array comprises a spot of a species of nucleic acid probe, preferably wherein the positional domain at each feature is unique. In some embodiments, the same positional domain may be used for a group of features, preferably a group of features in close proximity to each other. For instance, when multiple species of capture probes that contain different capture domains are used on the same substrate, it may be advantageous to use the same positional domain for each type of capture domain that is immobilized in a group of directly or indirectly adjacent features.

Directly adjacent features are neighbouring features on the array, for instance if the features are arrayed in a standard grid formation, any single feature will have 8 features that are directly adjacent. Indirectly adjacent features may be viewed as features that are in close proximity to each other, i.e. they are immobilized within a specific area on the array surface, but may have one or more features separating them, e.g. 1, 2, 3, 4 or 5 features may separate indirectly adjacent features. In some cases, e.g. on high density arrays, indirectly adjacent features may be separated by up to 10, 20, 30, 40 or 50 features.

A "species" of capture probe is defined with reference to the sequence of the capture domain and, if present, its positional domain; a single species of capture probe will have a unique capture domain sequence and/or unique combination of capture domain sequence and positional domain sequence. However, it is not required that each member of a species of capture probe has the same sequence in its entirety. In particular, since the capture domain may be or may comprise a random or degenerate sequence, the capture domains of individual probes within a species may vary. Accordingly, in some embodiments where the capture domains of the capture probes are the same, each feature comprises a single probe sequence. However in other embodiments where the capture domain varies, members of a species of probe will not have the exact same sequence, although the sequence of the positional domain of each member in the species will be the same. What is required is that each feature or position of the array carries a capture probe of a single species (specifically each feature or position carries a capture probe which has an identical positional tag, i.e. there is a single positional domain at each feature or position, although probes immobilized at directly or indirectly adjacent features may comprise the same positional domain if they comprise different capture domains). Each species has a different capture domain, positional domain and/or combination of capture domain and positional domain which identifies the species. However, each member of a species, may in some cases, as described in more detail herein, have a different capture domain, as the capture domain may be random or degenerate or may have a random or degenerate component. This means that within a given feature, or position, the capture domain of the probes may differ.

Thus in some, but not necessarily in all embodiments, the nucleotide sequence of any one probe molecule immobilized at a particular feature is the same as the other probe molecules immobilized at the same feature, but the nucleotide

sequence of the probes at each feature is different, distinct or distinguishable from the probes immobilized at every other feature. Preferably each feature comprises a different species of probe. However, in some embodiments it may be advantageous for a group of features to comprise the same species of probe, i.e. effectively to produce a feature covering an area of the array that is greater than a single feature, e.g. to lower the resolution of the array. In other embodiments of the array, the nucleotide sequence of the positional domain of any one probe molecule immobilized at a particular feature may be the same as the other probe molecules immobilized at the same feature but the capture domain may vary. The capture domain may nonetheless be designed to capture the same type of molecule, e.g. mRNA in general.

The positional domain (or tag), when present in the capture probe, comprises the sequence which is unique to each feature or a group of directly or indirectly adjacent features, and acts as a positional or spatial marker (the identification tag). In this way each region or domain of the tissue sample, e.g. each cell in the tissue, may be identified by spatial resolution across the array linking the nucleic acid, e.g. RNA (e.g. the transcripts) from a certain cell to a unique positional domain sequence in the capture probe. The positional domain of a capture probe in the array is one aspect of the methods of the invention that allows a specific transcript (or group of transcripts) to be correlated to a position in the tissue sample, for example it may be correlated to a cell in the sample. Thus, the positional domain of the capture domain may be seen as a nucleic acid tag (identification tag) and enables the position of a transcript to be correlated to a position in the sample indirectly, e.g. by analysis of the sequence of the captured transcript.

In some embodiments of the present invention, the methods allow a specific transcript (or group of transcripts) to be correlated to a position in the tissue sample directly, i.e. without the need for sequence analysis. For instance, a species of capture probe may be immobilized on a substrate (either in an array format or uniformly), wherein the capture domain of the capture probe is specific for a transcript (or group of transcripts). Only the specific transcript(s) will interact with (hybridize to) the capture probes. The transcripts will be captured on the substrate in a position relative to where the transcript is expressed in the tissue sample. The steps of generating and labelling the cDNA molecules using the captured transcript as a template and subsequent detection, e.g. imaging, of the labelled cDNA will allow the spatial expression pattern of the transcript to be determined directly. The

information obtained from the spatial detection, e.g. image, of the labelled cDNA may be correlated with an image the tissue sample to determine the precise areas of expression in the tissue sample.

It will be evident that multiple species of capture probe, each specific for a different transcript could be used to determine the expression pattern of multiple genes simultaneously. In a preferred embodiment the species of capture probe are arrayed on the substrate and the capture probes comprise a positional domain. In some embodiments groups of adjacent features (i.e. defining a small area on the array) may comprise capture probes with the same positional domain and different capture domains. The step of detecting, e.g. imaging, the labelled cDNA will provide a preliminary analysis of where the transcripts are expressed in the tissue sample. If more specific expression analysis is required or desirable, the immobilized cDNA molecules may be processed and analysed further, as described below. Sequence analysis of the immobilised cDNA allows the position of a particular transcript to be correlated to a position in the tissue sample, e.g. a cell, by virtue of the positional domain in the capture probe.

Alternatively or additionally, the labelled cDNA may be used to as a marker to select areas of cDNA immobilized on the substrate (which are representative of expression activity in the tissue sample) for further analysis. For instance, immobilized cDNA molecules may be removed from all areas of the substrate except the area of interest, e.g. by laser ablation, and the remaining immobilized cDNA molecules may be processed and analysed as described in more detail below. The removal of immobilized cDNA molecules from the surface of the substrate, e.g. from regions that correspond to cell or tissues in the tissue sample that are not of interest or from regions where the signal from the labelled cDNA molecules indicates that the transcripts are expressed above or below a specific threshold level, reduces the amount of further analysis required, i.e. fewer cDNA molecules for analysis means that less sequence analysis is required, which may result in a reduction in the amount of reagents and/or time required to perform the analysis.

Any suitable sequence may be used as the positional domain in the capture probes of the invention. By a suitable sequence, it is meant that the positional domain should not interfere with (i.e. inhibit or distort) the interaction between the RNA of the tissue sample and the capture domain of the capture probe. For example, the positional domain should be designed such that nucleic acid

molecules in the tissue sample do not hybridize specifically to the positional domain. Preferably, the nucleic acid sequence of the positional domain of the capture probes has less than 80% sequence identity to the nucleic acid sequences in the tissue sample. Preferably, the positional domain of the capture probe has less than 70%, 60%, 50% or less than 40% sequence identity across a substantial part of the nucleic acids molecules in the tissue sample. Sequence identity may be determined by any appropriate method known in the art, e.g. using the BLAST alignment algorithm.

In a preferred embodiment the positional domain of each species of capture probe contains a unique barcode sequence. The barcode sequences may be generated using random sequence generation. The randomly generated sequences may be followed by stringent filtering by mapping to the genomes of all common reference species and with pre-set Tm intervals, GC content and a defined distance of difference to the other barcode sequences to ensure that the barcode sequences will not interfere with the capture of the nucleic acid, e.g. RNA from the tissue sample and will be distinguishable from each other without difficulty.

As mentioned above, in some embodiments, the capture probe may comprise a universal domain (or linker domain or tag). The universal domain of the capture probe is located directly or indirectly upstream, i.e. closer to the 5' end of the capture probe nucleic acid molecule, of the capture domain or, if present, the positional domain. Preferably the universal domain is directly adjacent to the capture domain or, if present, the positional domain, i.e. there is no intermediate sequence between the capture domain and the universal domain or the positional domain and the universal domain. In embodiments where the capture probe comprises a universal domain, the domain will form the 5' end of the capture probe, which may be immobilized directly or indirectly on the substrate of the array.

The universal domain may be utilized in a number of ways in the methods of the invention. For example, in some embodiments the methods of the invention comprise a step of releasing (e.g. removing) at least part of the synthesised (i.e. extended) nucleic acid, e.g. cDNA molecules from the surface of the array. As described elsewhere herein, this may be achieved in a number of ways, of which one comprises cleaving the nucleic acid, e.g. cDNA molecule from the surface of the array. Thus, the universal domain may itself comprise a cleavage domain, i.e. a sequence that can be cleaved specifically, either chemically or preferably enzymatically.

Thus, the cleavage domain may comprise a sequence that is recognised by one or more enzymes capable of cleaving a nucleic acid molecule, i.e. capable of breaking the phosphodiester linkage between two or more nucleotides. For instance, the cleavage domain may comprise a restriction endonuclease (restriction enzyme) recognition sequence. Restriction enzymes cut double-stranded or single stranded DNA at specific recognition nucleotide sequences known as restriction sites and suitable enzymes are well known in the art. For example, it is particularly advantageous to use rare-cutting restriction enzymes, i.e. enzymes with a long recognition site (at least 8 base pairs in length), to reduce the possibility of cleaving elsewhere in the immobilized nucleic acid, e.g. cDNA molecule. In this respect, it will be seen that removing or releasing at least part of the nucleic acid, e.g. cDNA molecule requires releasing a part comprising the capture domain of the nucleic acid, e.g. cDNA, and all of the sequence downstream of the capture domain, e.g. all of the sequence that is 3' to the first nucleotide in the capture domain. Hence, cleavage of the nucleic acid, e.g. cDNA molecule should take place 5' to the capture domain. In preferred embodiments, removing or releasing at least part of the nucleic acid, e.g. cDNA, molecule requires releasing a part comprising the positional domain of the nucleic acid, e.g. cDNA and all of the sequence downstream of the positional domain, e.g. all of the sequence that is 3' to the first nucleotide in the positional domain. Hence, cleavage of the nucleic acid, e.g. cDNA molecule should take place 5' to the positional domain.

By way of example, the cleavage domain may comprise a poly-U sequence which may be cleaved by a mixture of Uracil DNA glycosylase (UDG) and the DNA glycosylase-lyase Endonuclease VIII, commercially known as the USER™ enzyme.

A further example of a cleavage domain can be utilised in embodiments where the capture probe is immobilized to the array substrate indirectly, i.e. via a surface probe. The cleavage domain may comprise one or more mismatch nucleotides, i.e. when the complementary parts of the surface probe and the capture probe are not 100% complementary. Such a mismatch is recognised, e.g. by the MutY and T7 endonuclease I enzymes, which results in cleavage of the nucleic acid molecule at the position of the mismatch.

In some embodiments of the invention, the capture domain of the capture probe comprises a cleavage domain, wherein the said cleavage domain is located at the 5' end of the capture domain. This cleavage domain may be viewed as a universal domain or part of the universal domain. In some embodiments of the

invention, the positional domain of the capture probe comprises a cleavage domain, wherein the said cleavage domain is located at the 5' end of the positional domain. This cleavage domain may be viewed as a universal domain or part of the universal domain.

The universal domain may comprise also an amplification domain. This may be in addition to, or instead of, a cleavage domain. In some embodiments of the invention, as described elsewhere herein, it may be advantageous to amplify the nucleic acid, e.g. cDNA molecules, for example after they have been released (e.g. removed or cleaved) from the substrate. It will be appreciated however, that the initial cycle of amplification, or indeed any or all further cycles of amplification may also take place *in situ* on the substrate. The amplification domain comprises a distinct sequence to which an amplification primer may hybridize. The amplification domain of the universal domain of the capture probe is preferably identical for each species of capture probe. Hence a single amplification reaction will be sufficient to amplify all of the nucleic acid, e.g. cDNA, molecules (which may or may not be released from the substrate prior to amplification).

Any suitable sequence may be used as the amplification domain in the capture probes of the invention. By a suitable sequence, it is meant that the amplification domain should not interfere with (i.e. inhibit or distort) the interaction between the nucleic acid, e.g. RNA of the tissue sample, and the capture domain of the capture probe. Furthermore, the amplification domain should comprise a sequence that is not the same or substantially the same as any sequence in the nucleic acid, e.g. RNA of the tissue sample, such that the primer used in the amplification reaction can hybridize only to the amplification domain under the amplification conditions of the reaction.

For example, the amplification domain should be designed such that nucleic acid molecules in the tissue sample do not hybridize specifically to the amplification domain or the complementary sequence of the amplification domain. Preferably, the nucleic acid sequence of the amplification domain of the capture probes and the complement thereof has less than 80% sequence identity to the nucleic acid sequences in the tissue sample. Preferably, the positional domain of the capture probe has less than 70%, 60%, 50% or less than 40% sequence identity across a substantial part of the nucleic acid molecules in the tissue sample. Sequence identity may be determined by any appropriate method known in the art, e.g. the using BLAST alignment algorithm.

Thus, alone, the universal domain of the capture probe may be seen as a universal domain oligonucleotide, which may be used in the synthesis of the capture probe in embodiments where the capture probe is immobilized on the array indirectly.

In some embodiments, the capture domain of the capture probe may be used as an amplification domain. For instance, in embodiments in which the capture probe does not contain a positional domain or universal domain. In a representative embodiment, the capture probe may be used to capture mRNA from a tissue sample, e.g. using a poly-T oligonucleotide. The signal from the labelled cDNA may be detected, e.g. imaged, and a portion of the array may be selected for further analysis. Thus, the unwanted cDNA molecules may be removed from the substrate, e.g. using laser ablation, and the remaining immobilized cDNA molecules may be released and/or amplified using the capture domain as a primer site, i.e. amplified using a poly-A oligonucleotide primer. The sequence analysis may provide positional information even when the captures probes do not contain positional domains because the sequence information is derived from nucleic acid molecules amplified only from a specific region of the substrate, which correlates to a specific region or portion of the tissue sample.

In one representative embodiment of the invention only the positional domain of each species of capture probe is unique. Hence, the capture domains and universal domains (if present) are in one embodiment the same for every species of capture probe for any particular array to ensure that the capture of the nucleic acid, e.g. RNA, from the tissue sample is uniform across the array. However, as discussed above, in some embodiments the capture domains may differ by virtue of including random or degenerate sequences or gene specific sequences.

In embodiments where the capture probe is immobilized on the substrate indirectly, e.g. via hybridisation to a surface probe, the capture probe may be synthesised on the substrate as described below.

The surface probes are immobilized on the substrate directly by or at, e.g. their 3' end. In embodiments where the probes are arrayed on the substrate each species of surface probe may be unique to each feature (distinct position) or groups of features (directly or indirectly adjacent features) of the array and is partly complementary to the capture probe, defined above.

Hence the surface probe comprises at its 5' end a domain (complementary capture domain) that is complementary to a part of the capture domain that does not bind to the nucleic acid, e.g. RNA, of the tissue sample. In other words, it comprises a domain that can hybridize to at least part of a capture domain oligonucleotide. The surface probe may further comprise a domain (complementary positional domain or complementary feature identification domain) that is complementary to the positional domain of the capture probe, if present. The complementary positional domain is located directly or indirectly downstream (i.e. at the 3' end) of the complementary capture domain, i.e. there may be an intermediary or linker sequence separating the complementary positional domain and the complementary capture domain. In embodiments where the capture probe is synthesized on the array surface, the surface probes of the array always comprise a domain (complementary universal domain) at the 3' end of the surface probe, i.e. directly or indirectly downstream of the positional domain (if present), which is complementary to the universal domain of the capture probe. In other words, it comprises a domain that can hybridize to at least part of the universal domain oligonucleotide.

In some embodiments of the invention the sequence of the surface probe shows 100% complementarity or sequence identity to the positional domain (if present) and the universal domain and to the part of the capture domain that does not bind to the nucleic acid, e.g. RNA, of the tissue sample. In other embodiments the sequence of the surface probe may show less than 100% sequence identity to the domains of the capture probe, e.g. less than 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91% or 90%. In a particularly preferred embodiment of the invention, the complementary universal domain shares less than 100% sequence identity to the universal domain of the capture probe.

In one embodiment of the invention, the capture probe is synthesized or generated on the substrate. In a representative embodiment (see figure 1), the substrate comprises surface probes as defined above. Oligonucleotides that correspond to the capture domain and universal domain of the capture probe are contacted with the substrate and allowed to hybridize to the complementary domains of the surface probes. Excess oligonucleotides may be removed by washing the substrate under standard hybridization conditions. The resultant substrate comprises partially single stranded probes, wherein both the 5' and 3' ends of the surface probe are double stranded and the complementary positional

domain is single stranded. The substrate may be treated with a polymerase enzyme to extend the 3' end of the universal domain oligonucleotide, in a template dependent manner, so as to synthesize the positional domain of the capture probe. The 3' end of the synthesized positional domain is then ligated, e.g. using a ligase enzyme, to the 5' end of the capture domain oligonucleotide to generate the capture probe. It will be understood in this regard that the 5' end of the capture domain oligonucleotide is phosphorylated to enable ligation to take place. As each species of surface probe comprises a unique complementary positional domain, each species of capture probe will comprise a unique positional domain.

It will be evident that in embodiments of the invention where the capture probe does not comprise a positional domain, the polymerase extension step described above may be omitted. Hence, the capture domain and universal domain may be allowed to hybridize to the complementary domains of the surface probes. Excess oligonucleotides may be removed by washing the substrate under standard hybridization conditions. The 3' end of the universal domain is then ligated, e.g. using a ligase enzyme, to the 5' end of the capture domain oligonucleotide to generate the capture probe.

The term "hybridisation" or "hybridises" as used herein refers to the formation of a duplex between nucleotide sequences which are sufficiently complementary to form duplexes via Watson-Crick base pairing. Two nucleotide sequences are "complementary" to one another when those molecules share base pair organization homology. "Complementary" nucleotide sequences will combine with specificity to form a stable duplex under appropriate hybridization conditions. For instance, two sequences are complementary when a section of a first sequence can bind to a section of a second sequence in an anti-parallel sense wherein the 3'-end of each sequence binds to the 5'-end of the other sequence and each A, T(U), G and C of one sequence is then aligned with a T(U), A, C and G, respectively, of the other sequence. RNA sequences can also include complementary G=U or U=G base pairs. Thus, two sequences need not have perfect homology to be "complementary" under the invention. Usually two sequences are sufficiently complementary when at least about 90% (preferably at least about 95%) of the nucleotides share base pair organization over a defined length of the molecule. The domains of the capture and surface probes thus contain a region of complementarity. Furthermore the capture domain of the capture probe contains a

region of complementarity for the nucleic acid, e.g. RNA (preferably mRNA), of the tissue sample.

The capture probe may also be synthesised on the substrate using polymerase extension (similarly to as described above) and a terminal transferase enzyme to add a "tail" which may constitute the capture domain. This is described further in Example 5 below. The use of terminal transferases to add nucleotide sequences to the end of an oligonucleotide is known in the art, e.g. to introduce a homopolymeric tail, e.g. a poly-T tail. Accordingly, in such a synthesis an oligonucleotide that corresponds to the universal domain of the capture probe may be contacted with the substrate and allowed to hybridize to the complementary domain of the surface probes. Excess oligonucleotides may be removed by washing the substrate under standard hybridization conditions. The resultant substrate comprises partially single stranded probes, wherein the 3' ends of the surface probes are double stranded and the complementary positional domain is single stranded. The substrate may be treated with a polymerase enzyme to extend the 3' end of the universal domain oligonucleotide, in a template dependent manner, so as to synthesize the positional domain of the capture probe. The capture domain, e.g. comprising a poly-T sequence, may then be introduced using a terminal transferase to add a poly-T tail to the positional domain to generate the capture probe. As described above, in embodiments where the capture probe does not comprise a positional domain, the polymerase extension step may be omitted and the capture domain may be introduced using a terminal transferase to add a poly-T tail to the universal domain to generate the capture probe.

The object substrate (often simply referred to as the "substrate") of, and for use in the methods of, the invention may contain multiple spots, or "features". Accordingly, in some embodiments of the invention the object substrate may be an array, i.e. an object substrate, e.g. slide or chip, on which the immobilized probes are arrayed on the surface of the substrate. A feature may be defined as an area or distinct position on the array substrate at which a single species of capture probe is immobilized. Hence each feature will comprise a multiplicity of probe molecules, of the same species. It will be understood in this context that whilst it is encompassed that each capture probe of the same species may have the same sequence, this need not necessarily be the case. Each species of capture probe may have the same positional domain (i.e. each member of a species and hence each probe in a feature may be identically "tagged"), but the sequence of each member of the

feature (species) may differ, because the sequence of a capture domain may differ. As described above, random, degenerate or sequence specific capture domains may be used. Thus the capture probes within a feature may comprise different random or degenerate sequences. The number and density of the features on the substrate, e.g. array, will determine the resolution of the array, i.e. the level of detail at which the transcriptome of the tissue sample can be analysed. Hence, a higher density of features will typically increase the resolution of the array.

As discussed above, the size and number of the features on the substrate, e.g. array, of the invention will depend on the nature of the tissue sample and required resolution. Thus, if it is desirable to determine a transcriptome only for regions of cells within a tissue sample (or the sample contains large cells) then the number and/or density of features on the array may be reduced (i.e. lower than the possible maximum number of features) and/or the size of the features may be increased (i.e. the area of each feature may be greater than the smallest possible feature), e.g. an array comprising few large features. Alternatively, if it is desirable to determine a transcriptome of individual cells within a sample, it may be necessary to use the maximum number of features possible, which would necessitate using the smallest possible feature size, e.g. an array comprising many small features.

In some embodiments the capture probes immobilized on the substrate are not in an array format, i.e. the capture probes may be distributed uniformly on the substrate. In these embodiments, it is not necessary for the capture probes to comprise a positional domain, because the capture probes are not immobilized at specific positions on the substrate. However, the capture probes may comprise a universal domain.

By distributed uniformly on the substrate it is meant that the capture probes are immobilized on at least a portion of the substrate and there is an even amount of capture probe immobilized in any specific area in that portion, i.e. the mean amount of probe immobilized per unit area is consistent across the portion of substrate on which the probe is immobilized. The mean amount of probe immobilized per unit area may be controlled by the concentration of probe contacted with the substrate to immobilize the probe or the conditions used to immobilize the probe. For instance, when the capture probe is immobilized to the substrate indirectly, e.g. by hybridisation to a surface probe, the hybridisation and/or wash conditions may be modified such that not all of the surface probes are bound

to a capture probe, i.e. in some embodiments not all of the surface probes are occupied by (i.e. bound to) a capture probe. Substrates on which the amount of probe immobilized per unit area is high will have a higher resolution.

A portion of the substrate may be at least 10% of the total substrate area. In some embodiments, a portion may be at least 20, 30, 40, 50, 60, 70, 80, 90, 95, 99 or 100% of the total substrate area. The portion of the substrate on which capture probes may be immobilized will be dependent on the size of the substrate and the size of the tissue sample to be contacted with the substrate. Advantageously, the area of the portion of the substrate on which probe is immobilized will be larger than the total area of the tissue sample. Preferably the area of the portion will be at least 1, 2, 3, 4 or 5% larger than the area of the tissue sample. In some embodiments the area of the portion will be least 10, 15, 20, 30, 40, 50% larger than the area of the tissue sample.

Whilst single cell resolution may be a preferred and advantageous feature of the present invention, it is not essential to achieve this, and resolution at the cell group level is also of interest, for example to detect or distinguish a particular cell type or tissue region, e.g. normal vs tumour cells.

In representative embodiments of the invention where the substrate is an array, the array may contain at least 2, 5, 10, 50, 100, 500, 750, 1000, 1500, 3000, 5000, 10000, 20000, 40000, 50000, 75000, 100000, 150000, 200000, 300000, 400000, 500000, 750000, 800000, 1000000, 1200000, 1500000, 1750000, 2000000, 2100000. 3000000, 3500000, 4000000, 4200000 or 4300000 features. Whilst 4300000 represents the maximum number of features presently available on a commercial array, it is envisaged that arrays with features in excess of this may be prepared and such arrays are of interest in the present invention. For instance, commercially available arrays allow for more than one array to be provided on a single substrate, e.g. slide or chip. Hence, the number of features on a substrate may be multiples of the above figures. For instance, the substrate, e.g. array, may comprise at least 8600000, 12900000 or 17200000 features. Given that array technology is continually developing, the array could contain even larger numbers of features, e.g. it has been postulated that is may be possible to include as many as 3.3×10^9 features on an array. As noted above, feature size may be decreased and this may allow greater numbers of features to be accommodated within the same or a similar area. By way of example these features may be comprised in an area of less than about $20cm^2$, $10cm^2$, $5cm^2$, $1cm^2$, $1mm^2$, or $100\mu m^2$.

Thus, in some embodiments of the invention the area of each feature may be from about 1 μm^2, 2 μm^2, 3 μm^2, 4 μm^2, 5 μm^2, 10 μm^2, 12 μm^2, 15 μm^2, 20 μm^2, 50 μm^2, 75 μm^2, 100 μm^2, 150 μm^2, 200 μm^2, 250 μm^2, 300 μm^2, 400 μm^2, or 500 μm^2. In some embodiments the area of a feature may be less than 1 μm^2, e.g. less than 0.5, 0.4, 0.2 or 0.1 μm^2. In some embodiments, e.g. when the object substrate comprises a single feature, the feature (e.g. the area on which a capture probe is immobilized on the substrate with a homogeneous distribution, i.e. uniformly) may be from about 100μm^2-20cm^2, 1mm^2-10cm^2, 0.5cm^2-5cm^2, 0.6cm^2-4cm^2, 0.7cm^2-3cm^2, 0.8cm^2-2cm^2 or 0.9cm^2-1cm^2. e.g. at least about 0.5cm^2, 0.6cm^2, 0.7cm^2, 0.8cm^2, 0.9cm^2, 1cm^2, 2cm^2, 3cm^2, 4cm^2 or 5cm^2.

It will be evident that a tissue sample from any organism could be used in the methods of the invention, e.g. plant, animal or fungal. The substrate, e.g. array, of the invention allows the capture of any nucleic acid, e.g. mRNA molecules, which are present in cells that are capable of transcription and/or translation. The substrates, e.g. arrays, and methods of the invention are particularly suitable for isolating and analysing the transcriptome of cells within a sample, wherein spatial resolution of the transcriptomes is desirable, e.g. where the cells are interconnected or in contact directly with adjacent cells. However, it will be apparent to a person of skill in the art that the methods of the invention may also be useful for the analysis of the transcriptome of different cells or cell types within a sample even if said cells do not interact directly, e.g. a blood sample. In other words, the cells do not need to present in the context of a tissue and can be applied to the array as single cells (e.g. cells isolated from a non-fixed tissue). Such single cells, whilst not necessarily fixed to a certain position in a tissue, are nonetheless applied to a certain position on the substrate, e.g. array, and can be individually identified. Thus, in the context of analysing cells that do not interact directly, or are not present in a tissue context, the spatial properties of the described methods may be applied to obtaining or retrieving unique or independent transcriptome information from individual cells. Thus, a tissue sample may be defined as a sample of tissue comprising one or more cells. It may include a suspension of cells.

The sample may thus be a harvested or biopsied tissue sample, or possibly a cultured sample. Representative samples include clinical samples e.g. whole blood or blood-derived products, blood cells, tissues, biopsies, or cultured tissues or cells etc. including cell suspensions. In some embodiments, the sample may be enriched for one or more types of cell, e.g. specific types of blood cell or tumour

cells. Techniques for cell isolation or enrichment are known in the art, and may include positive or negative selection based on expression of particular cell markers. Artificial tissues may for example be prepared from cell suspension (including for example blood cells). Cells may be captured in a matrix (for example a gel matrix e.g. agar, agarose, etc) and may then be sectioned in a conventional way. Such procedures are known in the art in the context of immunohistochemistry (see e.g. Andersson *et al* 2006, J. Histochem. Cytochem. 54(12): 1413-23. Epub 2006 Sep 6).

The methods of the invention may find particular utility in the identification of tumour cells, especially transcriptionally active, e.g. metastatic, tumour cells or tumour cells with metastatic potential. In this respect, tumour cells may be found in blood. Whilst about 90% of the population of circulating tumour cells may show no or little gene activity, because apoptosis or necrosis pathways are activated in this population, the remaining 10% are transcriptionally active. At least some of the transcriptionally active cells are likely to give rise to metastasis and it would be useful to be able to identify these cells. Accordingly, the tissue sample may be a suspension of cells containing tumour cells (e.g. a sample of tumour cells isolated from a blood sample, or a blood sample in which the tumour cells have been enriched, or a biopsy or tissue sample, or tumour cells enriched from such a sample). The method may be performed as described above, wherein the detection of labelled cDNA will correlate to transcriptionally active, e.g. metastatic or potentially metastatic, tumour cells. Thus, in some embodiments the invention may be seen as providing a method for the identification of tumour cells, especially transcriptionally active, e.g. potentially metastatic, tumour cells.

The mode of tissue preparation and how the resulting sample is handled may affect the transcriptomic analysis of the methods of the invention. Moreover, various tissue samples will have different physical characteristics and it is well within the skill of a person in the art to perform the necessary manipulations to yield a tissue sample for use with the methods of the invention. However, it is evident from the disclosures herein that any method of sample preparation may be used to obtain a tissue sample that is suitable for use in the methods of the invention. For instance any layer of cells with a thickness of approximately 1 cell or less may be used in the methods of the invention. In one embodiment, the thickness of the tissue sample may be less than 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 or 0.1 of the cross-section of a cell. However, since as noted above, the present invention is not

limited to single cell resolution and hence it is not a requirement that the tissue sample has a thickness of one cell diameter or less; thicker tissue samples may, if desired, be used. For example cryostat sections may be used, which may be e.g. 10-20 μm thick.

The tissue sample may be prepared in any convenient or desired way and the invention is not restricted to any particular type of tissue preparation. Fresh, frozen, fixed or unfixed tissues may be used. Any desired convenient procedure may be used for fixing or embedding the tissue sample, as described and known in the art. Thus any known fixatives or embedding materials may be used.

As a first representative example of a tissue sample for use in the invention, the tissue may prepared by deep freezing at temperature suitable to maintain or preserve the integrity (i.e. the physical characteristics) of the tissue structure, e.g. less than -20°C and preferably less than -25, -30, -40, -50, -60, -70 or -80 °C. The frozen tissue sample may be sectioned, i.e. thinly sliced, onto the substrate surface by any suitable means. For example, the tissue sample may be prepared using a chilled microtome, a cryostat, set at a temperature suitable to maintain both the structural integrity of the tissue sample and the chemical properties of the nucleic acids in the sample, e.g. to less than -15°C and preferably less than -20 or -25°C. Thus, the sample should be treated so as to minimize the degeneration or degradation of the nucleic acid, e.g. RNA in the tissue. Such conditions are well-established in the art and the extent of any degradation may be monitored through nucleic acid extraction, e.g. total RNA extraction and subsequent quality analysis at various stages of the preparation of the tissue sample.

In a second representative example, the tissue may be prepared using standard methods of formalin-fixation and paraffin-embedding (FFPE), which are well-established in the art. Following fixation of the tissue sample and embedding in a paraffin or resin block, the tissue samples may sectioned, i.e. thinly sliced, onto the substrate, e.g. array. As noted above, other fixatives and/or embedding materials can be used.

It will be apparent that the tissue sample section will need to be treated to remove the embedding material, e.g. to deparaffinize, i.e. to remove the paraffin or resin, from the sample prior to carrying out the methods of the invention. This may be achieved by any suitable method and the removal of paraffin or resin or other material from tissue samples is well established in the art, e.g. by incubating the sample (on the surface of the array) in an appropriate solvent e.g. xylene, e.g. twice

for 10 minutes, followed by an ethanol rinse, e.g. 99.5% ethanol for 2 minutes, 96% ethanol for 2 minutes, and 70% ethanol for 2 minutes.

It will be evident to the skilled person that the RNA in tissue sections prepared using methods of FFPE or other methods of fixing and embedding is more likely to be partially degraded than in the case of frozen tissue. However, without wishing to be bound by any particular theory, it is believed that this may be advantageous in the methods of the invention. For instance, if the RNA in the sample is partially degraded the average length of the RNA polynucleotides will be less and more randomized than a non-degraded sample. It is postulated therefore that partially degraded RNA would result in less bias in the various processing steps, described elsewhere herein, e.g. ligation of adaptors (amplification domains), amplification of the cDNA molecules and sequencing thereof.

Hence, in one embodiment of the invention the tissue sample, i.e. the section of the tissue sample contacted with the substrate, e.g. array, is prepared using FFPE or other methods of fixing and embedding. In other words the sample may be fixed, e.g. fixed and embedded. In an alternative embodiment of the invention the tissue sample is prepared by deep-freezing. In another embodiment a touch imprint of a tissue may be used, according to procedures known in the art. In other embodiments an unfixed sample may be used.

The thickness of the tissue sample section for use in the methods of the invention may be dependent on the method used to prepare the sample and the physical characteristics of the tissue. Thus, any suitable section thickness may be used in the methods of the invention. In representative embodiments of the invention the thickness of the tissue sample section will be at least 0.1μm, further preferably at least 0.2, 0.3, 0.4, 0.5, 0.7, 1.0, 1.5, 2, 3, 4, 5, 6, 7, 8, 9 or 10μm. In other embodiments the thickness of the tissue sample section is at least 10, 12, 13 , 14, 15, 20, 30, 40 or 50μm. However, the thickness is not critical and these are representative values only. Thicker samples may be used if desired or convenient e.g. 70 or 100 μm or more. Typically, the thickness of the tissue sample section is between 1-100 μm, 1-50 μm , 1-30 μm, 1-25 μm, 1-20 μm, 1-15 μm, 1-10 μm, 2-8μm, 3-7μm or 4-6μm, but as mentioned above thicker samples may be used.

On contact of the tissue sample section with the substrate, e.g. following removal of the embedding material e.g. deparaffinization, the nucleic acid, e.g. RNA, molecules in the tissue sample will bind to the immobilized capture probes on the substrate. In some embodiments it may be advantageous to facilitate the

hybridization of the nucleic acid, e.g. RNA molecules to the capture probes. Typically, facilitating the hybridization comprises modifying the conditions under which hybridization occurs. The primary conditions that can be modified are the time and temperature of the incubation of the tissue section on the array prior to the reverse transcription step, which is described elsewhere herein.

It will be evident that tissue samples from different sources may require different treatments to allow the nucleic acids to interact with, i.e. hybridize to, the capture probes immobilized on the substrate. For instance, it may be useful to permeabilize the tissue sample to facilitate the transfer of nucleic acid to the substrate surface. If the tissue sample is not permeabilized sufficiently the amount of nucleic acid captured on the substrate may be too low to enable further analysis (see Figure 2A), i.e. the signal from the labelled cDNA molecules may be of low intensity. Conversely, if the tissue sample is too permeable, the nucleic acid may diffuse away from its origin in the tissue sample. Hence, the nucleic acid may be captured on the substrate, but may not correlate accurately with its original spatial distribution in the tissue sample (see Figure 2C), i.e. the signal from the labelled cDNA molecules may have low spatial resolution. Hence, there must be a balance between permeabilizing the tissue sample enough to obtain a good signal intensity whilst maintaining the spatial resolution of the nucleic acid distribution in the tissue sample (see Figure 2B). The methods used to fix the tissue sample may also impact on the nucleic acid transfer from tissue sample to substrate.

Suitable methods and agents for permeabilizing and/or fixing cells and tissues are well known in the art and any appropriate method may be selected for use in the methods of the invention. In this respect, the methods of the invention may be used to determine the optimum conditions, e.g. the optimum combination of permeabilizing and/or fixative agents, for the capture of nucleic acids from a particular tissue sample. The inventors have found that some proteases are particularly useful in permeabilizing cells in a tissue sample, e.g. pepsin. Particularly useful fixatives include, e.g. methanol.

Thus, the methods and substrate of the invention are particularly useful for determining the optimum conditions for localised or spatial detection of the transcriptome of a tissue sample. In this respect, the step of labelling the cDNA generated on the substrate is particularly important because it enables the efficacy of the nucleic acid molecule capture to be assessed. Labelling the cDNA allows it to be detected, e.g. visualised, directly and the intensity and/or resolution of the

detected signal can be quantified. The signal intensity and/or resolution may be compared with the signal obtained from cDNA generated on a substrate prepared using different methods, e.g. tissue permeabilization and/or fixation methods. The methods that result in the best signal intensity and/or resolution may be selected for use in future analyses and/or may be optimised further. The methods do not require the capture probes to comprise positional domains and/or universal domains (although the capture probes may comprise these domains in some embodiments). Furthermore, the capture probes do not need to be arrayed on the substrate (although the capture probes may be arrayed in some embodiments). Consequently, the methods and substrates of the invention are particularly advantageous because they can be performed cheaply and using commonly available instrumentation or apparatus. Once optimum conditions have been determined for capturing nucleic acid molecules on a substrate from a particularly type of tissue sample and/or tissue sample prepared using a particular method, the optimum conditions can be used to analyse similar tissue samples, wherein further analysis, such as sequence analysis may be performed. The methods of the invention therefore obviate the need to use expensive substrates to optimise the nucleic acid capture conditions, i.e. arrays comprising multiple species of immobilized capture probes, which contain positional and/or universal domains.

As conditions for localised or spatial detection of nucleic acid molecules from a tissue sample on a substrate vary depending on the tissue sample, a typical range of parameters is discussed herein. For instance, on contacting the tissue sample section with the substrate, e.g. an array, the substrate may be incubated for at least 1 hour to allow the nucleic acid, e.g. RNA, to hybridize to the capture probes. This may be particularly useful for FFPE tissue samples before the paraffin is removed. Preferably the substrate may be incubated for at least 2, 3, 5, 10, 12, 15, 20, 22 or 24 hours or until the tissue sample section has dried. In other embodiments, e.g. for FFPE tissue samples after the paraffin has been removed or fresh frozen tissue, the substrate may be incubated for less time, e.g. at least 5 minutes, e.g. at least 10, 15, 20, 25 or 30 minutes. The substrate incubation time is not critical and any convenient or desired time may be used. Typical substrate, e.g. array, incubations may be up to 72 hours. Thus, the incubation may occur at any suitable temperature, for instance at room temperature, although in a preferred embodiment the tissue sample section is incubated on the array at a temperature of at least 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36 or 37°C. Incubation

temperatures of up to 55°C, e.g. 40, 45 or 50°C are commonplace in the art. In a particularly preferred embodiment the tissue sample section is allowed to dry on the substrate, e.g. array, at 37°C for 24 hours. In another preferred embodiment, the tissue sample section is allowed to dry on the substrate, e.g. array, at 50°C for 15 minutes. It will be understood therefore that the precise conditions and methods for contacting the tissue sample with the array are not critical and may vary according to the nature of the sample and the fixation. Once the tissue sample section has dried the substrate may be stored at room temperature before performing the reverse transcription step. It will be understood that the if the tissue sample section is allowed to dry on the substrate surface, it will need to be rehydrated before further manipulation of the captured nucleic acid can be achieved, e.g. the step of reverse transcribing the captured RNA.

Hence, the method of the invention may comprise a further step of rehydrating the tissue sample after contacting the sample with the substrate, e.g. array.

In some embodiments, the tissue sample, e.g. tissue section, may be treated or modified prior to the step of contacting the tissue sample with the substrate and/or prior to generating the cDNA molecules on the substrate, e.g. to select one or more portions of the tissue sample for analysis. For instance, the tissue sample may be dissected to isolate one or more portions for analysis. Alternatively viewed, the tissue sample may be dissected to discard one or more portions for which analysis is not required. Any suitable method for dissecting the tissue sample may be utilised in the methods of the invention. In some embodiments, the tissue sample is dissected using laser capture microdissection (LCM). Accordingly, the method of the invention may comprise a step of dissecting the tissue sample. This aspect of the method comprise a step of retaining one or more portions of tissue sample for analysis and/or discarding one or more portions of tissue sample.

In a preferred aspect of the invention, the tissue sample may be dissected on the substrate. For instance, the tissue sample may be contacted with (e.g. fixed to) a LCM membrane. The LCM membrane:tissue sample composite may be further contacted with the object substrate to form a "sandwich", in which the tissue sample is the central layer in the sandwich. One or more portions of the tissue sample may be dissected using a laser, wherein the one or more portions of the tissue sample that are not required for analysis are removed from the substrate

(e.g. peeled off along with the LCM membrane) and optionally discarded. The LCM membrane may be removed from the remaining one or more portions of tissue sample on the substrate and the method of the invention may be performed as described herein. In some embodiments, the one or more portions of tissue sample removed from the substrate may be used for analysis. Thus, in some embodiments capture probes may be immobilized on the surface of the LCM membrane and/or the substrate. Alternatively viewed, the LCM membrane may be an object substrate as described herein. Thus, the step of dissecting the tissue sample may comprise contacting the tissue sample with more than one substrate, wherein RNA is captured from one or more portions of the tissue sample on each substrate.

In some embodiments it may be advantageous to block (e.g. mask or modify) the capture probes prior to contacting the tissue sample with the substrate, particularly when the nucleic acid in the tissue sample, or the tissue sample itself, is subject to a process of modification prior to its capture on the substrate. Specifically, it may be advantageous to block or modify the free 3' end of the capture probe. It may be necessary to block or modify the capture probes, particularly the free 3' end of the capture probe, prior to contacting the tissue sample with the substrate to avoid modification of the capture probes, e.g. to avoid the removal or modification of the free 3' OH group on the end of the capture probes. Preferably the incorporation of a blocking domain may be incorporated into the capture probe when it is synthesised. However, the blocking domain may be incorporated to the capture probe after its synthesis.

In some embodiments the capture probes may be blocked by any suitable and reversible means that would prevent modification of the capture domains during the process of modifying the nucleic acid and/or the tissue sample, which occurs after the tissue sample has been contacted with the substrate. In other words, the capture probes may be reversibly masked or modified such that the capture domain of the capture probe does not comprise a free 3' end, i.e. such that the 3' end is removed or modified, or made inaccessible so that the capture probe is not susceptible to the process or reaction which is used to modify the nucleic acid and/or the tissue sample. Alternatively, if the capture probes are not blocked prior to the process or reaction used to modify the nucleic acid and/or tissue sample, capture probe may be modified after the process or reaction to reveal and/or restore the 3' end of the capture domain of the capture probe.

For example, blocking probes may be hybridised to the capture probes to mask the free 3' end of the capture domain, e.g. hairpin probes or partially double stranded probes, suitable examples of which are known in the art. The free 3' end of the capture domain may be blocked by chemical modification, e.g. addition of an azidomethyl group as a chemically reversible capping moiety such that the capture probes do not comprise a free 3' end. Suitable alternative capping moieties are well known in the art, e.g. the terminal nucleotide of the capture domain could be a reversible terminator nucleotide, which could be included in the capture probe during or after probe synthesis.

Alternatively or additionally, the capture domain of the capture probe could be modified so as to allow the removal of any modifications of the capture probe, e.g. additional nucleotides, that occur when the nucleic acid molecules and/or the tissue sample are modified. For instance, the capture probes may comprise an additional sequence downstream of the capture domain, i.e. 3' to capture domain, namely a blocking domain. This could be in the form of, e.g. a restriction endonuclease recognition sequence or a sequence of nucleotides cleavable by specific enzyme activities, e.g. uracil. Following the modification of the nucleic acid and/or the tissue sample, the capture probes could be subjected to an enzymatic cleavage, which would allow the removal of the blocking domain and any of the additional nucleotides that are added to the 3' end of the capture probe during the modification process. The removal of the blocking domain would reveal and/or restore the free 3' end of the capture domain of the capture probe. The blocking domain could be synthesised as part of the capture probe or could be added to the capture probe *in situ* (i.e. as a modification of an existing substrate), e.g. by ligation of the blocking domain.

The capture probes may be blocked using any combination of the blocking mechanisms described above.

Once the nucleic acid and/or tissue sample has been modified or processed to enable the nucleic acid to hybridise to the capture domain of the capture probe, the capture probe must be unblocked, e.g. by dissociation of the blocking oligonucleotide, removal of the capping moiety and/or blocking domain.

In order to correlate the transcriptome information, e.g. signal intensity and/or resolution of the labelled cDNA, sequence analysis etc., obtained from the substrate, e.g. by imaging the substrate and/or sequence analysis of the immobilized cDNA at one or more features of the array, with the region (i.e. an area

or cell) of the tissue sample the tissue sample is oriented in relation to the immobilized probes on the substrate, e.g. oriented in relation to the features on the array. In other words, the tissue sample is placed on the substrate, e.g. array, such that the position of a capture probe on the substrate, e.g. array, may be correlated with a position in the tissue sample. Thus it may be identified where in the tissue sample the position of each species of capture probe (or each feature of the array) corresponds. In other words, it may be identified to which location in the tissue sample the position of each species of capture probe corresponds. This may be done by virtue of positional markers present on the array, as described below.

Conveniently, but not necessarily, the tissue sample may be imaged following its contact with the array. This may be performed before or after the nucleic acid of the tissue sample is processed, e.g. before or after the cDNA generation step of the method, in particular the step of generating the first strand cDNA by reverse transcription. In some embodiments, the tissue sample is imaged before the immobilized cDNA is labelled. However, the tissue sample may be imaged at the same time as the labelled cDNA is imaged. In embodiments in which the cDNA is released from the surface of the substrate, the tissue sample may be imaged prior to the release of the captured and synthesised (i.e. extended) cDNA from the substrate, e.g. array. In a particularly preferred embodiment the tissue is imaged after the nucleic acid of the tissue sample has been processed, e.g. after the reverse transcription step, and any residual tissue is removed (e.g. washed) from the array prior to detecting, e.g. imaging, the labelled cDNA and/or the release of molecules from the substrate, e.g. array. In some embodiments, the step of processing the captured nucleic acid, e.g. the reverse transcription step, may act to remove residual tissue from the array surface, e.g. when using tissue preparing by deep-freezing. In such a case, imaging of the tissue sample may take place prior to the processing step, e.g. the cDNA synthesis step. Generally speaking, imaging may take place at any time after contacting the tissue sample with the substrate, but before any step which degrades or removes the tissue sample. As noted above, this may depend on the tissue sample.

Advantageously, the substrate, e.g. array, may comprise markers to facilitate the orientation of the tissue sample or the image thereof in relation to the immobilized capture probes on the substrate, e.g. the features of the array. Any suitable means for marking the array may be used such that they are detectable when the tissue sample is imaged. For instance, a molecule, e.g. a fluorescent

molecule, that generates a signal, preferably a visible signal, may be immobilized directly or indirectly on the surface of the array. Preferably, the array comprises at least two markers in distinct positions on the surface of the substrate, further preferably at least 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100 markers. Conveniently several hundred or even several thousand markers may be used. In some embodiments, tens of thousands of markers may be used. The markers may be provided in a pattern, for example the markers may make up an outer edge of the portion of the substrate on which the capture probes are immobilized, e.g. the markers may be a row of features on the outer border of the portion on which the capture probes are immobilized on the substrate, e.g. an entire outer row of features on an array. Other informative patterns may be used, e.g. lines sectioning the array. This may facilitate aligning an image of the tissue sample to the signal detected from the labelled cDNA molecules, (e.g. the image of the labelled cDNA molecules) i.e. to the portion on which the capture probes are immobilized on the substrate, e.g. an array, or indeed generally in correlating the features of the array to the tissue sample. Thus, the marker may be an immobilized molecule to which a signal giving molecule may interact to generate a signal. In a representative example, the substrate, e.g. array, may comprise a marker feature, e.g. a nucleic acid probe immobilized on the substrate to which a labelled nucleic acid may hybridize. For instance, a labelled nucleic acid molecule, or marker nucleic acid, may be linked or coupled to a chemical moiety capable of fluorescing when subjected to light of a specific wavelength (or range of wavelengths), i.e. excited. Such a marker nucleic acid molecule may be contacted with the array before, contemporaneously with or after the tissue sample is stained in order to visualize or image the tissue sample. However, the marker must be detectable when the tissue sample is imaged. Thus, in a preferred embodiment the marker may be detected using the same imaging conditions used to visualize the tissue sample. Furthermore, it is advantageous that the marker is detectable when the labelled cDNA is detected, e.g. imaged. Hence, in some embodiments the marker may be detected using the same conditions e.g. imaging conditions, used to detect the signal from the labelled cDNA.

In a particularly preferred embodiment of the invention, the substrate, e.g. array, comprises marker features to which a labelled, preferably fluorescently labelled, marker nucleic acid molecule, e.g. oligonucleotide, is hybridized.

The step of detecting, e.g. imaging, the labelled cDNA may use any convenient means known in the art, but typically will comprise microscopy e.g. light, bright field, dark field, phase contrast, fluorescence, reflection, interference, confocal microscopy or a combination thereof. However, the method used will be dependent on the method used to label the cDNA synthesized on the surface of the substrate. Numerous methods for labelling nucleic acid molecules, both single stranded and double stranded molecules are known in the art. The label must result in a visibly detectable signal. Whilst the label does not have to be directly signal giving, this is preferred as it reduces the number of processing steps require to generate a signal. If several steps are required to generate a signal from the labelled cDNA, the resulting signal may be inconsistent, i.e. signals from different areas of the substrate may be non-uniform.

A directly detectable label is one that can be directly detected without the use of additional reagents, while an indirectly detectable label is one that is detectable by employing one or more additional reagents, e.g., where the label is a member of a signal producing system made up of two or more components. In many embodiments, the label is a directly detectable label, where directly detectable labels of interest include, but are not limited to: fluorescent labels, coloured labels, radioisotopic labels, chemiluminescent labels, and the like. Any spectrophotometrically or optically-detectable label may be used. In other embodiments the label may provide a signal indirectly, i.e. it may require the addition of further components to generate signal. For instance, the label may be capable of binding a molecule that is conjugated to a signal giving molecule.

The label is incorporated into the synthesized part of the cDNA molecules, i.e. as part of the synthesized molecules, e.g. a labelled nucleotide, or binds to the newly synthesized part of the nucleic acid molecule. Hence, the capture probe, or a part thereof (such as a positional or universal domain), is not a label for the purpose of this aspect of the invention. The function of the label is to indicate areas on the substrate at which transcript has been captured and cDNA has been synthesized. Accordingly, a capture probe, or part thereof, cannot achieve this function as its presence on the surface of the substrate is not conditional on the presence of transcript captured on the surface of the substrate.

In preferred embodiments, the cDNA is labelled by the incorporation of a labelled nucleotide when the cDNA is synthesized. The labelled nucleotide may be incorporated in the first and/or second strand synthesis. In a particularly preferred

embodiment, the labelled nucleotide is a fluorescently labelled nucleotide. Thus, the labelled cDNA may be imaged by fluorescence microscopy. Whilst fluorescent labels require excitation to provide a detectable signal, as the source of excitation is derived from the instrument/apparatus used to detect the signal, fluorescent labels may be viewed as directly signal giving labels. Fluorescent molecules that may be used to label nucleotides are well known in the art, e.g. fluorescein, the cyanine dyes, such as Cy3, Cy5, Alexa 555, Bodipy 630/650, and the like. Other labels, such as those described below, may also be employed as are known in the art. In preferred embodiments fluorescently tagged CTP (such as Cy3-CTP, Cy5-CTP) is incorporated into the cDNA molecules synthesized on the surface of the array.

As mentioned above, labels may be incorporated into the synthesized cDNA by binding to the molecules, e.g., via intercalation. Representative detectable molecules that may find use in such embodiments include fluorescent nucleic acid stains, such as phenanthridinium dyes, including monomers or homo- or heterodimers thereof, that give an enhanced fluorescence when complexed with nucleic acids. Examples of phenanthridinium dyes include ethidium homodimer, ethidium bromide, propidium iodide, and other alkyl-substituted phenanthridinium dyes. In another embodiment of the invention, the nucleic acid stain is or incorporates an acridine dye, or a homo- or heterodimer thereof, such as acridine orange, acridine homodimer, ethidium-acridine heterodimer, or 9-amino-6-chloro-2-methoxyacridine. In yet another embodiment of the invention, the nucleic acid stain is an indole or imidazole dye, such as Hoechst 33258, Hoechst 33342, Hoechst 34580 (BIOPROBES 34, Molecular Probes, Inc. Eugene, Oreg., (May 2000)) DAPI (4',6-diamidino-2-phenylindole) or DIPI (4',6-(diimidazolin-2-yl)-2-phenylindole). Other permitted nucleic acid stains include, but are not limited to, 7-aminoactinomycin D, hydroxystilbamidine, LDS 751, selected psoralens (furocoumarins), styryl dyes, metal complexes such as ruthenium complexes, and transition metal complexes (incorporating Tb^{3+} and Eu^{3+}, for example). In certain embodiments of the invention, the nucleic acid stain is a cyanine dye or a homo- or heterodimer of a cyanine dye that gives an enhanced fluorescence when associated with nucleic acids. Any of the dyes described in U.S. Pat. No. 4,883,867 to Lee (1989), U.S. Pat. No. 5,582,977 to Yue et al. (1996), U.S. Pat. No. 5,321,130 to Yue et al. (1994), and U.S. Pat. No. 5,410,030 to Yue et al. (1995) (all four patents incorporated by reference) may be used, including nucleic acid stains commercially available under the trademarks TOTO, BOBO, POPO, YOYO, TO-

PRO, BO-PRO, PO-PRO and YO-PRO from Molecular Probes, Inc., Eugene, Oreg. Any of the dyes described in U.S. Pat. No. 5,436,134 to Haugland et al. (1995), U.S. Pat. No. 5,658,751 to Yue et al. (1997), and U.S. Pat. No. 5,863,753 to Haugland et al. (1999) (all three patents incorporated by reference) may be used, including nucleic acid stains commercially available under the trademarks SYBR Green, SYBR Gold, EvaGreen, SYTO, SYTOX, PICOGREEN, OLIGREEN, and RIBOGREEN from Molecular Probes, Inc., Eugene, Oreg. In yet other embodiments of the invention, the nucleic acid stain is a monomeric, homodimeric or heterodimeric cyanine dye that incorporates an aza- or polyazabenzazolium heterocycle, such as an azabenzoxazole, azabenzimidazole, or azabenzothiazole, that gives an enhanced fluorescence when associated with nucleic acids, including nucleic acid stains commercially available under the trademarks SYTO, SYTOX, JOJO, JO-PRO, LOLO, LO-PRO from Molecular Probes, Inc., Eugene, Oreg. The type of nucleic acid stain may be selected based on its capacity to bind to single or double stranded nucleic acid. In embodiments where the first cDNA strand is labelled, it may be preferable to use nucleic acid stains capable of labelling single stranded nucleic acid molecules as the RNA transcript captured on the substrate and used to template cDNA synthesis may be partially or fully degraded.

In a particularly advantageous embodiment of the invention, the methods may include a step of removing a portion of the nucleic acid molecules immobilized on the surface of the substrate. This step may be particularly advantageous for analysing the transcriptome of part or portion of a tissue sample, e.g. an area of interest, particular cell or tissue type etc. Removing nucleic acid molecules immobilized on the surface of the substrate that are not of interest may reduce costs and/or time required for further analysis steps. Fewer reagents are needed to perform sequence analysis on cDNA molecules from a portion of the tissue sample in comparison to the reagents required to perform sequence analysis on cDNA molecules from the whole of the tissue sample. Correspondingly, less sequence analysis is required for cDNA molecules from a portion of the tissue sample. A further benefit derived from removing a portion of the nucleic acid molecules from the surface of the substrate is that it is not necessary to include a positional domain in the capture probe to correlate the sequence analysis with a position in the tissue sample. In this respect, because cDNA molecules that are not of interest (i.e. from areas of the tissue sample that are not of interest) have been removed from the substrate, the sequences analysed will have been derived from the specific area of

the tissue sample that correlates to the portion of cDNA molecules that were not removed from the substrate. Accordingly, it is not necessary to immobilize the capture probes on the substrate in the form of an array. However, in many embodiments of the invention, the capture probes comprise a positional domain and/or are immobilized on the substrate in an array format, i.e. the substrate is an array.

The signal obtained from the step of detecting, e.g. imaging, the labelled cDNA molecules enables areas of the tissue sample to be selected for further analysis. In its simplest form, a single discrete portion of the immobilized cDNA on the substrate (corresponding to a portion of the tissue sample) may be selected for further analysis and all other immobilized nucleic acid molecules on the surface of the substrate may be removed. It will evident that more than one portion may be selected for further analysis, e.g. 2, 3, 4, 5, 6, 7, 8, 9, 10 or more portions. In some embodiments 15, 20, 25, 30, 40, 50 or more portions may be selected. Accordingly, cDNA molecules from one or more portions may be selected for removal, e.g. 2, 3, 4, 5, 6, 7, 8, 9, 10 or more portions. In some embodiments 15, 20, 25, 30, 40, 50 or more portions may be selected. In some embodiments, the portions selected for further analysis (or removal) may be based cell or tissue types, by reference an image of the tissue sample and correlating the position in the tissue sample with the position on the substrate. In still further embodiments, the portions selected for further analysis (or removal) may be based on the amount of cDNA immobilized in a particular area. For instance, portions of the substrate in which the intensity of the signal from the immobilized labelled cDNA molecules is above and/or below specific threshold limits may be targeted for removal (or further analysis). Hence, portions on the substrate that correlate to parts of the tissue sample with high levels of transcription may be removed to enrich the proportion of the transcriptome analysed for parts of the tissue sample with moderate or low levels of transcription. Similarly, the analysis may be focussed on parts of the tissue sample with high levels of transcription.

The step of removing a portion of the nucleic acid molecules immobilized on the surface of the substrate may be achieved using any convenient means. In some embodiments the immobilized cDNA molecules may be removed by laser ablation, e.g. the portion(s) of the substrate from which the cDNA molecules are to be removed may be identified by detecting, e.g. imaging, the labelled cDNA molecules and those areas subjected to treatment with a laser that is sufficient to remove the

cDNA molecules from the surface of the substrate. Advantageously, the laser ablation may also remove the tissue sample from the targeted areas. In some embodiments the tissue sample may be removed from the substrate, as described elsewhere herein, prior to the removal of a portion the immobilized cDNA molecules. Suitable instruments and apparatus for removing cDNA molecules from the surface of an array are known in the art, e.g. a MMI Cell cut instrument (Molecular Machines and Industries AG, Glattburg, Switzerland). Other means for removing cDNA molecules from the surface of the substrate may include cleavage, e.g. enzymatic cleavage. For instance, the portion(s) of the substrate for further analysis may be masked (e.g. using standard array mask apparatus) and the non-masked areas of the substrate subjected to a cleavage agent, e.g. an enzyme, to release the capture probes (and attached cDNA molecules) from the surface of the substrate, as described above. The portion of the substrate selected for further analysis may be unmasked and subsequent analysis performed according to the remaining method steps described herein.

The step of imaging the tissue may use any convenient histological means known in the art, e.g. light, bright field, dark field, phase contrast, fluorescence, reflection, interference, confocal microscopy or a combination thereof. Typically the tissue sample is stained prior to visualization to provide contrast between the different regions, e.g. cells, of the tissue sample. The type of stain used will be dependent on the type of tissue and the region of the cells to be stained. Such staining protocols are known in the art. In some embodiments more than one stain may be used to visualize (image) different aspects of the tissue sample, e.g. different regions of the tissue sample, specific cell structures (e.g. organelles) or different cell types. In other embodiments, the tissue sample may be visualized or imaged without staining the sample, e.g. if the tissue sample contains already pigments that provide sufficient contrast or if particular forms of microscopy are used.

In a preferred embodiment, the tissue sample is visualized or imaged using fluorescence microscopy. Accordingly, in some embodiments, the tissue sample and the labelled cDNA may be visualized or imaged at the same time or sequentially using the same imaging apparatus.

The tissue sample, i.e. any residual tissue that remains in contact with the substrate following the reverse transcription step and detecting, e.g. imaging, the labelled cDNA, and optionally imaging the tissue sample if imaging the tissue

sample is desired and was not carried out before reverse transcription, preferably is removed prior to the step of releasing the cDNA molecules from the substrate. Thus, the methods of the invention may comprise a step of washing the substrate. Removal of the residual tissue sample may be performed using any suitable means and will be dependent on the tissue sample. In the simplest embodiment, the substrate may be washed with water. The water may contain various additives, e.g. surfactants (e.g. detergents), enzymes etc to facilitate to removal of the tissue. In some embodiments, the substrate is washed with a solution comprising a proteinase enzyme (and suitable buffer) e.g. proteinase K. In other embodiments, the solution may comprise also or alternatively cellulase, hemicellulase or chitinase enzymes, e.g. if the tissue sample is from a plant or fungal source. In further embodiments, the temperature of the solution used to wash the substrate may be, e.g. at least 30°C, preferably at least 35, 40, 45, 50 or 55°C. It will be evident that the wash solution should minimize the disruption of the immobilized nucleic acid molecules. For instance, in some embodiments the nucleic acid molecules may be immobilized on the substrate indirectly, e.g. via hybridization of the capture probe and the RNA and/or the capture probe and the surface probe, thus the wash step should not interfere with the interaction between the molecules immobilized on the substrate, i.e. should not cause the nucleic acid molecules to be denatured.

Following the step of contacting the substrate with a tissue sample, under conditions sufficient to allow hybridization to occur between the nucleic acid, e.g. RNA (preferably mRNA), of the tissue sample to the capture probes, the step of securing (acquiring) the hybridized nucleic acid takes place. Securing or acquiring the captured nucleic acid involves extending the capture probe to produce a copy of the captured nucleic acid, e.g. generating cDNA from the captured (hybridized) RNA. It will be understood that this refers to the synthesis of a complementary strand of the hybridized nucleic acid, e.g. generating cDNA based on the captured RNA template (the RNA hybridized to the capture domain of the capture probe). Thus, in an initial step of extending the capture probe, e.g. the cDNA generation, the captured (hybridized) nucleic acid, e.g. RNA acts as a template for the extension, e.g. reverse transcription, step. In some embodiments, securing or acquiring the capture nucleic acid may be viewed as tagging or marking the captured nucleic acid with the positional domain specific to the feature on which the nucleic acid is captured. In many embodiments, the step of securing or acquiring

the captured nucleic acid involves the directly incorporating a signal giving label into the synthesized copy of the captured nucleic acid molecule.

Reverse transcription concerns the step of synthesizing cDNA (complementary or copy DNA) from RNA, preferably mRNA (messenger RNA), by reverse transcriptase. Thus cDNA can be considered to be a copy of the RNA present in a cell at the time at which the tissue sample was taken, i.e. it represents all or some of the genes that were expressed in said cell at the time of isolation.

The capture probe, specifically the capture domain of the capture probe, acts as a primer for producing the complementary strand of the nucleic acid hybridized to the capture probe, e.g. a primer for reverse transcription. Hence, the nucleic acid, e.g. cDNA, molecules generated by the extension reaction, e.g. reverse transcription reaction, incorporate the sequence of the capture probe, i.e. the extension reaction, e.g. reverse transcription reaction. Advantageously the molecules generated by the extension reaction incorporate directly a label such that the amount of transcript in the tissue sample that is in contact with the substrate may be determined, e.g. by measuring the intensity of the signal generated by the label. As mentioned above, in some embodiments each species of capture probe may comprise a positional domain (feature identification tag) that represents a unique sequence for each feature of the array. Thus, in some embodiments all of the nucleic acid, e.g. cDNA, molecules synthesized at a specific feature will comprise the same nucleic acid "tag".

The nucleic acid, e.g. cDNA, molecules synthesized at a specific position or area on the surface of the substrate, e.g. each feature of an array, may represent the genes expressed from the region or area of the tissue sample in contact with that position or area, e.g. feature. For instance, a tissue or cell type or group or sub-group thereof, and may further represent genes expressed under specific conditions, e.g. at a particular time, in a specific environment, at a stage of development or in response to stimulus etc. Hence, the cDNA at any single position or area, e.g. feature, may represent the genes expressed in a single cell, or if the position or area, e.g. feature, is in contact with the sample at a cell junction, the cDNA may represent the genes expressed in more than one cell. Similarly, if a single cell is in contact with a large area of the substrate, e.g. multiple features, then each position within the area, e.g. each feature, may represent a proportion of the genes expressed in said cell.

The step of extending the capture probe, e.g. reverse transcription, may be performed using any suitable enzymes and protocol of which many exist in the art, as described in detail below. However, it will be evident that it is not necessary to provide a primer for the synthesis of the first nucleic acid, e.g. cDNA, strand because the capture domain of the capture probe acts as the primer, e.g. reverse transcription primer.

Preferably, in the context of the present invention the secured nucleic acid (i.e. the nucleic acid covalently attached to the capture probe), e.g. cDNA, is treated to comprise double stranded DNA. Treatment of the captured nucleic acid to produce double stranded DNA may be achieved in a single reaction to generate only a second DNA, e.g. cDNA, strand, i.e. to produce double stranded DNA molecules without increasing the number of double stranded DNA molecules, or in an amplification reaction to generate multiple copies of the second strand, which may be in the form of single stranded DNA (e.g. linear amplification) or double stranded DNA, e.g. cDNA (e.g. exponential amplification).

The step of second strand DNA, e.g. cDNA, synthesis may take place *in situ* on the substrate, either as a discrete step of second strand synthesis, for example using random primers as described in more detail below, or in the initial step of an amplification reaction. Alternatively, the first strand DNA, e.g. cDNA (the strand comprising, i.e. incorporating, the capture probe) may be released from the array and second strand synthesis, whether as a discrete step or in an amplification reaction may occur subsequently, e.g. in a reaction carried out in solution.

Where second strand synthesis takes place on the substrate (i.e. *in situ*) the method may include an optional step of removing the captured nucleic acid, e.g. RNA, before the second strand synthesis, for example using an RNA digesting enzyme (RNase) e.g. RNase H. Procedures for this are well known and described in the art. However, this is generally not necessary, and in most cases the RNA degrades naturally. Removal of the tissue sample from the array will generally remove the RNA from the array. RNase H can be used if desired to increase the robustness of RNA removal. RNA removal may be useful in embodiments where the cDNA is labelled after it has been generated, e.g. labelled with a nucleic acid stain. Removal of the RNA may provide a consistent target to which the nucleic acid stain can interact (bind), i.e. all of the immobilized molecules will be single stranded after RNA removal. Prior to a step of RNA removal, the immobilized molecules may be a mixture of fully or partially double stranded molecules (RNA:DNA hybrids) and

single stranded molecules (where the RNA has already degraded). Some nucleic acid stains may provide a stronger signal when interacting with double stranded nucleic acid, when compared to the signal from single stranded nucleic acid. Thus, it is preferable when using a nucleic acid stain to label the immobilised cDNA that the molecules are either all fully single stranded or double stranded.

In tissue samples that comprise large amounts of RNA, the step of generating the double stranded cDNA may yield a sufficient amount of cDNA that it may be sequenced directly (following release from the substrate). In this case, second strand cDNA synthesis may be achieved by any means known in the art and as described below. The second strand synthesis reaction may be performed on the substrate directly, i.e. whilst the cDNA is immobilized on the substrate, or preferably after the cDNA has been released from the substrate, as described below.

In other embodiments it will be necessary to enhance, i.e. amplify, the amount of secured nucleic acid, e.g. synthesized cDNA, to yield quantities that are sufficient for DNA sequencing. In this embodiment, the first strand of the secured nucleic acid, e.g. cDNA molecules, which comprise also the capture probe of the substrate, acts as a template for the amplification reaction, e.g. a polymerase chain reaction. The first reaction product of the amplification will be a second strand of DNA, e.g. cDNA, which itself will act as a template for further cycles of the amplification reaction.

In either of the above described embodiments, the second strand of DNA, e.g. cDNA, will comprise a complement of the capture probe. If the capture probe comprises a universal domain, and particularly an amplification domain within the universal domain, then this may be used for the subsequent amplification of the DNA, e.g. cDNA, e.g. the amplification reaction may comprise a primer with the same sequence as the amplification domain, i.e. a primer that is complementary (i.e. hybridizes) to the complement of the amplification domain. In view of the fact that the amplification domain is upstream of the positional domain (if present) of the capture probe (in the secured nucleic acid, e.g. the first cDNA strand), the complement of the positional domain (if the position domain is present in the capture probe) will be incorporated in the second strand of the DNA, e.g. cDNA molecules.

In embodiments where the second strand of DNA, e.g. cDNA, is generated in a single reaction, the second strand synthesis may be achieved by any suitable

means. For instance, the first strand cDNA, preferably, but not necessarily, released from the substrate, may be incubated with random primers, e.g. hexamer primers, and a DNA polymerase, preferably a strand displacement polymerase, e.g. klenow (exo$^-$), under conditions sufficient for templated DNA synthesis to occur. This process will yield double stranded cDNA molecules of varying lengths and is unlikely to yield full-length cDNA molecules, i.e. cDNA molecules that correspond to entire mRNA from which they were synthesized. The random primers will hybridise to the first strand cDNA molecules at a random position, i.e. within the sequence rather than at the end of the sequence.

If it is desirable to generate full-length DNA, e.g. cDNA, molecules, i.e. molecules that correspond to the whole of the captured nucleic acid, e.g. RNA molecule (if the nucleic acid, e.g. RNA, was partially degraded in the tissue sample then the captured nucleic acid, e.g. RNA, molecules will not be "full-length" transcripts), then the 3' end of the secured nucleic acid, e.g. first stand cDNA, molecules may be modified. For example, a linker or adaptor may be ligated to the 3' end of the cDNA molecules. This may be achieved using single stranded ligation enzymes such as T4 RNA ligase or Circligase™ (Epicentre Biotechnologies).

Alternatively, a helper probe (a partially double stranded DNA molecule capable of hybridising to the 3' end of the first strand cDNA molecule), may be ligated to the 3' end of the secured nucleic acid, e.g. first strand cDNA, molecule using a double stranded ligation enzyme such as T4 DNA ligase. Other enzymes appropriate for the ligation step are known in the art and include, e.g. Tth DNA ligase, Taq DNA ligase, *Thermococcus* sp. (strain 9°N) DNA ligase (9°N™ DNA ligase, New England Biolabs), and Ampligase™ (Epicentre Biotechnologies). The helper probe comprises also a specific sequence from which the second strand DNA, e.g. cDNA, synthesis may be primed using a primer that is complementary to the part of the helper probe that is ligated to the secured nucleic acid, e.g. first cDNA strand. A further alternative comprises the use of a terminal transferase active enzyme to incorporate a polynucleotide tail, e.g. a poly-A tail, at the 3' end of the secured nucleic acid, e.g. first strand of cDNA, molecules. The second strand synthesis may be primed using a poly-T primer, which may also comprise a specific amplification domain for further amplification. Other methods for generating "full-length" double stranded DNA, e.g. cDNA, molecules (or maximal length second strand synthesis) are well-established in the art.

In some embodiments, second strand synthesis may use a method of template switching, e.g. using the SMART™ technology from Clontech®. SMART (Switching Mechanism at 5' End of RNA Template) technology is well established in the art and is based that the discovery that reverse transcriptase enzymes, e.g. Superscript® II (Invitrogen), are capable of adding a few nucleotides at the 3' end of an extended cDNA molecule, i.e. to produce a DNA/RNA hybrid with a single stranded DNA overhang at the 3' end. The DNA overhang may provide a target sequence to which an oligonucleotide probe can hybridise to provide an additional template for further extension of the cDNA molecule. Advantageously, the oligonucleotide probe that hybridises to the cDNA overhang contains an amplification domain sequence, the complement of which is incorporated into the synthesised first strand cDNA product. Primers containing the amplification domain sequence, which will hybridise to the complementary amplification domain sequence incorporated into the cDNA first strand, can be added to the reaction mix to prime second strand synthesis using a suitable polymerase enzyme and the cDNA first strand as a template. This method avoids the need to ligate adaptors to the 3' end of the cDNA first strand. Whilst template switching was originally developed for full-length mRNAs, which have a 5' cap structure, it has since been demonstrated to work equally well with truncated mRNAs without the cap structure. Thus, template switching may be used in the methods of the invention to generate full length and/or partial or truncated cDNA molecules. Thus, in a preferred embodiment of the invention, the second strand synthesis may utilise, or be achieved by, template switching. In a particularly preferred embodiment, the template switching reaction, i.e. the further extension of the cDNA first strand to incorporate the complementary amplification domain, is performed *in situ* (whilst the capture probe is still attached, directly or indirectly, to the substrate, e.g. array). Preferably, the second strand synthesis reaction is also performed *in situ*.

As mentioned above, in some embodiments the immobilized cDNA may be labelled by incorporating label into the second strand of the cDNA, e.g. incorporating labelled nucleotides into the cDNA second strand. This may be in addition to, or as an alternative to, incorporating labelled nucleotides into the first cDNA strand.

In embodiments where it may be necessary or advantageous to enhance, enrich or amplify the DNA, e.g. cDNA, molecules, amplification domains may be incorporated in the DNA, e.g. cDNA, molecules. As discussed above, a first

amplification domain may be incorporated into the secured nucleic acid molecules, e.g. the first strand of the cDNA molecules, when the capture probe comprises a universal domain comprising an amplification domain. In these embodiments, the second strand synthesis may incorporate a second amplification domain. For example, the primers used to generate the second strand cDNA, e.g. random hexamer primers, poly-T primer, the primer that is complementary to the helper probe, may comprise at their 5' end an amplification domain, i.e. a nucleotide sequence to which an amplification primer may hybridize. Thus, the resultant double stranded DNA may comprise an amplification domain at or towards each 5' end of the double stranded DNA, e.g. cDNA, molecules. These amplification domains may be used as targets for primers used in an amplification reaction, e.g. PCR. Alternatively, the linker or adaptor which is ligated to the 3' end of the secured nucleic acid molecules, e.g. first strand cDNA molecules, may comprise a second universal domain comprising a second amplification domain. Similarly, a second amplification domain may be incorporated into the first strand cDNA molecules by template switching.

In embodiments where the capture probe does not comprise a universal domain, particularly comprising an amplification domain, the second strand of the cDNA molecules may be synthesised in accordance with the above description. The resultant double stranded DNA molecules may be modified to incorporate an amplification domain at the 5' end of the first DNA, e.g. cDNA, strand (a first amplification domain) and, if not incorporated in the second strand DNA, e.g. cDNA synthesis step, at the 5' end of the second DNA, e.g. cDNA, strand (a second amplification domain). Such amplification domains may be incorporated, e.g. by ligating double stranded adaptors to the ends of the DNA, e.g. cDNA, molecules. Enzymes appropriate for the ligation step are known in the art and include, e.g. Tth DNA ligase, Taq DNA ligase, *Thermococcus* sp. (strain 9°N) DNA ligase (9°N™ DNA ligase, New England Biolabs), Ampligase™ (Epicentre Biotechnologies) and T4 DNA ligase. In a preferred embodiment the first and second amplification domains comprise different sequences.

From the above, it is therefore apparent that universal domains, which may comprise an amplification domain, may be added to the secured (i.e. extended) DNA molecules, for example to the cDNA molecules, or their complements (e.g. second strand) by various methods and techniques and combinations of such techniques known in the art, e.g. by use of primers which include such a domain,

ligation of adaptors, use of terminal transferase enzymes and/or by template switching methods. As is clear from the discussion herein, such domains may be added before or after release of the DNA molecules from the array.

It will be apparent from the above description that all of the DNA, e.g. cDNA, molecules from a single substrate that have been synthesized by the methods of the invention may all comprise the same first and second amplification domains. Consequently, a single amplification reaction, e.g. PCR, may be sufficient to amplify all of the DNA, e.g. cDNA, molecules. Thus in a preferred embodiment, the method of the invention may comprise a step of amplifying the DNA, e.g. cDNA, molecules. In one embodiment the amplification step is performed after the release of the DNA, e.g. cDNA molecules from the substrate. In other embodiments amplification may be performed on the substrate (i.e. *in situ* on the substrate). It is known in the art that amplification reactions may be carried out on substrates, such as arrays, and on-chip thermocyclers exist for carrying out such reactions. Thus, in one embodiment arrays which are known in the art as sequencing platforms or for use in any form of sequence analysis (e.g. in or by next generation sequencing technologies) may be used as the basis of the substrates of the present invention (e.g. Illumina bead arrays etc.)

For the synthesis of the second strand of DNA, e.g. cDNA, it is preferable to use a strand displacement polymerase (e.g. Φ29 DNA polymerase, Bst (exo^-) DNA polymerase, klenow (exo^-) DNA polymerase) if the cDNA released from the substrate of the array comprises a partially double stranded nucleic acid molecule. For instance, the released nucleic acids will be at least partially double stranded (e.g. DNA:RNA hybrid) in embodiments where the capture probe is immobilized indirectly on the substrate of the array via a surface probe and the step of releasing the DNA, e.g. cDNA molecules comprises a cleavage step. The strand displacement polymerase is necessary to ensure that the second cDNA strand synthesis incorporates the complement of the capture probe including the positional domain (feature identification domain), if present, into the second DNA, e.g. cDNA strand.

It will be evident that the step of releasing at least part of the DNA, e.g. cDNA molecules or their amplicons from the surface of the substrate may be achieved using a number of methods. In some embodiments, it will be evident that the primary aim of the release step is to yield molecules into which the positional domain of the capture probe (or its complement) is incorporated (or included), such

that the DNA, e.g. cDNA, molecules or their amplicons are "tagged" according to their feature (or position) on the array. However, as discussed above, a positional domain is not essential, particularly where only a portion of the nucleic acid molecules are released because the other nucleic acid molecules have been removed from the surface of the substrate (and discarded) in an earlier step. The release step thus removes DNA, e.g. cDNA, molecules or amplicons thereof from the substrate, which DNA, e.g. cDNA, molecules or amplicons include the positional information that can be correlated to the tissue sample. For instance, in some embodiments the released DNA comprises a positional domain or its complement (by virtue of it having been incorporated into the secured nucleic acid, e.g. the first strand cDNA by, e.g. extension of the capture probe, and optionally copied in the second strand DNA if second strand synthesis takes place on the array, or copied into amplicons if amplification takes place on the array). Hence, in order to yield sequence analysis data that can be correlated specifically with the various regions in the tissue sample it is advantageous that the released molecules comprise the positional domain of the capture probe (or its complement). However, if the DNA molecules have been released from a specific portion of the substrate, it will be evident that the sequence analysis can be correlated with the region(s) in the tissue sample that were not removed from the substrate.

Since the released molecule may be a first and/or second strand DNA, e.g. cDNA, molecule or amplicon, and since the capture probe may be immobilised indirectly on the substrate, it will be understood that whilst the release step may comprise a step of cleaving a DNA, e.g. cDNA molecule from the array, the release step does not require a step of nucleic acid cleavage; a DNA, e.g. cDNA molecule or an amplicon may simply be released by denaturing a double-stranded molecule, for example releasing the second cDNA strand from the first cDNA strand, or releasing an amplicon from its template or releasing the first strand cDNA molecule (i.e. the extended capture probe) from a surface probe. Accordingly, a DNA, e.g. cDNA, molecule may be released from the substrate by nucleic acid cleavage and/or by denaturation (e.g. by heating to denature a double-stranded molecule). Where amplification is carried out *in situ* on the substrate, this will of course encompass releasing amplicons by denaturation in the cycling reaction.

In some embodiments, the DNA, e.g. cDNA, molecules are released by enzymatic cleavage of a cleavage domain, which may be located in the universal domain or positional domain of the capture probe. As mentioned above, in some

embodiments the cleavage domain must be located upstream (at the 5' end) of the positional domain such that the released DNA, e.g. cDNA, molecules comprise the positional (identification) domain. Suitable enzymes for nucleic acid cleavage include restriction endonucleases, e.g. RsaI. Other enzymes, e.g. a mixture of Uracil DNA glycosylase (UDG) and the DNA glycosylase-lyase Endonuclease VIII (USER™ enzyme) or a combination of the MutY and T7 endonuclease I enzymes, are preferred embodiments of the methods of the invention.

In an alternative embodiment, the DNA, e.g. cDNA, molecules may be released from the surface of the substrate by physical means. For instance, in embodiments where the capture probe is indirectly immobilized on the substrate, e.g. via hybridization to the surface probe, it may be sufficient to disrupt the interaction between the nucleic acid molecules. Methods for disrupting the interaction between nucleic acid molecules, e.g. denaturing double stranded nucleic acid molecules, are well known in the art. A straightforward method for releasing the DNA, e.g. cDNA, molecules (i.e. of stripping the substrate, e.g. array, of the synthesized DNA, e.g. cDNA molecules) is to use a solution that interferes with the hydrogen bonds of the double stranded molecules. In a preferred embodiment of the invention, the DNA, e.g. cDNA, molecules may be released by applying heated water, e.g. water or buffer of at least 85°C, preferably at least 90, 91, 92, 93, 94, 95, 96, 97, 98, 99°C. As an alternative or addition to the use of a temperature sufficient to disrupt the hydrogen bonding, the solution may comprise salts, surfactants etc. that may further destabilize the interaction between the nucleic acid molecules, resulting in the release of the DNA, e.g. cDNA, molecules.

It will be understood that the application of a high temperature solution, e.g. 90-99°C water may be sufficient to disrupt a covalent bond used to immobilize the capture probe or surface probe to the substrate. Hence, in a preferred embodiment, the DNA, e.g. cDNA, molecules may be released by applying hot water to the substrate to disrupt covalently immobilized capture or surface probes.

It is implicit that the released DNA, e.g. cDNA, molecules (the solution comprising the released DNA, e.g. cDNA, molecules) are collected for further manipulation, e.g. second strand synthesis and/or amplification. Nevertheless, the method of the invention may be seen to comprise a step of collecting or recovering the released DNA, e.g. cDNA, molecules. As noted above, in the context of *in situ* amplification the released molecules may include amplicons of the secured nucleic acid, e.g. cDNA.

In embodiments of methods of the invention, it may be desirable to remove any unextended capture probes. This may be, for example, after the step of releasing DNA molecules from the substrate. Any desired or convenient method may be used for such removal including, for example, use of an enzyme to degrade the unextended probes, e.g. exonuclease.

The DNA, e.g. cDNA molecules, or amplicons, that have been released from the substrate, which may have been modified as discussed above, are analysed to investigate (e.g. determine their sequence, although as noted above actual sequence determination is not required - any method of analysing the sequence may be used). Thus, any method of nucleic acid analysis may be used. The step of sequence analysis may identify the positional domain and hence allow the analysed molecule to be localised to a position in the tissue sample. Similarly, the nature or identity of the analysed molecule may be determined. In this way the nucleic acid, e.g. RNA, at given position in the substrate, and hence in the tissue sample may be determined. Hence the analysis step may include or use any method which identifies the analysed molecule (and hence the "target" molecule) and its positional domain. Generally such a method will be a sequence-specific method. For example, the method may use sequence-specific primers or probes, particularly primers or probes specific for the positional domain and/or for a specific nucleic acid molecule to be detected or analysed e.g. a DNA molecule corresponding to a nucleic acid, e.g. RNA or cDNA molecule to be detected. Typically in such a method sequence-specific amplification primers e.g. PCR primers may be used.

In some embodiments it may be desirable to analyse a subset or family of target related molecules, e.g. all of the sequences that encode a particular group of proteins which share sequence similarity and/or conserved domains, e.g. a family of receptors. Hence, the amplification and/or analysis methods described herein may use degenerate or gene family specific primers or probes that hybridise to a subset of the captured nucleic acids or nucleic acids derived therefrom, e.g. amplicons. In a particularly preferred embodiment, the amplification and/or analysis methods may utilise a universal primer (i.e. a primer common to all of the captured sequences) in combination with a degenerate or gene family specific primer specific for a subset of target molecules.

Thus in one embodiment, amplification-based, especially PCR-based, methods of sequence analysis are used.

However, the steps of modifying and/or amplifying the released DNA, e.g. cDNA, molecules may introduce additional components into the sample, e.g. enzymes, primers, nucleotides etc. Hence, the methods of the invention may further comprise a step of purifying the sample comprising the released DNA, e.g. cDNA molecules or amplicons prior to the sequence analysis, e.g. to remove oligonucleotide primers, nucleotides, salts etc that may interfere with the sequencing reactions. Any suitable method of purifying the DNA, e.g. cDNA molecules may be used.

As noted above, sequence analysis of the released DNA molecules may be direct or indirect. Thus the sequence analysis material or substrate (which may be viewed as the molecules which are subjected to the sequence analysis step or process) may directly be the molecules which is released from the object substrate, e.g. array, or it may be a molecule which is derived therefrom. Thus, for example in the context of sequence analysis step which involves a sequencing reaction, the sequencing template may be the molecule which is released from the object substrate, e.g. array, or it may be a molecule derived therefrom. For example, a first and/or second strand DNA, e.g. cDNA, molecule released from the substrate, e.g. array, may be directly subjected to sequence analysis (e.g. sequencing), i.e. may directly take part in the sequence analysis reaction or process (e.g. the sequencing reaction or sequencing process, or be the molecule which is sequenced or otherwise identified). In the context of *in situ* amplification the released molecule may be an amplicon. Alternatively, the released molecule may be subjected to a step of second strand synthesis or amplification before sequence analysis (e.g. sequencing or identification by other means). The sequence analysis substrate (e.g. template) may thus be an amplicon or a second strand of a molecule which is directly released from the object substrate, e.g. array.

Both strands of a double stranded molecule may be subjected to sequence analysis (e.g. sequenced) but the invention is not limited to this and single stranded molecules (e.g. cDNA) may be analysed (e.g. sequenced). For example various sequencing technologies may be used for single molecule sequencing, e.g. the Helicos or Pacbio technologies, or nanopore sequencing technologies which are being developed. Thus, in one embodiment the first strand of DNA, e.g. cDNA may be subjected to sequencing. The first strand DNA, e.g. cDNA may need to be modified at the 3' end to enable single molecule sequencing. This may be done by

procedures analogous to those for handling the second DNA, e.g. cDNA strand. Such procedures are known in the art.

In a preferred aspect of the invention the sequence analysis will identify or reveal a portion of captured nucleic acid, e.g. RNA, sequence and optionally the sequence of the positional domain. In some embodiments the sequence of the positional domain (or tag) will identify the feature to which the nucleic acid, e.g. mRNA, molecule was captured. The sequence of the captured nucleic acid, e.g. RNA, molecule may be compared with a sequence database of the organism from which the sample originated to determine the gene to which it corresponds. By determining which region (e.g. cell) of the tissue sample was in contact with the position or area, e.g. feature, of the substrate from which the captured nucleic acid was released, it is possible to determine which region of the tissue sample was expressing said gene. This analysis may be achieved for all of the DNA, e.g. cDNA, molecules generated by the methods of the invention, yielding a spatial transcriptome of the tissue sample. However, in some embodiments only a selection of the transcripts present in the tissue sample may be captured on the substrate (e.g. if the capture domain of the capture probe comprises a sequence specific for a gene or set of genes) and/or only a portion of the captured molecules may be selected for further analysis, e.g. sequence analysis (e.g. a portion of the captured molecules may be removed from the substrate prior to the sequence analysis step).

By way of a representative example, sequencing data may be analysed to sort the sequences into specific species of capture probe, e.g. according to the sequence of the positional domain. This may be achieved by, e.g. using the FastX toolkit FASTQ Barcode splitter tool to sort the sequences into individual files for the respective capture probe positional domain (tag) sequences. The sequences of each species, i.e. from each feature, may be analyzed to determine the identity of the transcripts. For instance, the sequences may be identified using e.g. Blastn software, to compare the sequences to one or more genome databases, preferably the database for the organism from which the tissue sample was obtained. The identity of the database sequence with the greatest similarity to the sequence generated by the methods of the invention will be assigned to said sequence. In general, only hits with a certainty of at least $1e^{-6}$, preferably $1e^{-7}$, $1e^{-8}$, or $1e^{-9}$ will be considered to have been successfully identified.

It will be apparent that any nucleic acid sequencing method may be utilised in the methods of the invention. However, the so-called "next generation sequencing" techniques will find particular utility in the present invention. High-throughput sequencing is particularly useful in the methods of the invention because it enables a large number of nucleic acids to be partially sequenced in a very short period of time. In view of the recent explosion in the number of fully or partially sequenced genomes, it is not essential to sequence the full length of the generated DNA, e.g. cDNA molecules to determine the gene to which each molecule corresponds. For example, the first 100 nucleotides from each end of the DNA, e.g. cDNA molecules should be sufficient to identify the gene expressed and, in embodiments in which the capture probes are arrayed on the substrate, the feature to which the nucleic acid, e.g. mRNA, was captured (i.e. its location on the array). In some embodiments, the sequence reaction from the "capture probe end" of the DNA, e.g. cDNA molecules, yields the sequence of the positional domain and at least about 20 bases, preferably 30 or 40 bases of transcript specific sequence data. However, in embodiments in which the capture probe does not contain a positional domain, the sequence reaction from the "capture probe end" of the DNA, may yield at least about 70 bases, preferably 80, 90, or 100 bases of transcript specific sequence data. The sequence reaction from the "non-capture probe end" may yield at least about 70 bases, preferably 80, 90, or 100 bases of transcript specific sequence data.

As a representative example, the sequencing reaction may be based on reversible dye-terminators, such as used in the Illumina™ technology. For example, DNA molecules are first attached to primers on, e.g. a glass or silicon slide and amplified so that local clonal colonies are formed (bridge amplification). Four types of ddNTPs are added, and non-incorporated nucleotides are washed away. Unlike pyrosequencing, the DNA can only be extended one nucleotide at a time. A camera takes images of the fluorescently labelled nucleotides then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing a next cycle. This may be repeated until the required sequence data is obtained. Using this technology, thousands of nucleic acids may be sequenced simultaneously on a single slide.

Other high-throughput sequencing techniques may be equally suitable for the methods of the invention, e.g. pyrosequencing. In this method the DNA is amplified inside water droplets in an oil solution (emulsion PCR), with each droplet

containing a single DNA template attached to a single primer-coated bead that then forms a clonal colony. The sequencing machine contains many picolitre-volume wells each containing a single bead and sequencing enzymes. Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA and the combined data are used to generate sequence read-outs.

An example of a technology in development is based on the detection of hydrogen ions that are released during the polymerisation of DNA. A microwell containing a template DNA strand to be sequenced is flooded with a single type of nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide it is incorporated into the growing complementary strand. This causes the release of a hydrogen ion that triggers a hypersensitive ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence multiple nucleotides will be incorporated in a single cycle. This leads to a corresponding number of released hydrogen ions and a proportionally higher electronic signal.

Thus, it is clear that future sequencing formats are slowly being made available, and with shorter run times as one of the main features of those platforms it will be evident that other sequencing technologies will be useful in the methods of the invention, e.g. nanopore sequencing methods.

An essential feature of the present invention, as described above, is a step of securing a complementary strand of the captured nucleic acid molecules to the capture probe, e.g. reverse transcribing the captured RNA molecules. The reverse transcription reaction is well known in the art and in representative reverse transcription reactions, the reaction mixture includes a reverse transcriptase, dNTPs and a suitable buffer. The reaction mixture may comprise other components, e.g. RNase inhibitor(s). The primers and template are the capture domain of the capture probe and the captured RNA molecules, as described above. In the subject methods, each dNTP will typically be present in an amount ranging from about 10 to 5000 μM, usually from about 20 to 1000 μM. It will be evident that an equivalent reaction may be performed to generate a complementary strand of a captured DNA molecule, using an enzyme with DNA polymerase activity. Reactions of this type are well known in the art and are described in more detail below.

In some embodiments, a labelled dNTP may be present in the reaction mix, thereby incorporating a label into the synthesized DNA molecule. In a

representative embodiment, the labelled dNTP is a fluorescently labelled dNTP, e.g. Cy3-dCTP.

The desired reverse transcriptase activity may be provided by one or more distinct enzymes, wherein suitable examples are: M-MLV, MuLV, AMV, HIV, ArrayScript™, MultiScribe™, ThermoScript™, and SuperScript® I, II, and III enzymes.

The reverse transcriptase reaction may be carried out at any suitable temperature, which will be dependent on the properties of the enzyme. Typically, reverse transcriptase reactions are performed between 37-55°C, although temperatures outside of this range may also be appropriate. The reaction time may be as little as 1, 2, 3, 4 or 5 minutes or as much as 48 hours. Typically the reaction will be carried out for between 3-12 hours, such as 5-120 minutes, 5-60, 5-45 or 5-30 minutes or 1-10 or 1-5 minutes according to choice. The reaction time is not critical and any desired reaction time may be used. For instance, overnight incubations are commonplace.

As indicated above, certain embodiments of the methods include an amplification step, where the copy number of generated DNA, e.g. cDNA molecules is increased, e.g. in order to enrich the sample to obtain a better representation of the nucleic acids, e.g. transcripts, captured from the tissue sample. The amplification may be linear or exponential, as desired, where representative amplification protocols of interest include, but are not limited to: polymerase chain reaction (PCR); isothermal amplification, etc.

The polymerase chain reaction (PCR) is well known in the art, being described in U.S. Pat. Nos.: 4,683,202; 4,683,195; 4,800,159; 4,965,188 and 5,512,462, the disclosures of which are herein incorporated by reference. In representative PCR amplification reactions, the reaction mixture that includes the above released DNA, e.g. cDNA molecules from the substrate, e.g. array, which are combined with one or more primers that are employed in the primer extension reaction, e.g., the PCR primers that hybridize to the first and/or second amplification domains (such as forward and reverse primers employed in geometric (or exponential) amplification or a single primer employed in a linear amplification). The oligonucleotide primers with which the released DNA, e.g. cDNA molecules (hereinafter referred to as template DNA for convenience) is contacted will be of sufficient length to provide for hybridization to complementary template DNA under annealing conditions (described in greater detail below). The length of the primers

will depend on the length of the amplification domains, but will generally be at least 10 bp in length, usually at least 15 bp in length and more usually at least 16 bp in length and may be as long as 30 bp in length or longer, where the length of the primers will generally range from 18 to 50 bp in length, usually from about 20 to 35 bp in length. The template DNA may be contacted with a single primer or a set of two primers (forward and reverse primers), depending on whether primer extension, linear or exponential amplification of the template DNA is desired.

In addition to the above components, the reaction mixture produced in the subject methods typically includes a polymerase and deoxyribonucleoside triphosphates (dNTPs). The desired polymerase activity may be provided by one or more distinct polymerase enzymes. In many embodiments, the reaction mixture includes at least a Family A polymerase, where representative Family A polymerases of interest include, but are not limited to: Thermus aquaticus polymerases, including the naturally occurring polymerase (Taq) and derivatives and homologues thereof, such as Klentaq (as described in Barnes et al, Proc. Natl. Acad. Sci USA (1994) 91:2216-2220); Thermus thermophilus polymerases, including the naturally occurring polymerase (Tth) and derivatives and homologues thereof, and the like. In certain embodiments where the amplification reaction that is carried out is a high fidelity reaction, the reaction mixture may further include a polymerase enzyme having 3'-5' exonuclease activity, e.g., as may be provided by a Family B polymerase, where Family B polymerases of interest include, but are not limited to: Thermococcus litoralis DNA polymerase (Vent) as described in Perler et al., Proc. Natl. Acad. Sci. USA (1992) 89:5577-5581; Pyrococcus species GB-D (Deep Vent); Pyrococcus furiosus DNA polymerase (Pfu) as described in Lundberg et al., Gene (1991) 108:1-6, Pyrococcus woesei (Pwo) and the like. Where the reaction mixture includes both a Family A and Family B polymerase, the Family A polymerase may be present in the reaction mixture in an amount greater than the Family B polymerase, where the difference in activity will usually be at least 10-fold, and more usually at least about 100-fold. Usually the reaction mixture will include four different types of dNTPs corresponding to the four naturally occurring bases present, i.e. dATP, dTTP, dCTP and dGTP. In the subject methods, each dNTP will typically be present in an amount ranging from about 10 to 5000 μM, usually from about 20 to 1000 μM.

The reaction mixtures prepared in the reverse transcriptase and/or amplification steps of the subject methods may further include an aqueous buffer

medium that includes a source of monovalent ions, a source of divalent cations and a buffering agent. Any convenient source of monovalent ions, such as KCl, K-acetate, NH_4-acetate, K-glutamate, NH_4Cl, ammonium sulphate, and the like may be employed. The divalent cation may be magnesium, manganese, zinc and the like, where the cation will typically be magnesium. Any convenient source of magnesium cation may be employed, including $MgCl_2$, Mg-acetate, and the like. The amount of Mg^{2+} present in the buffer may range from 0.5 to 10 mM, but will preferably range from about 3 to 6 mM, and will ideally be at about 5 mM. Representative buffering agents or salts that may be present in the buffer include Tris, Tricine, HEPES, MOPS and the like, where the amount of buffering agent will typically range from about 5 to 150 mM, usually from about 10 to 100 mM, and more usually from about 20 to 50 mM, where in certain preferred embodiments the buffering agent will be present in an amount sufficient to provide a pH ranging from about 6.0 to 9.5, where most preferred is pH 7.3 at 72 °C. Other agents which may be present in the buffer medium include chelating agents, such as EDTA, EGTA and the like.

In preparing the reverse transcriptase, DNA extension or amplification reaction mixture of the steps of the subject methods, the various constituent components may be combined in any convenient order. For example, in the amplification reaction the buffer may be combined with primer, polymerase and then template DNA, or all of the various constituent components may be combined at the same time to produce the reaction mixture.

As discussed above, a preferred embodiment of the invention the DNA, e.g. cDNA molecules may be modified by the addition of amplification domains to the ends of the nucleic acid molecules, which may involve a ligation reaction. A ligation reaction is also required for the *in situ* synthesis of the capture probe on the substrate, when the capture probe is immobilized indirectly on the substrate surface.

As is known in the art, ligases catalyze the formation of a phosphodiester bond between juxtaposed 3'-hydroxyl and 5'-phosphate termini of two immediately adjacent nucleic acids. Any convenient ligase may be employed, where representative ligases of interest include, but are not limited to: Temperature sensitive and thermostable ligases. Temperature sensitive ligases include, but are not limited to, bacteriophage T4 DNA ligase, bacteriophage T7 ligase, and E. coli ligase. Thermostable ligases include, but are not limited to, Taq ligase, Tth ligase,

and Pfu ligase. Thermostable ligase may be obtained from thermophilic or hyperthermophilic organisms, including but not limited to, prokaryotic, eukaryotic, or archael organisms. Certain RNA ligases may also be employed in the methods of the invention.

In this ligation step, a suitable ligase and any reagents that are necessary and/or desirable are combined with the reaction mixture and maintained under conditions sufficient for ligation of the relevant oligonucleotides to occur. Ligation reaction conditions are well known to those of skill in the art. During ligation, the reaction mixture in certain embodiments may be maintained at a temperature ranging from about 4°C to about 50°C, such as from about 20°C to about 37°C for a period of time ranging from about 5 seconds to about 16 hours, such as from about 1 minute to about 1 hour. In yet other embodiments, the reaction mixture may be maintained at a temperature ranging from about 35°C to about 45°C, such as from about 37°C to about 42°C, e.g., at or about 38°C, 39°C, 40°C or 41°C, for a period of time ranging from about 5 seconds to about 16 hours, such as from about 1 minute to about 1 hour, including from about 2 minutes to about 8 hours. In a representative embodiment, the ligation reaction mixture includes 50 mM Tris pH7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 mg/ml BSA, 0.25 units/ml RNase inhibitor, and T4 DNA ligase at 0.125 units/ml. In yet another representative embodiment, 2.125 mM magnesium ion, 0.2 units/ml RNase inhibitor; and 0.125 units/ml DNA ligase are employed. The amount of adaptor in the reaction will be dependent on the concentration of the DNA, e.g. cDNA in the sample and will generally be present at between 10-100 times the molar amount of DNA, e.g. cDNA.

By way of a representative example the method of the invention may comprise the following steps:

(a) contacting an object substrate with a tissue sample, wherein at least one species of capture probe is directly or indirectly immobilized on said substrate such that the probes are oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, such that RNA of the tissue sample hybridises to said capture probes;

(b) imaging the tissue sample on the substrate;

(c) reverse transcribing the captured mRNA molecules to generate cDNA molecules, wherein labelled nucleotides are incorporated into the synthesized part of the cDNA molecules;

(d) washing the substrate to remove residual tissue;

(e) imaging the substrate such that the signal from the labelled cDNA molecules is detected;

(f) removing the labelled cDNA from at least one portion of the surface of the substrate;

(g) releasing at least part of the remaining labelled cDNA molecules from the surface of the array;

(h) performing second strand cDNA synthesis on the released cDNA molecules;

and

(i) analysing the sequence of (e.g. sequencing) the cDNA molecules.

By way of an alternative representative example the method of the invention may comprise the following steps:

(a) contacting an object substrate with a tissue sample, wherein at least one species of capture probe is directly or indirectly immobilized on said substrate such that the probes are oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, such that RNA of the tissue sample hybridises to said capture probes;

(b) optionally rehydrating the tissue sample;

(c) reverse transcribing the captured mRNA molecules to generate cDNA molecules;

(d) imaging the tissue sample on the substrate;

(e) washing the substrate to remove residual tissue;

(f) labelling the cDNA molecules with a nucleic acid stain;

(g) imaging the substrate such that the signal from the labelled cDNA molecules is detected;

(h) removing the labelled cDNA from at least one portion of the surface of the substrate;

(i) releasing at least part of the remaining labelled cDNA molecules from the surface of the array;

(j) amplifying the released cDNA molecules;

(k) optionally purifying the cDNA molecules to remove components that may interfere with the sequencing reaction;

and

(l) analysing the sequence of (e.g. sequencing) the cDNA molecules.

By way of yet a further representative example the method of the invention may comprise the following steps:

(a) contacting an object substrate with a tissue sample, wherein at least one species of capture probe is directly or indirectly immobilized on said substrate such that the probes are oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, such that RNA of the tissue sample hybridises to said capture probes;

(b) imaging the tissue sample on the substrate;

(c) reverse transcribing the captured mRNA molecules to generate cDNA molecules, wherein labelled nucleotides are incorporated into the synthesized part of the cDNA molecules;

(d) washing the substrate to remove residual tissue;

(e) imaging the substrate such that the signal from the labelled cDNA molecules is detected;

(f) repeating steps (a)-(e), with a second object substrate, using different conditions in step (a);

(g) comparing the intensity and/or resolution of the signal from the labelled cDNA molecules immobilized on said first and second object substrate; and

(h) selecting the conditions that provide the optimum signal intensity and/or resolution of the labelled cDNA molecules.

By way of a further alternative representative example the method of the invention may comprise the following steps:

(a) contacting an object substrate with a tissue sample, wherein multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on said substrate and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the substrate, and

(ii) a capture domain;

such that RNA of the tissue sample hybridises to said capture probes;

(b) optionally rehydrating the tissue sample;

(c) imaging the tissue sample on the substrate;

(d) reverse transcribing the captured mRNA molecules to generate cDNA molecules, wherein labelled nucleotides are incorporated into the synthesized part of the cDNA molecules;

(e) washing the substrate to remove residual tissue;

(f) imaging the substrate such that the signal from the labelled cDNA molecules is detected;

(g) removing the labelled cDNA from at least one portion of the surface of the substrate;

(h) releasing at least part of the remaining labelled cDNA molecules from the surface of the array;

(i) amplifying the released cDNA molecules;

(j) optionally purifying the cDNA molecules to remove components that may interfere with the sequencing reaction;

and

(k) analysing the sequence of (e.g. sequencing) the cDNA molecules.

The present invention includes any suitable combination of the steps in the above described methods. It will be understood that the invention also encompasses variations of these methods, for example where amplification is performed *in situ* on the substrate, e.g. array. Also encompassed are methods which omit the step of imaging the tissue sample.

The invention may also be seen to include a method for making or producing an object substrate (i) for use in capturing mRNA from a tissue sample that is contacted with said substrate; (ii) for use in determining and/or analysing a (e.g. the partial or global) transcriptome of a tissue sample; or (iii) for use in determining the optimum conditions for localised or spatial detection of nucleic acid from a tissue sample contacted with a substrate, said method comprising immobilizing, directly or indirectly, at least one species of capture probe to a substrate such that each probe is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer.

Optionally the probes are immobilized uniformly on the object substrate, i.e. the probes are not arrayed as distinct features. In a particular embodiment of the invention, the probes are identical.

In some embodiments of the invention the probes are capable of hybridizing to (i.e. capturing) all mRNA, i.e. RNA molecules with a polyA tail. Hence, in particularly preferred embodiments of the invention the probes may comprise

regions of consecutive dTTP or dUTP nucleotides, e.g. oligoT and/or oligoU nucleotides, as described in more detail above.

The method of immobilizing the capture probes on the object substrate may be achieved using any suitable means as described herein. Where the capture probes are immobilized on the array indirectly the capture probe may be synthesized on the object substrate. Said method may comprise any one or more of the following steps:

(a) immobilizing directly or indirectly multiple surface probes to a substrate, wherein the surface probes comprise:

(i) a domain capable of hybridizing to part of the capture domain oligonucleotide (a part not involved in capturing the nucleic acid, e.g. RNA); and

(ii) a complementary universal domain;

(b) hybridizing to the surface probes immobilized on the substrate capture domain oligonucleotides and universal domain oligonucleotides;

(c) ligating the universal domain to the capture domain oligonucleotide to produce the capture oligonucleotide.

Thus, in one particular embodiment the method may be viewed as a method for making or producing an object substrate comprising a substrate on which one or more species of capture probe, comprising a capture domain, is directly or indirectly immobilized such that each probe is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein the probes are immobilised on the object substrate with a homogeneous distribution and said substrate is for use in:

(i) capturing RNA from a tissue sample that is contacted with said object substrate; or

(ii) localised or spatial detection of RNA in a tissue sample,

said method comprising:

(a) immobilizing directly or indirectly multiple surface probes to a substrate, wherein the surface probes comprise:

(i) a domain capable of hybridizing to part of a capture domain oligonucleotide; and

(ii) a domain that is complementary to a universal domain oligonucleotide;

(b) hybridizing to the surface probes immobilized on the substrate, capture domain oligonucleotides and universal domain oligonucleotides;

(c) ligating the universal domain oligonucleotides to the capture domain oligonucleotides to produce the capture probes.

The features of the object substrate produced by the above method of producing the array of the invention, may be further defined in accordance with the above description.

Thus, in one embodiment the invention provides an object substrate for use in the localised or spatial detection of RNA in a tissue sample, comprising a planar substrate on which one or more species of capture probe, comprising a capture domain, is directly or indirectly immobilized such that each probe is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer wherein the probes are immobilised on the object substrate with a homogeneous distribution and wherein the capture probe is selected from an oligonucleotide comprising a poly-T, poly-U and/or random oligonucleotide sequence.

The invention will be further described with reference to the following non-limiting Examples with reference to the following drawings in which:

Figure 1 shows 3' to 5' surface probe composition and synthesis of 5' to 3' oriented capture probes that are indirectly immobilized at the array surface.

Figure 2 shows how the methods of the invention can be used to determine the optimum conditions for capturing RNA from a tissue sample on a substrate, wherein: (A) shows a tissue sample that is not permeabilized sufficiently to allow the capture of RNA on the substrate; (B) shows a tissue sample that is permeabilized to allow capture of RNA on the substrate whilst retaining spatial information; and (C) shows a tissue sample that is too permeable such that the RNA has been allowed to diffuse away from its origin in the tissue sample and captured RNA does not correlate accurately with its original spatial distribution in the tissue sample.

Figure 3 shows a bar chart demonstrating the efficiency of enzymatic cleavage (USER or RsaI) from in-house manufactured arrays and by 99°C water from Agilent manufactured arrays, as measured by hybridization of fluorescently labelled probes to the array surface after probe release.

Figure 4 shows a table that lists the reads sorted for their origin across the low density in-house manufactured DNA-capture array as seen in the schematic representation.

Figure 5 shows a schematic illustration of the principle of the method described in Example 4, i.e. use of microarrays with immobilized DNA oligos

(capture probes) carrying spatial labelling tag sequences (positional domains). Each feature of oligos of the microarray carries a 1) a unique labelling tag (positional domain) and 2) a capture sequence (capture domain).

Figure 6 shows the composition of 5' to 3' oriented capture probes used on high-density capture arrays.

Figure 7 shows a Matlab visualization of captured transcripts from total RNA extracted from mouse olfactory bulb.

Figure 8 shows Olfr (olfactory receptor) transcripts as visualized across the capture array using Matlab visualization after capture from mouse olfactory bulb tissue.

Figure 9 shows a pattern of printing for in-house 41-ID-tag microarrays.

Figure 10 shows a Matlab visualization of captured ID-tagged transcripts from mouse olfactory bulb tissue on 41-ID-tag in-house arrays overlaid with the tissue image. For clarity, the specific features on which particular genes were identified have been circled.

Figure 11 shows a pattern of cDNA synthesis on array surface from a section of mouse brain olfactory bulb, as visualized by an Agilent microarray scanner.

Figure 12 shows a pattern of cDNA synthesis on array surface from a section of mouse brain cortex, as visualized by an Agilent microarray scanner.

Figure 13 shows a pattern of cDNA synthesis on array surface from a section of mouse brain olfactory bulb, before ablation (a) and after ablation of non-wanted areas (b), as imaged on the MMI cellcut instrument.

Figure 14 shows the resulting library made from cleaved fluorescently labelled cDNA library of non-ablated areas as visualized with a DNA high sensitivity chip on an Agilent Bioanalyzer.

Figure 15 shows a pattern of cDNA synthesis on a high-density feature array surface from a section of mouse brain olfactory bulb. A frame (visible at the top and to the right in the image) consisting of features containing a single DNA probe sequence was labelled by hybridization of a complementary oligonucleotide labelled with Cy3.

Figure 16 shows a bar chart of Cy3 intensities on arrays using different amounts of Cy3 labelled dCTP.

Figure 17 shows images of cDNA synthesis on array surfaces and corresponding images of the tissue sections on the array surface using various

alternative sample types, wherein: (A) depicts the phase contrast image (left) and corresponding Cy3 labelled cDNA footprint (right) of a zebra fish sample; (B) depicts the phase contrast image (left) and corresponding Cy3 labelled cDNA footprint (right) of a fruit fly (Drosophila) sample; (C) depicts the phase contrast image (left) and corresponding Cy3 labelled cDNA footprint (right) of a prostate tumour section; and (D) depicts the phase contrast image (left) and corresponding Cy3 labelled cDNA footprint (right) of mouse fibroblast cells.

Examples

Example 1

Preparation of the array

The following experiments demonstrate how oligonucleotide probes may be attached to an array substrate by either the 5' or 3' end to yield an array with capture probes capable of hybridizing to mRNA.

Preparation of in-house printed microarray with 5' to 3' oriented probes

20 RNA-capture oligonucleotides with individual tag sequences (Tag 1-20, Table 1 were spotted on glass slides to function as capture probes. The probes were synthesized with a 5'-terminus amino linker with a C6 spacer. All probes where synthesized by Sigma-Aldrich (St. Louis, MO, USA). The RNA-capture probes were suspended at a concentration of 20 µM in 150 mM sodium phosphate, pH 8.5 and were spotted using a Nanoplotter NP2.1/E (Gesim, Grosserkmannsdorf, Germany) onto CodeLink™ Activated microarray slides (7.5cm x 2.5cm; Surmodics, Eden Prairie, MN, USA). After printing, surface blocking was performed according to the manufacturer's instructions. The probes were printed in 16 identical arrays on the slide, and each array contained a pre-defined printing pattern. The 16 sub-arrays were separated during hybridization by a 16-pad mask (ChipClip™ Schleicher & Schuell BioScience, Keene, NH, USA).

Table 1

Name	Sequence	5' mod	3' mod	Length
Sequences for free 3' capture probes				
TAP-ID1	UUAAGTACAAATCTCGACTGCCACTCTGAACCTTCTCCTTCTCCTTCACCTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 1)	Amino-C6		72
Enzymatic recog	UUAAGTACAA (SEQ ID NO: 2)			10
Universal amp handle P	ATCTCGACTGCCACTCTGAA (SEQ ID NO: 3)			20
ID1	CCTTCTCCTTCTCCTTCACC (SEQ ID NO: 4)			20
Capture sequence	TTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 5)			22
ID1	CCTTCTCCTTCTCCTTCACC (SEQ ID NO: 6)			20
ID2	CCTTGCTGCTTCTCCTCCTC (SEQ ID NO: 7)			20
ID3	ACCTCCTCCGCCTCCTCCTC (SEQ ID NO: 8)			20
ID4	GAGACATACCACCAAGAGAC (SEQ ID NO: 9)			20
ID5	GTCCTCTATTCCGTCACCAT (SEQ ID NO: 10)			20
ID6	GACTGAGCTCGAACATATGG (SEQ ID NO: 11)			20
ID7	TGGAGGATTGACACAGAACG (SEQ ID NO: 12)			20
ID8	CCAGCCTCTCCATTACATCG (SEQ ID NO: 13)			20
ID9	AAGATCTACCAGCCAGCCAG (SEQ ID NO: 14)			20
ID10	CGAACTTCCACTGTCTCCTC (SEQ ID NO: 15)			20

ID11	TTGCGCCTTCTCCAATACAC (SEQ ID NO: 16)			20
ID12	CTCTTCTTAGCATGCCACCT (SEQ ID NO: 17)			20
ID13	ACCACTTCTGCATTACCTCC (SEQ ID NO: 18)			20
ID14	ACAGCCTCCTCTTCTTCCTT (SEQ ID NO: 19)			20
ID15	AATCCTCTCCTTGCCAGTTC (SEQ ID NO: 20)			20
ID16	GATGCCTCCACCTGTAGAAC (SEQ ID NO: 21)			20
ID17	GAAGGAATGGAGGATATCGC (SEQ ID NO: 22)			20
ID18	GATCCAAGGACCATCGACTG (SEQ ID NO: 23)			20
ID19	CCACTGGAACCTGACAACCG (SEQ ID NO: 24)			20
ID20	CTGCTTCTTCCTGGAACTCA (SEQ ID NO: 25)			20
Sequences for free 5' surface probes and on-chip free 3' capture probe synthesis				
Free 5' surface probe - A	GCGTTCAGAGTGGCAGTCGAGATCACGCGGCAATCATATCGGACAGATCGGAAGAGCGTAGTGTAG (SEQ ID NO: 26)		Amino C7	66
Free 5' surface probe - U	GCGTTCAGAGTGGCAGTCGAGATCACGCGGCAATCATATCGGACGGCTGCTGGTAAATAGAGATCA (SEQ ID NO: 27)		Amino C7	66
Nick	GCG			3
LP'	TTCAGAGTGGCAGTCGAGATCAC (SEQ ID NO: 28)			23
ID'	GCGGCAATCATATCGGAC (SEQ ID NO: 29)			18
A' 22bp MutY mismatch	AGATCGGAAGAGCGTAGTGTAG (SEQ ID NO: 30)			22
U' 22bp MutY mismatch	GGCTGCTGGTAAATAGAGATCA (SEQ ID NO: 31)			
Hybridized sequences for capture probe synthesis				
Illumina amp handle A	ACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 32)			33
Universa ampl handle U	AAGTGTGGAAAGTTGATCGCTATTTACCAGCAGCC (SEQ ID NO: 33)			35

Capture_LP_Poly-dTVN	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 34)	Phosphorylated	45
Capture_LP_Poly-d24T	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 35)	Phosphorylated	47
Additional secondary universal amplification handles			
Illumina amp handle B	AGACGTGTGCTCTTCCGATCT (SEQ ID NO: 36)		21
Universal amp handle X	ACGTCTGTGAATAGCCGCAT (SEQ ID NO: 37)		20
B_R6 handle (or X)	AGACGTGTGCTCTTCCGATCTNNNNNNNNN (SEQ ID NO: 38)		27 (26)
B_R8 handle (or X)	AGACGTGTGCTCTTCCGATCTNNNNNNNNNNN (SEQ ID NO: 39)		29 (28)
B_polyTVN (or X)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 40)		43 (42)
B_poly24T (or X)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 41)		45 (44)
Amplification handle to incorporate A handle into P handle products			
A_P handle	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATCTCGACTGCCACTCTGAA (SEQ ID NO: 42)		53

Preparation of in-house printed microarray with 3' to 5' oriented probes and synthesis of 5' to 3' oriented capture probes

Printing of surface probe oligonucleotides was performed as in the case with 5' to 3' oriented probes above, with an amino-C7 linker at the 3' end, as shown in Table 1.

To hybridize primers for capture probe synthesis, hybridization solution containing 4xSSC and 0.1% SDS, 2µM extension primer (the universal domain oligonucleotide) and 2µM thread joining primer (the capture domain oligonucleotide) was incubated for 4 min at 50°C. Meanwhile the in-house array was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50µL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake; 2) 0.2xSSC for 1 min at 300 rpm shake; and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry and placed back in the ChipClip.

For extension and ligation reaction (to generate the positional domain of the capture probe) 50µL of enzyme mix containing 10xAmpligase buffer, 2.5 U AmpliTaq DNA Polymerase Stoffel Fragment (Applied Biosystems), 10 U Ampligase (Epicentre Biotechnologies), dNTPs 2 mM each (Fermentas) and water, was pipetted to each well. The array was subsequently incubated at 55°C for 30 min. After incubation the array was washed according to the previously described array washing method but the first step has the duration of 10 min instead of 6 min.

The method is depicted in Figure 1.

Tissue Preparation

The following experiments demonstrate how tissue sample sections may be prepared for use in the methods of the invention.

Preparation of fresh frozen tissue and sectioning onto capture probe arrays

Fresh non-fixed mouse brain tissue was trimmed if necessary and frozen down in -40°C cold isopentane and subsequently mounted for sectioning with a cryostat at 10µm. A slice of tissue was applied onto each capture probe array to be used.

Preparation of formalin-fixed paraffin-embedded (FFPE) tissue

Mouse brain tissue was fixed in 4% formalin at 4°C for 24h. After that it was incubated as follows: 3x incubation in 70% ethanol for 1 hour; 1x incubation in 80%

ethanol for 1 hour; 1x incubation in 96% ethanol for 1 hour; 3x incubation in 100% ethanol for 1 hour; and 2x incubation in xylene at room temperature for 1 h.

The dehydrated samples were then incubated in liquid low melting paraffin 52-54°C for up to 3 hours, during which the paraffin was changed once to wash out residual xylene. Finished tissue blocks were then stored at RT. Sections were then cut at 4µm in paraffin with a microtome onto each capture probe array to be used.

The sections were dried at 37°C on the array slides for 24 hours and stored at RT.

Deparaffinization of FFPE tissue

Formalin fixed paraffinized mouse brain 10µm sections attached to CodeLink slides were deparaffinised in xylene twice for: 10 min, 99.5% ethanol for 2 min; 96% ethanol for 2 min; 70% ethanol for 2 min; and were then air dried.

cDNA synthesis

The following experiments demonstrate that mRNA captured on the array from the tissue sample sections may be used as template for cDNA synthesis.

cDNA synthesis on chip

A 16 well mask and Chip Clip slide holder from Whatman was attached to a CodeLink slide. The SuperScript™III One-step RT-PCR System with Platinum®Taq DNA Polymerase from Invitrogen was used when performing the cDNA synthesis. For each reaction 25 µl 2x reaction mix (SuperScript™III One-step RT-PCR System with Platinum®Taq DNA Polymerase, Invitrogen), 22.5 µl H_2O and 0.5 µl 100xBSA were mixed and heated to 50°C. SuperScript III/Platinum Taq enzyme mix was added to the reaction mix, 2 µl per reaction, and 50 µl of the reaction mix was added to each well on the chip. The chip was incubated at 50°C for 30 min (Thermomixer Comfort, Eppendorf).

The reaction mix was removed from the wells and the slide was washed with: 2xSSC, 0.1% SDS at 50°C for 10 min; 0.2xSSC at room temperature for 1 min; and 0.1xSSC at room temperature for 1 min. The chip was then spin dried.

In the case of FFPE tissue sections, the sections could now be stained and visualized before removal of the tissue, see below section on visualization.

Visualization

Hybridization of fluorescent marker probes prior to staining

Prior to tissue application fluorescent marker probes were hybridized to features comprising marker oligonucleotides printed on the capture probe array. The fluorescent marker probes aid in the orientation of the resulting image after tissue visualization, making it possible to combine the image with the resulting expression profiles for individual capture probe "tag" (positional domain) sequences obtained after sequencing. To hybridize fluorescent probes a hybridization solution containing 4xSSC and 0.1% SDS, 2µM detection probe (P) was incubated for 4 min at 50°C. Meanwhile the in-house array was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50µL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry.

General histological staining of FFPE tissue sections prior to or post cDNA synthesis

FFPE tissue sections immobilized on capture probe arrays were washed and rehydrated after deparaffinization prior to cDNA synthesis as described previously, or washed after cDNA synthesis as described previously. They are then treated as follows: incubate for 3 minutes in Hematoxylin; rinse with deionized water; incubate 5 minutes in tap water; rapidly dip 8 to 12 times in acid ethanol; rinse 2x1 minute in tap water; rinse 2 minutes in deionized water; incubate 30 seconds in Eosin; wash 3x5 minutes in 95% ethanol; wash 3x5 minutes in 100% ethanol; wash 3x10 minutes in xylene (can be done overnight); place coverslip on slides using DPX; dry slides in the hood overnight.

General immunohistochemistry staining of a target protein in FFPE tissue sections prior to or post cDNA synthesis

FFPE tissue sections immobilized on capture probe arrays were washed and rehydrated after deparaffinization prior to cDNA synthesis as described previously, or washed after cDNA synthesis as described previously. They were then treated as follows without being allowed to dry during the whole staining process: sections were incubated with primary antibody (dilute primary antibody in blocking solution comprising 1xTris Buffered Saline (50mM Tris, 150mM NaCl, pH 7.6), 4% donkey serum and 0.1% triton-x) in a wet chamber overnight at RT; rinse three times with 1xTBS; incubate section with matching secondary antibody

conjugated to a fluorochrome (FITC, Cy3 or Cy5) in a wet chamber at RT for 1 hour. Rinse 3x with 1xTBS, remove as much as possible of TBS and mount section with ProLong Gold +DAPI (Invitrogen) and analyze with fluorescence microscope and matching filter sets.

Removal of residual tissue

Frozen tissue

For fresh frozen mouse brain tissue the washing step directly following cDNA synthesis was enough to remove the tissue completely.

FFPE tissue

The slides with attached formalin fixed paraffinized mouse brain tissue sections were attached to ChipClip slide holders and 16 well masks (Whatman). For each 150 µl Proteinase K Digest Buffer from the RNeasy FFPE kit (Qiagen), 10 µl Proteinase K Solution (Qiagen) was added. 50 µl of the final mixture was added to each well and the slide was incubated at 56°C for 30 min.

Capture probe (cDNA) release

Capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes)

A 16 well mask and CodeLink slide was attached to the ChipClip holder (Whatman). 50µl of a mixture containing 1x FastStart High Fidelity Reaction Buffer with 1.8 mM MgCl2 (Roche), 200 µM dNTPs (New England Biolabs) and 0.1U/1 µl USER Enzyme (New England Biolabs) was heated to 37°C and was added to each well and incubated at 37°C for 30 min with mixing (3 seconds at 300 rpm, 6 seconds at rest) (Thermomixer comfort; Eppendorf). The reaction mixture containing the released cDNA and probes was then recovered from the wells with a pipette.

Capture probe release with uracil cleaving USER enzyme mixture in TdT (terminal transferase) buffer (covalently attached probes)

50µl of a mixture containing: 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com); 0.1µg/µl BSA (New England Biolabs); and 0.1U/µl USER Enzyme (New England Biolabs) was heated to 37°C and was added to each well and incubated at 37°C for 30 min with mixing (3 seconds at 300 rpm, 6 seconds at rest)

(Thermomixer comfort; Eppendorf). The reaction mixture containing the released cDNA and probes was then recovered from the wells with a pipette.

Capture probe release with boiling hot water (covalently attached probes)

A 16 well mask and CodeLink slide was attached to the ChipClip holder (Whatman). 50µl of 99°C water was pipetted into each well. The 99°C water was allowed to react for 30 minutes. The reaction mixture containing the released cDNA and probes was then recovered from the wells with a pipette.

Capture probe release with heated PCR buffer (hybridized *in situ* synthesized capture probes, i.e. capture probes hybridized to surface probes)

50µl of a mixture containing: 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com); 0.1µg/µl BSA (New England Biolabs); and 0.1U/µl USER Enzyme (New England Biolabs) was preheated to 95°C. The mixture was then added to each well and incubated for 5 minutes at 95°C with mixing (3 seconds at 300 rpm, 6 seconds at rest) (Thermomixer comfort; Eppendorf). The reaction mixture containing the released probes was then recovered from the wells.

Capture probe release with heated TdT (terminal transferase) buffer (hybridized *in situ* synthesized capture probes, i.e. capture probes hybridized to surface probes)

50µl of a mixture containing: 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com); 0.1µg/µl BSA (New England Biolabs); and 0.1U/µl USER Enzyme (New England Biolabs) was preheated to 95°C. The mixture was then added to each well and incubated for 5 minutes at 95°C with mixing (3 seconds at 300 rpm, 6 seconds at rest) (Thermomixer comfort; Eppendorf). The reaction mixture containing the released probes was then recovered from the wells.

The efficacy of treating the array with the USER enzyme and water heated to 99°C can be seen in Figure 3. Enzymatic cleavage using the USER enzyme and the RsaI enzyme was performed using the "in-house" arrays described above (Figure 3). Hot water mediated release of DNA surface probes was also performed using commercial arrays manufactured by Agilent.

Probe collection and linker introduction

The experiments demonstrate that first strand cDNA released from the array surface may be modified to produce double stranded DNA and subsequently amplified.

Whole Transcriptome Amplification by the Picoplex whole genome amplification kit (capture probe sequences including positional domain (tag) sequences not retained at the edge of the resulting dsDNA)

Capture probes were released with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached capture probes) or with heated PCR buffer (hybridized *in situ* synthesized capture probes, i.e. capture probes hybridized to surface probes).

The released cDNA was amplified using the Picoplex (Rubicon Genomics) random primer whole genome amplification method, which was carried out according to manufacturers instructions.

Whole Transcriptome Amplification by dA tailing with Terminal Transferase (TdT) (capture probe sequences including positional domain (tag) sequences retained at the end of the resulting dsDNA)

Capture probes were released with uracil cleaving USER enzyme mixture in TdT (terminal transferase) buffer (covalently attached capture probes) or with heated TdT (terminal transferase) buffer (hybridized *in situ* synthesized capture probes, i.e. capture probes hybridized to surface probes).

38µl of cleavage mixture was placed in a clean 0.2ml PCR tube. The mixture contained: 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com), 0.1µg/µl BSA (New England Biolabs); 0.1U/µl USER Enzyme (New England Biolabs) (not for heated release); released cDNA (extended from surface probes); and released surface probes. To the PCR tube, 0.5µl RNase H (5U/µl, final concentration of 0.06U/µl), 1µl TdT (20U/µl, final concentration of 0.5U/µl), and 0.5µl dATPs (100mM, final concentration of 1.25mM), were added. For dA tailing, the tube was incubated in a thermocycler (Applied Biosystems) at 37°C for 15 min followed by an inactivation of TdT at 70°C for 10 min. After dA tailing, a PCR master mix was prepared. The mix contained: 1x Faststart HiFi PCR Buffer (pH 8.3) with 1.8mM $MgCl_2$ (Roche); 0.2mM of each dNTP (Fermentas); 0.2µM of each primer, A (complementary to the amplification domain of the capture probe) and B_(dT)24 (Eurofins MWG Operon) (complementary to the poly-A tail to be added to the 3' end

of the first cDNA strand); and 0.1U/µl Faststart HiFi DNA polymerase (Roche). 23µl of PCR Master mix was placed into nine clean 0.2ml PCR tubes. 2µl of dA tailing mixture were added to eight of the tubes, while 2µl water (RNase/DNase free) was added to the last tube (negative control). PCR amplification was carried out with the following program: Hot start at 95˚C for 2 minutes, second strand synthesis at 50˚C for 2 minutes and 72˚C for 3 minutes, amplification with 30 PCR cycles at 95˚C for 30 seconds, 65˚C for 1 minutes, 72˚C for 3 minutes, and a final extension at 72˚C for 10 minutes.

Post-reaction cleanup and analysis

Four amplification products were pooled together and were processed through a Qiaquick PCR purification column (Qiagen) and eluted into 30µl EB (10mM Tris-Cl, pH 8.5). The product was analyzed on a Bioanalyzer (Agilent). A DNA 1000 kit was used according to manufacturers instructions.

Sequencing

Illumina sequencing

dsDNA library for Illumina sequencing using sample indexing was carried out according to manufacturers instructions. Sequencing was carried out on an HiSeq2000 platform (Illumina).

Bioinformatics

Obtaining digital transcriptomic information from sequencing data from whole transcriptome libraries amplified using the dA tailing terminal transferase approach

The sequencing data was sorted through the FastX toolkit FASTQ Barcode splitter tool into individual files for the respective capture probe positional domain (tag) sequences. Individually tagged sequencing data was then analyzed through mapping to the mouse genome with the Tophat mapping tool. The resulting SAM file was processed for transcript counts through the HTseq-count software.

Obtaining digital transcriptomic information from sequencing data from whole transcriptome libraries amplified using the Picoplex whole genome amplification kit approach

The sequencing data was converted from FASTQ format to FASTA format using the FastX toolkit FASTQ-to-FASTA converter. The sequencing reads was aligned to the capture probe positional domain (tag) sequences using Blastn and

the reads with hits better than $1e^{-6}$ to one of tag sequences were sorted out to individual files for each tag sequence respectively. The file of tag sequence reads was then aligned using Blastn to the mouse transcriptome, and hits were collected.

Combining visualization data and expression profiles

The expression profiles for individual capture probe positional domain (tag) sequences are combined with the spatial information obtained from the tissue sections through staining. Thereby the transcriptomic data from the cellular compartments of the tissue section can be analyzed in a directly comparative fashion, with the availability to distinguish distinct expression features for different cellular subtypes in a given structural context

Example 2

Stained FFPE mouse brain tissue (olfactory bulb) sections were placed on top of a bar-coded transcriptome capture array, according to the general procedure described in Example 1. As compared with the experiment with fresh frozen tissue in Example 1, better morphology was observed with the FFPE tissue.

Example 3

Whole Transcriptome Amplification by Random primer second strand synthesis followed by universal handle amplification (capture probe sequences including tag sequences retained at the end of the resulting dsDNA)

Following capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes)

OR

Following capture probe release with heated PCR buffer (hybridized *in situ* synthesized capture probes)

1μl RNase H (5U/μl) was added to each of two tubes, final concentration of 0.12U/μl, containing 40μl 1x Faststart HiFi PCR Buffer (pH 8.3) with 1.8mM $MgCl_2$ (Roche, www.roche-applied-science.com), 0.2mM of each dNTP (Fermentas, www.fermentas.com), 0.1μg/μl BSA (New England Biolabs, www.neb.com), 0.1U/μl USER Enzyme (New England Biolabs), released cDNA (extended from surface probes) and released surface probes. The tubes were incubated at 37°C for 30 min followed by 70°C for 20 min in a thermo cycler (Applied Biosystems, www.appliedbiosystems.com). 1μl Klenow Fragment (3' to 5' exo minus) (Illumina,

www.illumina.com) and 1μl handle coupled random primer (10μM) (Eurofins MWG Operon, www.eurofinsdna.com) was added to the two tubes (B_R8 (octamer) to one of the tubes and B_R6 (hexamer) to the other tube), final concentration of 0.23μM. The two tubes were incubated at 15°C for 15 min, 25°C for 15 min, 37°C for 15 min and finally 75°C for 20 min in a thermo cycler (Applied Biosystems). After the incubation, 1μl of each primer, A_P and B (10μM) (Eurofins MWG Operon), was added to both tubes, final concentration of 0.22μM each. 1μl Faststart HiFi DNA polymerase (5U/μl) (Roche) was also added to both tubes, final concentration of 0.11U/μl. PCR amplification were carried out in a thermo cycler (Applied Biosystems) with the following program: Hot start at 94˚C for 2 min, followed by 50 cycles at 94˚C for 15 seconds, 55˚C for 30 seconds, 68˚C for 1 minute, and a final extension at 68˚C for 5 minutes. After the amplification, 40μl from each of the two tubes were purified with Qiaquick PCR purification columns (Qiagen, www.qiagen.com) and eluted into 30μl EB (10mM Tris-Cl, pH 8.5). The Purified products were analyzed with a Bioanalyzer (Agilent, www.home.agilent.com), DNA 7500 kit were used. This Example demonstrates the use of random hexamer and random octamer second strand synthesis, followed by amplification to generate the population from the released cDNA molecules.

Example 4

Amplification of ID-specific and gene specific products after cDNA synthesis and probe collection

Following capture probe release with uracil cleaving USER enzyme mixture in PCR buffer (covalently attached probes).

The cleaved cDNA was amplified in final reaction volumes of 10 µl. 7 µl cleaved template, 1 µl ID-specific forward primer (2 µM), 1 µl gene-specific reverse primer (2 µM) and 1 µl FastStart High Fidelity Enzyme Blend in 1.4x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ to give a final reaction of 10 µl with 1x FastStart High Fidelity Reaction Buffer with 1.8 mM $MgCl_2$ and 1 U FastStart High Fidelity Enzyme Blend. PCR amplification were carried out in a thermo cycler (Applied Biosystems) with the following program: Hot start at 94˚C for 2 min, followed by 50 cycles at 94˚C for 15 seconds, 55˚C for 30 seconds, 68˚C for 1 minute, and a final extension at 68˚C for 5 minutes.

Primer sequences, resulting in a product of approximately 250 bp,

Beta-2 microglobulin (B2M) primer
5'-TGGGGGTGAGAATTGCTAAG-3' (SEQ ID NO: 43)
ID-1 primer
5'-CCTTCTCCTTCTCCTTCACC-3' (SEQ ID NO: 44)
ID-5 primer
5'-GTCCTCTATTCCGTCACCAT-3' (SEQ ID NO: 45)
ID-20 primer
5'-CTGCTTCTTCCTGGAACTCA-3' (SEQ ID NO: 46)

The results show successful amplification of ID-specific and gene-specific products using two different ID primers (i.e. specific for ID tags positioned at different locations on the microarray and the same gene specific primer from a brain tissue covering all the probes). Accordingly this experiment establishes that products may be identified by an ID tag-specific or target nucleic acid specific amplification reaction. It is further established that different ID tags may be distinguished. A second experiment, with tissue covering only half of the ID probes (i.e. capture probes) on the array resulted in a positive result (PCR product) for spots that were covered with tissue.

Example 5

Alternative synthesis of 5' to 3' oriented capture probes using polymerase extension and terminal transferase tailing

To hybridize primers for capture probe synthesis hybridization solution containing 4xSSC and 0.1% SDS and 2 μM extension primer (A_primer) was incubated for 4 min at 50°C. Meanwhile the in-house array (see Example 1) was attached to a ChipClip (Whatman). The array was subsequently incubated at 50°C for 30 min at 300 rpm shake with 50 μL of hybridization solution per well.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry and placed back in the ChipClip.

1μl Klenow Fragment (3' to 5' exo minus) (Illumina, www.illumina.com) together with 10x Klenow buffer, dNTPs 2 mM each (Fermentas) and water, was mixed into a 50 μl reaction and was pipetted into each well.

The array was incubated at 15°C for 15 min, 25°C for 15 min, 37°C for 15 min and finally 75°C for 20 min in an Eppendorf Thermomixer.

After incubation, the array was removed from the ChipClip and washed with the 3 following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 6 min at 300 rpm shake, 2) 0.2xSSC for 1 min at 300 rpm shake and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was then spun dry and placed back in the ChipClip.

For dT tailing a 50 µl reaction mixture containing 1x TdT buffer (20mM Tris-acetate (pH 7.9), 50mM Potassium Acetate and 10mM Magnesium Acetate) (New England Biolabs, www.neb.com), 0.1µg/µl BSA (New England Biolabs), 0.5µl RNase H (5U/µl) , 1µl TdT (20U/µl) and 0.5µl dTTPs (100mM) was prepared. The mixture was added to the array surface and the array was incubated in a thermo cycler (Applied Biosystems) at 37°C for 15 min followed by an inactivation of TdT at 70°C for 10 min.

Example 6

Spatial transcriptomics using 5' to 3' high probe density arrays and formalin-fixed frozen (FF-frozen) tissue with USER system cleavage and amplification via terminal transferase

Array preparation

Pre-fabricated high-density microarrays chips were ordered from Roche-Nimblegen (Madison, WI, USA). Each capture probe array contained 135,000 features of which 132,640 features carried a capture probe comprising a unique ID-tag sequence (positional domain) and a capture region (capture domain). Each feature was 13x13 µm in size. The capture probes were composed 5' to 3' of a universal domain containing five dUTP bases (a cleavage domain) and a general amplification domain, an ID tag (positional domain) and a capture region (capture domain) (Figure 6 and Table 4). Each array was also fitted with a frame of marker probes carrying a generic 30 bp sequence (Table 4) to enable hybridization of fluorescent probes to help with orientation during array visualization.

Tissue preparation – preparation of formalin-fixed frozen tissue

The animal (mouse) was perfused with 50ml PBS and 100ml 4% formalin solution. After excision of the olfactory bulb, the tissue was put into a 4% formalin bath for post-fixation for 24 hrs. The tissue was then sucrose treated in 30% sucrose dissolved in PBS for 24 hrs to stabilize morphology and to remove excess

formalin. The tissue was frozen at a controlled rate down to -40°C and kept at -20°C between experiments. Similar preparation of tissue postfixed for 3 hrs or without post-fixation was carried out for a parallel specimen. Perfusion with 2% formalin without post-fixation was also used successfully. Similarly the sucrose treatment step could be omitted. The tissue was mounted into a cryostat for sectioning at 10µm. A slice of tissue was applied onto each capture probe array to be used. Optionally for better tissue adherence, the array chip was placed at 50°C for 15 minutes.

Optional control - Total RNA preparation from sectioned tissue

Total RNA was extracted from a single tissue section (10µm) using the RNeasy FFPE kit (Qiagen) according to manufacturers instructions. The total RNA obtained from the tissue section was used in control experiments for a comparison with experiments in which the RNA was captured on the array directly from the tissue section. Accordingly, in the case where total RNA was applied to the array the staining, visualization and degradation of tissue steps were omitted.

On-chip reactions

The hybridization of marker probe to the frame probes, reverse transcription, nuclear staining, tissue digestion and probe cleavage reactions were all performed in a 16 well silicone gasket (ArrayIt, Sunnyvale, CA, USA) with a reaction volume of 50 µl per well. To prevent evaporation, the cassettes were covered with plate sealers (In Vitro AB, Stockholm, Sweden).

Optional - tissue permeabilization prior to cDNA synthesis

For permeabilization using Proteinase K, proteinase K (Qiagen, Hilden, Germany) was diluted to 1 µg/ml in PBS. The solution was added to the wells and the slide incubated at room temperature for 5 minutes, followed by a gradual increase to 80°C over 10 minutes. The slide was washed briefly in PBS before the reverse transcription reaction.

Alternatively for permeabilization using microwaves, after tissue attachment, the slide was placed at the bottom of a glass jar containing 50ml 0.2xSSC (Sigma-Aldrich) and was heated in a microwave oven for 1 minute at 800W. Directly after microwave treatment the slide was placed onto a paper tissue and was dried for 30 minutes in a chamber protected from unnecessary air exposure. After drying, the slide was briefly dipped in water (RNase/DNase free) and finally spin-dried by a centrifuge before cDNA synthesis was initiated.

cDNA synthesis

For the reverse transcription reaction the SuperScript III One-Step RT-PCR System with Platinum Taq (Life Technologies/Invitrogen, Carlsbad, CA, USA) was used. Reverse transcription reactions contained 1x reaction mix, 1x BSA (New England Biolabs, Ipswich, MA, USA) and 2 µl SuperScript III RT/Platinum Taq mix in a final volume of 50 µl. This solution was heated to 50°C before application to the tissue sections and the reaction was performed at 50°C for 30 minutes. The reverse transcription solution was subsequently removed from the wells and the slide was allowed to air dry for 2 hours.

Tissue visualization

After cDNA synthesis, nuclear staining and hybridization of the marker probe to the frame probes (probes attached to the array substrate to enable orientation of the tissue sample on the array) was done simultaneously. A solution with DAPI at a concentration of 300nM and marker probe at a concentration of 170 nM in PBS was prepared. This solution was added to the wells and the slide was incubated at room temperature for 5 minutes, followed by brief washing in PBS and spin drying.

Alternatively the marker probe was hybridized to the frame probes prior to placing the tissue on the array. The marker probe was then diluted to 170 nM in hybridization buffer (4xSSC, 0.1% SDS). This solution was heated to 50°C before application to the chip and the hybridization was performed at 50°C for 30 minutes at 300 rpm. After hybridization, the slide was washed in 2xSSC, 0.1% SDS at 50°C and 300 rpm for 10 minutes, 0.2xSSC at 300 rpm for 1 minute and 0.1xSSC at 300 rpm for 1 minute. In that case the staining solution after cDNA synthesis only contained the nuclear DAPI stain diluted to 300nM in PBS. The solution was applied to the wells and the slide was incubated at room temperature for 5 minutes, followed by brief washing in PBS and spin drying.

The sections were microscopically examined with a Zeiss Axio Imager Z2 and processed with MetaSystems software.

Tissue removal

The tissue sections were digested using Proteinase K diluted to 1.25 µg/µl in PKD buffer from the RNeasy FFPE Kit (both from Qiagen) at 56°C for 30 minutes with an interval mix at 300 rpm for 3 seconds, then 6 seconds rest. The slide was subsequently washed in 2xSSC, 0.1% SDS at 50°C and 300 rpm for 10 minutes, 0.2xSSC at 300 rpm for 1 minute and 0.1xSSC at 300 rpm for 1 minute.

Probe release

The 16-well Hybridization Cassette with silicone gasket (ArrayIt) was preheated to 37°C and attached to the Nimblegen slide. A volume of 50µl of cleavage mixture preheated to 37°C, consisting of Lysis buffer at an unknown concentration (Takara), 0.1U/µl USER Enzyme (NEB) and 0.1µg/µl BSA was added to each of wells containing surface immobilized cDNA. After removal of bubbles the slide was sealed and incubated at 37°C for 30 minutes in a Thermomixer comfort with cycled shaking at 300 rpm for 3 seconds with 6 seconds rest in between. After the incubation 45µl cleavage mixture was collected from each of the used wells and placed into 0.2ml PCR tubes.

Library preparation

Exonuclease treatment

After cooling the solutions on ice for 2 minutes, Exonuclease I (NEB) was added, to remove unextended cDNA probes, to a final volume of 46.2µl and a final concentration of 0.52U/µl. The tubes were incubated in a thermo cycler (Applied Biosystems) at 37°C for 30 minutes followed by inactivation of the exonuclease at 80°C for 25 minutes.

dA-tailing by terminal transferase

After the exonuclease step, 45µl polyA-tailing mixture, according to manufacturers instructions consisting of TdT Buffer (Takara), 3mM dATP (Takara) and manufacturers TdT Enzyme mix (TdT and RNase H) (Takara), was added to each of the samples. The mixtures were incubated in a thermocycler at 37°C for 15 minutes followed by inactivation of TdT at 70°C for 10 minutes.

Second-strand synthesis and PCR-amplification

After dA-tailing, 23µl PCR master mix was placed into four new 0.2ml PCR tubes per sample, to each tube 2µl sample was added as a template. The final PCRs consisted of 1x Ex Taq buffer (Takara), 200µM of each dNTP (Takara), 600nM A_primer (MWG), 600nM B_dT20VN_primer (MWG) and 0.025U/µl Ex Taq polymerase (Takara)(Table 4). A second cDNA strand was created by running one cycle in a thermocycler at 95°C for 3 minutes, 50°C for 2 minutes and 72°C for 3 minutes. Then the samples were amplified by running 20 cycles (for library preparation) or 30 cycles (to confirm the presence of cDNA) at 95°C for 30

seconds, 67°C for 1 minute and 72°C for 3 minutes, followed by a final extension at 72°C for 10 minutes.

Library cleanup

After amplification, the four PCRs (100µl) were mixed with 500µl binding buffer (Qiagen) and placed in a Qiaquick PCR purification column (Qiagen) and spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane. The membrane was then washed with wash buffer (Qiagen) containing ethanol and finally eluted into 50µl of 10mM Tris-Cl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification (purification by superparamagnetic beads conjugated to carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The amplified cDNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and were then eluted into 15µl of 10mM Tris-Cl, pH 8.5.

Library Quality analysis

Samples amplified for 30 cycles were analyzed with an Agilent Bioanalyzer (Agilent) in order to confirm the presence of an amplified cDNA library, the DNA High Sensitivity kit or DNA 1000 kit were used depending on the amount of material.

Sequencing library preparation

Library indexing

Samples amplified for 20 cycles were used further to prepare sequencing libraries. An index PCR master mix was prepared for each sample and 23µl was placed into six 0.2ml tubes. 2µl of the amplified and purified cDNA was added to each of the six PCRs as template making the PCRs containing 1x Phusion master mix (Fermentas), 500nM InPE1.0 (Illumina), 500nM Index 1-12 (Illumina), and 0.4nM InPE2.0 (Illumina). The samples were amplified in a thermocycler for 18 cycles at 98°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 minute, followed by a final extension at 72°C for 5 minutes.

Sequencing library cleanup

After amplification, the six PCRs (150µl) were mixed with 750µl binding buffer and placed in a Qiaquick PCR purification column and spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane (because of the

large sample volume (900µl), the sample was split in two (each 450µl) and was bound in two separate steps). The membrane was then washed with wash buffer containing ethanol and finally eluted into 50µl of 10mM Tris-Cl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification with an MBS robot. A final PEG concentration of 7.8% was used in order to remove fragments below 300-350bp. The amplified cDNA was allowed to bind to the CA-beads for 10 min and were then eluted into 15µl of 10mM Tris-Cl, pH 8.5. Samples were analyzed with an Agilent Bioanalyzer in order to confirm the presence and size of the finished libraries, the DNA High Sensitivity kit or DNA 1000 kit were used according to manufacturers instructions depending on the amount of material.

Sequencing

The libraries were sequenced on the Illumina Hiseq2000 or Miseq depending on desired data throughput according to manufacturers instructions. Optionally for read 2, a custom sequencing primer B_r2 was used to avoid sequencing through the homopolymeric stretch of 20T.

Data analysis

Read 1 was trimmed 42 bases at 5' end. Read 2 was trimmed 25 bases at 5' end (optionally no bases were trimmed from read 2 if the custom primer was used). The reads were then mapped with bowtie to the repeat masked *Mus musculus* 9 genome assembly and the output was formatted in the SAM file format. Mapped reads were extracted and annotated with UCSC refGene gene annotations. Indexes were retrieved with 'indexFinder' (an inhouse software for index retrieval). A mongo DB database was then created containing information about all caught transcripts and their respective index position on the chip.

A matlab implementation was connected to the database and allowed for spatial visualization and analysis of the data (Figure 7).

Optionally the data visualization was overlaid with the microscopic image using the fluorescently labelled frame probes for exact alignment and enabling spatial transcriptomic data extraction.

Example 7

Spatial transcriptomics using 3' to 5' high probe density arrays and FFPE tissue with MutY system cleavage and amplification via TdT

Array preparation

Pre-fabricated high-density microarrays chips were ordered from Roche-Nimblegen (Madison, WI, USA). Each used capture probe array contained 72k features out of which 66,022 contained a unique ID-tag complementary sequence. Each feature was 16x16 µm in size. The capture probes were composed 3' to 5' in the same way as the probes used for the in-house printed 3' to 5' arrays with the exeception to 3 additional bases being added to the upper (P') general handle of the probe to make it a long version of P', LP' (Table 4). Each array was also fitted with a frame of probes carrying a generic 30 bp sequence to enable hybridization of fluorescent probes to help with orientation during array visualization.

Synthesis of 5' to 3' oriented capture probes

The synthesis of 5' to 3' oriented capture probes on the high-density arrays was carried out as in the case with in-house printed arrays, with the exception that the extension and ligation steps were carried out at 55°C for 15 mins followed by 72°C for 15 mins. The A-handle probe (Table 4) included an A/G mismatch to allow for subsequent release of probes through the MutY enzymatic system described below. The P-probe was replaced by a longer LP version to match the longer probes on the surface.

Preparation of formalin-fixed paraffin-embedded tissue and deparaffinization

This was carried out as described above in the in-house protocol.

cDNA synthesis and staining

cDNA synthesis and staining was carried out as in the protocol for 5' to 3' oriented high-density Nimblegen arrays with the exception that biotin labelled dCTPs and dATPs were added to the cDNA synthesis together with the four regular dNTPs (each was present at 25x times more than the biotin labelled ones).

Tissue removal

Tissue removal was carried out in the same way as in the protocol for 5' to 3' oriented high-density Nimblegen arrays described in Example 6.

Probe cleavage by MutY

A 16-well Incubation chamber with silicone gasket (ArrayIT) was preheated to 37°C and attached to the Codelink slide. A volume of 50µl of cleavage mixture preheated to 37°C, consisting of 1x Endonucelase VIII Buffer (NEB), 10U/µl MutY (Trevigen), 10U/µl Endonucelase VIII (NEB), 0.1µg/µl BSA was added to each of wells containing surface immobilized cDNA. After removal of bubbles the slide was sealed and incubated at 37°C for 30 minutes in a Thermomixer comfort with cycled

shaking at 300rpm for 3 seconds with 6 seconds rest in between. After the incubation, the plate sealer was removed and 40µl cleavage mixture was collected from each of the used wells and placed into a PCR plate.

Library preparation

Biotin-streptavidin mediated library cleanup

To remove unextended cDNA probes and to change buffer, the samples were purified by binding the biotin labeled cDNA to streptavidin coated C1-beads (Invitrogen) and washing the beads with 0.1M NaOH (made fresh). The purification was carried out with an MBS robot (Magnetic Biosolutions), the biotin labelled cDNA was allowed to bind to the C1-beads for 10 min and was then eluted into 20µl of water by heating the bead-water solution to 80°C to break the biotin-streptavidin binding.

dA-tailing by terminal transferase

After the purification step, 18µl of each sample was placed into new 0.2ml PCR tubes and mixed with 22µl of a polyA-tailing master mix leading to a 40µl reaction mixture according to manufacturers instructions consisting of lysis buffer (Takara, Cellamp Whole Transcriptome Amplification kit), TdT Buffer (Takara), 1.5mM dATP (Takara) and TdT Enzyme mix (TdT and RNase H) (Takara). The mixtures were incubated in a thermocycler at 37°C for 15 minutes followed by inactivation of TdT at 70°C for 10 minutes.

Second-strand synthesis and PCR-amplification

After dA-tailing, 23µl PCR master mix was placed into four new 0.2ml PCR tubes per sample, to each tube 2µl sample was added as a template. The final PCRs consisted of 1x Ex Taq buffer (Takara), 200µM of each dNTP (Takara), 600nM A_primer (MWG), 600nM B_dT20VN_primer (MWG) and 0.025U/µl Ex Taq polymerase (Takara). A second cDNA strand was created by running one cycle in a thermo cycler at 95°C for 3 minutes, 50°C for 2 minutes and 72°C for 3 minutes. Then the samples were amplified by running 20 cycles (for library preparation) or 30 cycles (to confirm the presence of cDNA) at 95°C for 30 seconds, 67°C for 1 minute and 72°C for 3 minutes, followed by a final extension at 72°C for 10 minutes.

Library cleanup

After amplification, the four PCRs (100µl) were mixed with 500µl binding buffer (Qiagen) and placed in a Qiaquick PCR purification column (Qiagen) and

spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane. The membrane was then washed with wash buffer (Qiagen) containing ethanol and finally eluted into 50µl of 10mM Tris-HCl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification (purification by superparamagnetic beads conjugated to carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The amplified cDNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and were then eluted into 15µl of 10mM Tris-HCl, pH 8.5.

Second PCR-amplification

The final PCRs consisted of 1x Ex Taq buffer (Takara), 200µM of each dNTP (Takara), 600nM A_primer (MWG), 600nM B_ primer (MWG) and 0.025U/µl Ex Taq polymerase (Takara). The samples were heated to 95°C for 3 minutes, and then amplified by running 10 cycles at 95°C for 30 seconds, 65°C for 1 minute and 72°C for 3 minutes, followed by a final extension at 72°C for 10 minutes.

Second library cleanup

After amplification, the four PCRs (100µl) were mixed with 500µl binding buffer (Qiagen) and placed in a Qiaquick PCR purification column (Qiagen) and spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane. The membrane was then washed with wash buffer (Qiagen) containing ethanol and finally eluted into 50µl of 10mM Tris-Cl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification (purification by super-paramagnetic beads conjugated to carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The amplified cDNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and were then eluted into 15µl of 10mM Tris-HCl, pH 8.5.

Sequencing library preparation

Library indexing

Samples amplified for 20 cycles were used further to prepare sequencing libraries. An index PCR master mix was prepared for each sample and 23µl was placed into six 0.2ml tubes. 2µl of the amplified and purified cDNA was added to each of the six PCRs as template making the PCRs containing 1x Phusion master

mix (Fermentas), 500nM InPE1.0 (Illumina), 500nM Index 1-12 (Illumina), and 0.4nM InPE2.0 (Illumina). The samples were amplified in a thermo cycler for 18 cycles at 98°C for 30 seconds, 65°C for 30 seconds and 72°C for 1 minute, followed by a final extension at 72°C for 5 minutes.

Sequencing library cleanup

After amplification, the samples was purified and concentrated by CA-purification with an MBS robot. A final PEG concentration of 7.8% was used in order to remove fragments below 300-350bp. The amplified cDNA was allowed to bind to the CA-beads for 10 min and were then eluted into 15μl of 10mM Tris-HCl, pH 8.5.

10μl of the amplified and purified samples were placed on a Caliper XT chip and fragments between 480bp and 720bp were cut out with the Caliper XT (Caliper). Samples were analyzed with an Agilent Bioanalyzer in order to confirm the presence and size of the finished libraries, the DNA High Sensitivity kit was used.

Sequencing and Data analysis

Sequencing and Bioinformatic was carried out in the same way as in the protocol for 5' to 3' oriented high-density Nimblegen arrays described in Example 6. However, in the data analysis, read 1 was not used in the mapping of transcripts. Specific Olfr transcripts could be sorted out using the Matlab visualization tool (Figure 8).

Example 8

Spatial transcriptomics using in house printed 41-tag microarray with 5' to 3' oriented probes and formalin-fixed frozen (FF-frozen) tissue with permeabilization through ProteinaseK or microwaving with USER system cleavage and amplification via TdT

Array preparation

In-house arrays were printed as previously described but with a pattern of 41 unique ID-tag probes with the same composition as the probes in the 5' to 3' oriented high-density array in Example 6 (Figure 9).

All other steps were carried out in the same way as in the protocol described in Example 6.

Example 9

Alternative method for performing the cDNA synthesis step

cDNA synthesis on chip as described above can also be combined with template switching to create a second strand by adding a template switching primer to the cDNA synthesis reaction (Table 4). The second amplification domain is introduced by coupling it to terminal bases added by the reverse transcriptase at the 3' end of the first cDNA strand, and primes the synthesis of the second strand. The library can be readily amplified directly after release of the double-stranded complex from the array surface.

Example 10

The following experiments demonstrate how the secured (captured) cDNA molecules can be labelled and detected on the surface of the object substrate, e.g. array.

Preparation of in-house printed microarray with 5' to 3' oriented probes

The RNA-capture oligonucleotide (Table 2) was printed on glass slides to function as the capture probe. The probe was synthesized with a 5'-terminus amino linker with a C6 spacer. All probes where synthesized by Sigma-Aldrich (St. Louis, MO, USA). The RNA-capture probe was suspended at a concentration of 20 µM in 150 mM sodium phosphate, pH 8.5 and spotted using a pipette onto CodeLink™ Activated microarray slides (7.5cm x 2.5cm; Surmodics, Eden Prairie, MN, USA). Each array was printed with 10µl of capture probe-containing solution, and left to dry. After printing, surface blocking was performed according to the manufacturer's instructions. The probes were printed in 16 arrays on the slide. The 16 sub-arrays were separated during reaction steps by a 16-pad mask (Arrayit Corporation, Sunnyvale, CA, USA).

Preparation of fresh frozen tissue and sectioning onto capture probe arrays

Fresh non-fixed mouse brain tissue was trimmed if necessary and frozen in -40°C cold isopentane and subsequently mounted for sectioning with a cryostat at 10µm. A slice of tissue was applied onto each probe array.

Fixation of tissue section using formalin

50µl of 4% paraformaldehyde dissolved in PBS was added directly to the probe array to cover the tissue section. The array was incubated at room temperature for 10 minutes and then washed for 10 seconds in PBS. The array was then incubated at 50°C for 15 minutes.

Permeabilization of the tissue sample using pepsin and HCl

Pepsin was diluted to 0.1% in 0.1M HCl and was preheated to 37°C. The array was attached to an ArrayIt 16-well mask and holder. 50µl of the pepsin/HCl mixture was added to each well. The array was incubated for 10 minutes at 37°C and the wells were washed with 100µl 0.1x SSC by pipetting.

cDNA synthesis with Cy3-dNTP

For each well a cDNA synthesis mixture (80µl) was prepared containing 4µl each of dATP/dGTP/dTTP (10mM), 4µl dCTP(2.5mM), 4µl Cy3-dCTP(1mM), 4µl DTT (0.1M), 1x BSA, 20U/µl Superscript III, 5U/µl RNaseOUT, 1x first strand buffer (Superscript III, Invitrogen) and MilliQ water. 70 µl of the reaction mixture was added to each well. The reactions were covered with a plastic sealer and incubated at 37°C overnight.

Washing

After incubation the array was removed from the ArrayIt mask and holder and washed using the following steps: 1) 50°C 2xSSC solution with 0.1% SDS for 10 min at 300 rpm shake; 2) 0.2xSSC for 1 min at 300 rpm shake; and 3) 0.1xSSC for 1 min at 300 rpm shake. The array was spun dry.

Tissue removal

The array was attached to an ArrayIt slide holder and 16 well mask (ArrayIt Corporation). 10 µl Proteinase K Solution (Qiagen) was added for each 150 µl Proteinase K Digest Buffer from the RNeasy FFPE kit (Qiagen). 50 µl of the final mixture was added to each well and the array was incubated at 56°C for 1 hour. The array was washed as described above.

Imaging

The array was imaged at 532 nm using an Agilent microarray scanner at 100% exposure and 5µm resolution (Figure 11 and Figure 12).

Table 2

Probe 1

Amino-C6-UUACACTCTTTCCCTACACGACGCTCTTCCGATCT GTCCGATATGATTGCCGCTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 47)

Example 11

The following experiments demonstrate how a portion of the secured (captured) cDNA molecules can be removed from selected parts of the surface of

the object substrate, e.g. array. This enables isolation of cDNA from limited parts of the tissue section.

The method described in Example 10 was performed up to and including the step of tissue removal. Followed by the steps described below.

cDNA removal by laser ablation

The array was mounted into a MMI Cellcut instrument (Molecular Machines and Industries AG, Glattbrugg, Switzerland) and the sections of the fluorescently labelled cDNA to be removed were marked for ablation by laser.

Imaging

The codelink glass chip was imaged at 532 nm using an Agilent microarray scanner at 100% exposure and 5µm resolution. Removal of ablated areas was verified (Figure 13a and b). The array was washed according to the procedure described in Example 10.

Release of remaining cDNA from the array

The array was attached to an ArrayIt slide holder and 16 well mask (ArrayIt Corporation). A cleavage mixture (50µl) containing 1x Exo I buffer (New England Biolabs, Ipswich, MA, USA), 1x BSA, RNase/DNase free water, 5U of USER enzyme mix (New England Biolabs) was added to each well. The reactions were covered with a plastic sealer and incubated at 37°C for 1 hour using interval mixing of 3 seconds at 300rpm and 6 seconds rest. After the incubation, 45µl cleavage mixture was collected from each of the used wells and placed into 0.2ml PCR tubes.

Exonuclease treatment

After cooling the solutions on ice for 2 minutes, Exonuclease I (NEB) was added, to remove unextended cDNA probes, to a final volume of 46.2µl and a final concentration of 0.52U/µl. The tubes were incubated in a thermocycler (Applied Biosystems) at 37°C for 30 minutes followed by inactivation of the exonuclease at 80°C for 25 minutes.

dA-tailing by terminal transferase

After the exonuclease step, 45µl polyA-tailing mixture consisting of TdT Buffer (Takara), 3mM dATP (Takara) and manufacturers TdT Enzyme mix (TdT and RNase H) (Takara), was added to each of the samples according to manufacturer's instructions. The mixtures were incubated in a thermo cycler at 37°C for 15 minutes followed by inactivation of TdT at 70°C for 10 minutes.

Second-strand synthesis and PCR-amplification

After dA-tailing, 23µl PCR master mix was placed into four new 0.2ml PCR tubes per sample. 2µl of sample was added to each tube as a template. The final PCRs consisted of 1x Ex Taq buffer (Takara), 200µM of each dNTP (Takara), 600nM A_primer (MWG), 600nM B_dT20VN_primer (MWG) and 0.025U/µl Ex Taq polymerase (Takara)(Table 3). A second cDNA strand was created by running one cycle in a thermocycler at 95°C for 3 minutes, 50°C for 2 minutes and 72°C for 3 minutes. The samples were amplified by running 20 cycles (for library preparation) or 30 cycles (to confirm the presence of cDNA) at 95°C for 30 seconds, 67°C for 1 minute and 72°C for 3 minutes, followed by a final extension at 72°C for 10 minutes.

Library cleanup

After amplification, the four PCRs (100µl) were mixed with 500µl binding buffer (Qiagen) and placed in a Qiaquick PCR purification column (Qiagen) and spun for 1 minute at 17,900 x g in order to bind the amplified cDNA to the membrane. The membrane was then washed with wash buffer (Qiagen) containing ethanol and finally eluted into 50µl of 10mM Tris-Cl, pH 8.5.

The purified and concentrated sample was further purified and concentrated by CA-purification (purification by superparamagnetic beads conjugated to carboxylic acid) with an MBS robot (Magnetic Biosolutions). A final PEG concentration of 10% was used in order to remove fragments below 150-200bp. The amplified cDNA was allowed to bind to the CA-beads (Invitrogen) for 10 min and were then eluted into 15µl of 10mM Tris-Cl, pH 8.5.

Library Quality analysis

Samples amplified for 30 cycles were analyzed with an Agilent Bioanalyzer (Agilent) in order to confirm the presence of an amplified cDNA library, the DNA High Sensitivity kit or DNA 1000 kit were used depending on the amount of material (Figure 14).

Table 3

Second strand synthesis and first PCR Amplification handles	
A_primer	ACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 48)
B_dt20VN_primer	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 49)

Table 4. Oligos used for spatial transcriptomics

Example 6

Nimblegen 5' to 3' arrays with free 3' end Array probes

	5' to 3'
Probe1 (SEQ ID NO: 50)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCCGATATGATTGCCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe2 (SEQ ID NO: 51)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGAGCCGGGTTCATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe3 (SEQ ID NO: 52)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTGAGGCACTCTGTTGGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe4 (SEQ ID NO: 53)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGATTAGTCGCCATTCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe5 (SEQ ID NO: 54)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTGAGGGTAGATGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe6 (SEQ ID NO: 55)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGGCCAATACTGTTATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe7 (SEQ ID NO: 56)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTACCCTGATTCGACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe8 (SEQ ID NO: 57)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCCACTTTCGCCGTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe9 (SEQ ID NO: 58)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCAACTTTGAGCAAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe10 (SEQ ID NO: 59)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCAATTCGGAATTCCGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe11 (SEQ ID NO: 60)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCCCAAGGTAATACATTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe12 (SEQ ID NO: 61)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCATTTCCTATTCGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe13 (SEQ ID NO: 62)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGCTAAATCTAACCGCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe14 (SEQ ID NO: 63)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAATTAAATTCTGATGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe15 (SEQ ID NO: 64)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCATTACATAGGTGCTAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe16 (SEQ ID NO: 65)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGACTTGCGCTCGCACTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe17 (SEQ ID NO: 66)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGTATCTCCCAAGTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe18 (SEQ ID NO: 67)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGCGCCTGTAATCCGCATTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe19 (SEQ ID NO: 68)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGCCACTCTTTAGGTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe20 (SEQ ID NO: 69)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTATGCAAGTGATTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe21 (SEQ ID NO: 70)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAAGCCACGTTTATACGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe22 (SEQ ID NO: 71)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTACCTGATTGCTGTATAACTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe23 (SEQ ID NO: 72)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGCGCATCTATCCTCTATTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe24 (SEQ ID NO: 73)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCACGCGTAGGACTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN

Probe25 (SEQ ID NO: 74)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCGACTAAGTATGTAGCGCTTTTTTTTTTTTTTTTTTTTVN
Frame probe	
Layout1 (SEQ ID NO: 75)	AAATTTCGTCTGCTATCGCGCTTCTGTACC
Fluorescent marker probe	
PS_1 (SEQ ID NO: 76)	GGTACAGAAGCGCGATAGCAG – Cy3
Second strand synthesis and first PCR Amplification handles	
A_primer (SEQ ID NO: 77)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
B_dt20VN_primer (SEQ ID NO: 78)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTVN
Custom sequencing primer	
B_r2 (SEQ ID NO: 79)	TCA GAC GTG TGC TCT TCC GAT CTT TTT TTT TTT TTT TTT T
Example 7	
Nimblegen 3' to 5' arrays with free 5' end Array probes	
	5' to 3'
Probe1 (SEQ ID NO: 80)	GCGTTCAGAGTGGCAGTCGAGATCACGCGGCAATCATATCGGACAGATCGGAAGAGCGTAGTGTAG
Probe2 (SEQ ID NO: 81)	GCGTTCAGAGTGGCAGTCGAGATCACAAGATGAACCCGGCTCATAGATCGGAAGAGCGTAGTGTAG
Probe3 (SEQ ID NO: 82)	GCGTTCAGAGTGGCAGTCGAGATCACTCCCAACAGAGTGCCTCAAGATCGGAAGAGCGTAGTGTAG
Probe4 (SEQ ID NO: 83)	GCGTTCAGAGTGGCAGTCGAGATCACCGAATGGCGACTAATCATAGATCGGAAGAGCGTAGTGTAG
Probe5 (SEQ ID NO: 84)	GCGTTCAGAGTGGCAGTCGAGATCACAAACATCTACCCTCAAGTAGATCGGAAGAGCGTAGTGTAG
Probe6 (SEQ ID NO: 85)	GCGTTCAGAGTGGCAGTCGAGATCACGATAACAGTATTGGCCATAGATCGGAAGAGCGTAGTGTAG
Probe7 (SEQ ID NO: 86)	GCGTTCAGAGTGGCAGTCGAGATCACGGTCGAATCAGGGTAGCGAGATCGGAAGAGCGTAGTGTAG
Probe8 (SEQ ID NO: 87)	GCGTTCAGAGTGGCAGTCGAGATCACACTACGGCGAAAGTGGGCAGATCGGAAGAGCGTAGTGTAG
Probe9 (SEQ ID NO: 88)	GCGTTCAGAGTGGCAGTCGAGATCACATCTTGCTCAAAGTTGCTAGATCGGAAGAGCGTAGTGTAG
Probe10 (SEQ ID NO: 89)	GCGTTCAGAGTGGCAGTCGAGATCACCCGGAATTCCGAATTGGCAGATCGGAAGAGCGTAGTGTAG
Probe11 (SEQ ID NO: 90)	GCGTTCAGAGTGGCAGTCGAGATCACATGTATTACCTTGGGCGAAGATCGGAAGAGCGTAGTGTAG
Probe12 (SEQ ID NO: 91)	GCGTTCAGAGTGGCAGTCGAGATCACCTCGAATAGGAAATGCGAAGATCGGAAGAGCGTAGTGTAG
Probe13 (SEQ ID NO: 92)	GCGTTCAGAGTGGCAGTCGAGATCACGGCGGTTAGATTTAGCAAAGATCGGAAGAGCGTAGTGTAG
Probe14 (SEQ ID NO: 93)	GCGTTCAGAGTGGCAGTCGAGATCACCCATCAGAATTTAATTCCAGATCGGAAGAGCGTAGTGTAG

Probe15 (SEQ ID NO: 94)	GCGTTCAGAGTGGCAGTCGAGATCACCTTAGCACCTATGTAATGAGATCGGAAGAGCGTAGTGTAG
Probe16 (SEQ ID NO: 95)	GCGTTCAGAGTGGCAGTCGAGATCACGTGCGAGCGCAAGTCAATAGATCGGAAGAGCGTAGTGTAG
Probe17 (SEQ ID NO: 96)	GCGTTCAGAGTGGCAGTCGAGATCACGAACTTGGGAGATACTATAGATCGGAAGAGCGTAGTGTAG
Probe18 (SEQ ID NO: 97)	GCGTTCAGAGTGGCAGTCGAGATCACTGCGGATTACAGGCGCACAGATCGGAAGAGCGTAGTGTAG
Probe19 (SEQ ID NO: 98)	GCGTTCAGAGTGGCAGTCGAGATCACCTACCTAAAGAGTGGCGCAGATCGGAAGAGCGTAGTGTAG
Probe20 (SEQ ID NO: 99)	GCGTTCAGAGTGGCAGTCGAGATCACAAGCCAATCACTTGCATAAGATCGGAAGAGCGTAGTGTAG
Probe21 (SEQ ID NO: 100)	GCGTTCAGAGTGGCAGTCGAGATCACCGTATAAACGTGGCTTGGAGATCGGAAGAGCGTAGTGTAG
Probe22 (SEQ ID NO: 101)	GCGTTCAGAGTGGCAGTCGAGATCACGTTATACAGCAATCAGGTAGATCGGAAGAGCGTAGTGTAG
Probe23 (SEQ ID NO: 102)	GCGTTCAGAGTGGCAGTCGAGATCACTAGAGGATAGATGCGCTGAGATCGGAAGAGCGTAGTGTAG
Probe24 (SEQ ID NO: 103)	GCGTTCAGAGTGGCAGTCGAGATCACACTAGTCCTACGCGTGGAAGATCGGAAGAGCGTAGTGTAG
Probe25 (SEQ ID NO: 104)	GCGTTCAGAGTGGCAGTCGAGATCACGCGCTACATACTTAGTCGAGATCGGAAGAGCGTAGTGTAG
Frame probe	
Layout1 (SEQ ID NO: 105)	AAATTTCGTCTGCTATCGCGCTTCTGTACC
Capture probe	
LP_Poly-dTVN (SEQ ID NO: 106)	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTTTTVN
Amplification handle probe	
A-handle (SEQ ID NO: 107)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Second strand synthesis and first PCR amplification handles	
A_primer (SEQ ID NO: 108)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
B_dt20VN_primer (SEQ ID NO: 109)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTVN
Second PCR	
A_primer (SEQ ID NO: 110)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
B_primer (SEQ ID NO: 111)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Example 9	
Template switching	
Templateswitch_longB (SEQ ID NO: 112)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATrGrGrG

Example 12

The following experiment demonstrates that the step of labelling the secured (captured) cDNA molecules can be performed using arrays that comprise a high density of capture probes comprising unique positional domains.

Array preparation

Pre-fabricated high-density microarray chips were ordered from Roche-Nimblegen (Madison, WI, USA). Each chip contained multiple probe arrays, each with 270,000 features of which 135,000 features carried a capture probe comprising a unique ID-tag sequence (positional domain) and a capture region (capture domain). Each feature was 13x13 µm in size. The capture probes were composed 5' to 3' of: a universal domain containing five dUTP bases (a cleavage domain) and a general 5' amplification domain; an ID tag (positional domain); and a capture region (capture domain) (see Table 5). Each array was also fitted with a frame of marker probes carrying a generic 30bp sequence (Table 5) to enable hybridization of fluorescent probes to help with orientation during array visualization.

The probe-arrays were separated during reaction steps by a 16-pad mask (Arrayit Corporation, Sunnyvale, CA, USA).

The method described in Example 10 was performed up to and including the step of tissue removal.

Imaging

A solution with frame marker probe (Table 5) at a concentration of 170 nM in PBS was prepared. This solution was added to the wells and the slide was incubated at room temperature for 5 minutes, followed by brief washing in PBS and spin drying.

The high-density microarray glass chip was imaged at 532 nm using an Agilent microarray scanner at 100% exposure and 5µm resolution (Figure 15).

Table 5 Nimblegen 5' to 3' arrays with free 3' end Array probes	
	5' to 3'
Probe1 (SEQ ID NO: 113)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCCGATATGATTGCCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe2 (SEQ ID NO: 114)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGAGCCGGGTTCATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe3 (SEQ ID NO: 115)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTGAGGCACTCTGTTGGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe4 (SEQ ID NO: 116)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGATTAGTCGCCATTCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe5 (SEQ ID NO: 117)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTGAGGGTAGATGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe6 (SEQ ID NO: 118)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGGCCAATACTGTTATCTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe7 (SEQ ID NO: 119)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTACCCTGATTCGACCTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe8 (SEQ ID NO: 120)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCCACTTTCGCCGTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe9 (SEQ ID NO: 121)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCAACTTTGAGCAAGATTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe10 (SEQ ID NO: 122)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCAATTCGGAATTCCGGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe11 (SEQ ID NO: 123)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCCCAAGGTAATACATTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe12 (SEQ ID NO: 124)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCATTTCCTATTCGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe13 (SEQ ID NO: 125)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGCTAAATCTAACCGCCTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe14 (SEQ ID NO: 126)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAATTAAATTCTGATGGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe15 (SEQ ID NO: 127)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCATTACATAGGTGCTAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe16 (SEQ ID NO: 128)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGACTTGCGCTCGCACTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe17 (SEQ ID NO: 129)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGTATCTCCCAAGTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe18 (SEQ ID NO: 130)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGCGCCTGTAATCCGCATTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe19 (SEQ ID NO: 131)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGCCACTCTTTAGGTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe20 (SEQ ID NO: 132)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTATGCAAGTGATTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe21 (SEQ ID NO: 133)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAAGCCACGTTTATACGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe22 (SEQ ID NO: 134)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTACCTGATTGCTGTATAACTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe23 (SEQ ID NO: 135)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGCGCATCTATCCTCTATTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe24 (SEQ ID NO: 136)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCACGCGGTAGGACTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe25 (SEQ ID NO: 137)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCGACTAAGTATGTAGCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTVN
Frame probe (SEQ ID NO: 138)	AAATTTCGTCTGCTATCGCGCTTCTGTACC
Fluorescent marker probe (SEQ ID NO: 139)	GGTACAGAAGCGCGATAGCAG – Cy3

Example 13

The following experiment demonstrates that the step of labelling the secured (captured) cDNA molecules is effective at lower concentrations of fluorescently labelled nucleotides.

The method described in Example 10 was performed up to the cDNA synthesis step, which was replaced by the step described below.

cDNA synthesis with Cy3-dNTP

For each well an 80µl cDNA synthesis mixture was prepared containing 4µl each of dATP/dGTP/dTTP (10mM), 4 ul DTT (0.1M), 1x BSA, 20U/µl Superscript III, 5U/ul RNaseOUT, 1x first strand buffer (Superscript III, Invitrogen) and MilliQ water. The concentration of dCTP and Cy3-dCTP was varied so that in five parallel experiments 1/1, 1/2, 1/4, 1/8 and 1/16 of the original concentration of 2.5mM for dCTP and 1mM for Cy3-dCTP was used. 4µl of dCTP and 4µl of Cy3-dCTP with these concentrations were pipetted into each respective synthesis mixture. 70µl of the reaction mixture was added to each well. The reactions were covered with a plastic sealer and incubated at 37°C overnight.

The washing, tissue removal and imaging steps were performed according to Example 10.

Signal quantitation

The signal intensities resulting from imaging of the codelink glass chip in the Agilent microarray scanner were analyzed using the Genepix pro software. Average signal intensities were calculated from multiple selected areas within each Cy3 footprint and plotted in a diagram (Figure 16).

Example 14

The following experiment demonstrates that a variety of tissues can be used in the methods of the invention.

The method described in Example 10 was performed up to the Imaging step. The tissue preparation, fixation and imaging steps were replaced by the steps described below.

Preparation of fresh frozen tissue and sectioning onto capture probe arrays

Fresh non-fixed drosophila or zebrafish tissue was frozen with dry ice and subsequently mounted for sectioning with a cryostat at 10µm. A slice of tissue was applied onto each probe array to be used.

Fixation of tissue section using methanol

The sections were fixed in a pre-chilled methanol bath at -20°C for 10 min. After fixation, the slide was briefly washed for 10 seconds 1xPBS and heat dried at 50°C for 15 min in an eppendorf thermomixer.

Imaging

The probe array chip with zebrafish (Figure 17A) or drosophila (Figure 17B) tissue was imaged before removal using phase-contrast imaging. Imaging after removal of tissue use the Cy3 compatible channel on a MMI Cellcut instrument mounted on an Olympus IX 81 microscope.

Example 15

The following experiment demonstrates further that a variety of tissues can be used in the methods of the invention.

The method described in Example 10 was performed up to the Imaging step. The tissue preparation, fixation, permeabilization, tissue removal and imaging steps were replaced by the steps described below.

Preparation of fresh frozen tissue and sectioning onto capture probe arrays

Fresh prostate cancer tissue was trimmed if necessary, embedded in Neg-50 (Thermo Scientific) and snap frozen in liquid nitrogen. The tissue was subsequently mounted for sectioning with a cryostat at 10µm. A slice of tissue was applied onto each probe array to be used.

Fixation of tissue section using formalin

The chip was attached to an ArrayIt 16-well mask and holder. 70µl of 4% paraformaldehyde dissolved in PBS was added to the probe array well to cover the tissue section. The reaction was incubated at room temperature for 10 minutes. The mask was removed and the chip was washed for 10 seconds in PBS. The chip was then incubated at 50°C for 15 minutes.

Permeabilization using pepsin and HCl

Pepsin was diluted to 0.1% in 0.1M HCl and was preheated to 37°C. The chip was attached to an ArrayIt 16-well mask and holder. 70µl of the pepsin/HCl mixture was added to each well. The reaction was incubated for 10 minutes at 37°C. The wells were washed with 100µl 0.1x SSC by pipetting.

Tissue removal

The array was attached to an ArrayIt slide holder and 16 well mask (ArrayIt Corporation). 20 µl Proteinase K Solution (Qiagen) was added for each 150 µl Proteinase K Digest Buffer from the RNeasy FFPE kit (Qiagen). 50 µl of the final mixture was added to each well and the array was incubated at 56°C for 1 hour. The array was washed as described in Example 10.

Imaging

The probe array chip with prostate tissue was imaged before removal using phase-contrast imaging. Imaging after removal of tissue use the Cy3 compatible channel on a MMI Cellcut instrument mounted on an Olympus IX 81 microscope (Figure 17C).

Example 16

The following experiment demonstrates that a cell sample, i.e. a suspension of cells, can be used as the tissue sample in the methods of the invention.

The method described in Example 10 was performed up to the Imaging step. The tissue preparation, fixation, permeabilization, cDNA synthesis, tissue removal and imaging steps were replaced by the steps described below.

Tissue preparation: Application of cells onto capture probe arrays

Approximately 1000-2000 mouse fibroblast cells (cell line (NIH 373)) in a 5µl volume (in 0.1x SSC) were pipetted onto the probe array and distributed using the pipette tip. The chip was then incubated at 37°C for 5.5 minutes.

Fixation of cells using formalin

The chip was attached to an ArrayIt 16-well mask and holder. 100ul of 4% paraformaldehyde dissolved in PBS was added to the probe array well to cover the cells. The reaction was incubated at room temperature for 10 minutes. The probe array well was washed once with 100µl 0.1x SSC by pipetting. The chip was then incubated at 50°C for 15 minutes.

Permeabilization using pepsin and HCl

Pepsin was diluted to 0.1% in 0.1M HCl and was preheated to 37°C. The chip was attached to an ArrayIt 16-well mask and holder. 70µl of the pepsin/HCl mixture was added to each well. The reaction was incubated for 1 minute at 37°C. The wells were washed with 100µl 0.1x SSC by pipetting.

cDNA synthesis with Cy3-dNTP

For each well an 80µl cDNA synthesis mixture was prepared containing 4µl each of dATP/dGTP/dTTP (10mM), 2µl dCTP(2.5mM), 2µl Cy3-dCTP(1mM), 4µl DTT (0.1M), 1x BSA, 20U/µl Superscript III, 5U/µl RNaseOUT, 1x first strand buffer (Superscript III, Invitrogen) and MilliQ water. 70 µl of the reaction mixture was added to each well. The reactions were covered with a plastic sealer and incubated at 37°C overnight.

Imaging

The probe array chip with cells was imaged before removal using phase-contrast imaging. Imaging after removal of tissue use the Cy3 compatible channel on a MMI Cellcut instrument mounted on an Olympus IX 81 microscope (Figure 17D).

Claims

1. A method for localized or spatial detection and/or analysis of RNA in a tissue sample or a portion thereof, comprising:

(a) providing an object substrate on which at least one species of capture probe, comprising a capture domain, is directly or indirectly immobilized such that the probes are oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer;

(b) contacting said substrate with a tissue sample and allowing RNA of the tissue sample to hybridise to the capture probes;

(c) generating cDNA molecules from the captured RNA molecules using said capture probes as RT primers;

(d) labelling the cDNA molecules generated in step (c), wherein said labelling step may be contemporaneous with, or subsequent to, said generating step;

(e) detecting a signal from the labelled cDNA molecules; and optionally

(f) imaging the tissue sample, wherein the tissue sample is imaged before or after step (c).

2. The method of claim 1, being a method for determining the optimum conditions for localised or spatial detection of RNA in a tissue sample on an object substrate, comprising further steps:

(g) repeating steps (a)-(e), and optionally step (f), with a second object substrate, using different conditions in step (b);

(h) comparing the intensity and/or resolution of the signal from the labelled cDNA molecules immobilized on said first and second object substrate; and optionally

(i) selecting the conditions that provide the optimum signal intensity and/or resolution of the labelled cDNA molecules.

3. The method of claim 1, being a method for determining and/or analysing RNA or a transcriptome of a tissue sample or a portion thereof comprising further steps:

(g') removing the labelled cDNA from at least one portion of the surface of the object substrate;

(h') optionally amplifying the remaining cDNA molecules immobilized on the surface of the object substrate;

(i') releasing at least part of the remaining cDNA molecules and/or optionally their amplicons from the surface of the object substrate, wherein said released molecules may be a first strand and/or second strand cDNA molecule or an amplicon thereof;

(j') directly or indirectly analysing the sequence of the released molecules.

4. The method of claim 1, being a method for determining and/or analysing RNA or a transcriptome of a tissue sample or a portion thereof comprising:

(a") providing an object substrate (e.g. an array) on which multiple species of capture probes are directly or indirectly immobilized such that each species occupies a distinct position on the object substrate and is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein each species of said capture probe comprises a nucleic acid molecule with 5' to 3':

(i) a positional domain that corresponds to the position of the capture probe on the object substrate, and

(ii) a capture domain;

(b") contacting said object substrate with a tissue sample such that the position of a capture probe on the object substrate may be correlated with a position in the tissue sample and allowing RNA of the tissue sample to hybridise to the capture domain in said capture probes;

(c") generating cDNA molecules from the captured RNA molecules using said capture probes as RT primers,

(d") labelling the cDNA molecules generated in step (c'), wherein said labelling step may be contemporaneous with, or subsequent to, said generating step;

(e") detecting a signal from the labelled cDNA molecules;

(f") optionally imaging the tissue sample, wherein the tissue sample is imaged before or after step (c").

(g") optionally removing the labelled cDNA from at least one portion of the surface of the object substrate;

(h") optionally amplifying the cDNA molecules immobilized on the surface of the object substrate;

(i") releasing at least part of the cDNA molecules and/or optionally their amplicons from the surface of the object substrate, wherein said released molecules may be a first strand and/or second strand cDNA molecule or an amplicon thereof and wherein said part includes the positional domain or a complement thereof;

(j") directly or indirectly analysing the sequence of the released molecules.

5. The method of claim 3 or 4, further comprising step (k) correlating said sequence analysis information with an image of said tissue sample, wherein the tissue sample is imaged before or after step (c).

6. The method of any one of claims 1 to 5, further comprising a step of correlating the signal detected from the labelled cDNA molecules with an image of said tissue sample, wherein the tissue sample is imaged before or after step (c).

7. The method of any one of claims 1 to 6, wherein the label is incorporated into the synthesized part of the cDNA molecules.

8. The method of claim 7, wherein the label is conjugated to a nucleotide and the step of labelling comprises the incorporation of labelled nucleotides into the synthesized cDNA molecule, preferably wherein the labelled nucleotides are fluorescently labelled nucleotides.

9. The method of any one of claims 1 to 8, wherein the step of detecting a signal from the labelled cDNA molecules comprises imaging the substrate such that the signal from the labelled cDNA molecules is detected, preferably wherein the substrate is imaged using light, bright field, dark field, phase contrast, fluorescence, reflection, interference or confocal microscopy or a combination thereof.

10. The method of claim 9, being a method for the identification of transcriptionally active tumour cells, wherein the tissue sample is a cell suspension comprising tumour cells and the image of the labelled cDNA corresponds to transcriptionally active cells.

11. The method of any one of claims 1 to 10, wherein the capture probes are DNA molecules.

12. The method of any one of claim 3 to 11, wherein the at least one portion of labelled cDNA molecules is removed from of the surface of the object substrate by laser ablation.

13. The method of any one of claims 1 to 13, wherein the capture probes further comprise a positional domain which is 5' relative to the capture domain, wherein said positional domain comprises a sequence that corresponds to the position of the capture probe on the object substrate, preferably wherein the positional domain of each species of capture probe comprises a unique barcode sequence.

14. The method of any one of claims 1 to 13, wherein the capture probes further comprise a universal domain which is 5' relative to the capture domain or, if present the positional domain, wherein said universal domain comprises:

(i) an amplification domain, for amplifying the generated DNA molecules; and/or

(ii) a cleavage domain for releasing the generated DNA molecules from the surface of the object substrate.

15. The method of any one of claims 1 to 14, wherein the capture domain comprises a poly-T or poly-U DNA oligonucleotide comprising at least 10 deoxythymidine and/or deoxyuridine residues and/or a random or degenerate oligonucleotide sequence.

16. The method of any one of claims 1 to 15, wherein the capture probes are directly immobilized on the object substrate surface by their 5' end.

17. The method of any one of claims 1 to 15, wherein the capture probes are indirectly immobilized on the object substrate surface by hybridization to a surface probe, wherein the capture domain of the capture probes comprises an upstream sequence that is capable of hybridizing to 5' end of surface probes that are immobilized on the object substrate.

18. The method of claim 17, wherein the surface probes are immobilized to the object substrate surface by their 3' end.

19. The method of claim 17 or 18, wherein the surface probes comprise a sequence that is complementary to:

(i) at least part of the capture domain; and

(ii) at least part of the universal amplification domain.

20. The method of claim 19, wherein the surface probes further comprise a sequence that is complementary to the positional domain.

21. The method of any one of claims 4 to 20, wherein the molecules are released from the surface of the object substrate by:

(i) nucleic acid cleavage;

(ii) denaturation; and/or

(iii) physical means.

22. The method of claim 21(i), wherein the molecules are released by enzymatic cleavage of a cleavage domain, which is located in the universal domain or positional domain of the capture probe.

23. The method of claim 21(ii) or (iii), wherein the molecules are released by applying hot water or buffer to the object substrate.

24. The method of any one of claims 1 to 23, wherein the object substrate is composed of a material selected from the group consisting of glass, silicon, poly-L-lysine coated material, nitrocellulose, polystyrene, cyclic olefin copolymers (COCs), cyclic olefin polymers (COPs), polypropylene, polyethylene and polycarbonate.

25. The method of any one of claims 1 to 36, wherein the tissue sample is a tissue section or a cell suspension.

26. The method of any one of claims 4 to 25, further comprising a step of washing the object substrate to remove residual tissue.

27. The method of any one of claims 1 to 26, wherein the object substrate comprises at least one positional marker to enable orientation of the tissue sample on the object substrate, preferably wherein the positional marker is capable of hybridizing to a labelled, preferably fluorescently labelled, marker nucleic acid molecule.

28. The method of any one of claims 1 to 27, wherein the tissue sample is imaged using light, bright field, dark field, phase contrast, fluorescence, reflection, interference or confocal microscopy or a combination thereof, preferably using fluorescence microscopy.

29. The method of any one of claims 1 to 28, comprising a step of modifying the tissue sample prior to the step of contacting the tissue sample with the substrate and/or prior to the step of generating the cDNA molecules on the substrate.

30. The method of claim 29, wherein the step of modifying the tissue sample comprises dissecting the tissue sample, preferably wherein the tissue sample is dissected using laser capture microdissection (LCM).

31. A method for making or producing an object substrate comprising a substrate on which one or more species of capture probe, comprising a capture domain, is directly or indirectly immobilized such that each probe is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer, wherein the probes are immobilised on the object substrate with a homogeneous distribution and said substrate is for use in:

(i) capturing RNA from a tissue sample that is contacted with said object substrate; or

(ii) localised or spatial detection of RNA in a tissue sample,

said method comprising:

(a) immobilizing directly or indirectly multiple surface probes to a substrate, wherein the surface probes comprise:

(i) a domain capable of hybridizing to part of a capture domain oligonucleotide; and

(ii) a domain that is complementary to a universal domain oligonucleotide;

(b) hybridizing to the surface probes immobilized on the substrate, capture domain oligonucleotides and universal domain oligonucleotides;

(c) ligating the universal domain oligonucleotides to the capture domain oligonucleotides to produce the capture probes.

32. An object substrate for use in the localised or spatial detection of RNA in a tissue sample comprising a planar substrate on which one or more species of capture probe, comprising a capture domain, is directly or indirectly immobilized such that each probe is oriented to have a free 3' end to enable said probe to function as a reverse transcriptase (RT) primer wherein the probes are immobilised on the object substrate with a homogeneous distribution and wherein the capture probe is selected from an oligonucleotide comprising a poly-T, poly-U and/or random oligonucleotide sequence.

33. The method of 31 or the object substrate of claim 32, wherein the capture probes are as defined in any one of claims 11, 14, 15 or 17, the surface probes are as defined in claims 18 or 19, the object substrate is as defined in claim 24 or 27 and/or the tissue sample is as defined in claim 25.

Figure 1

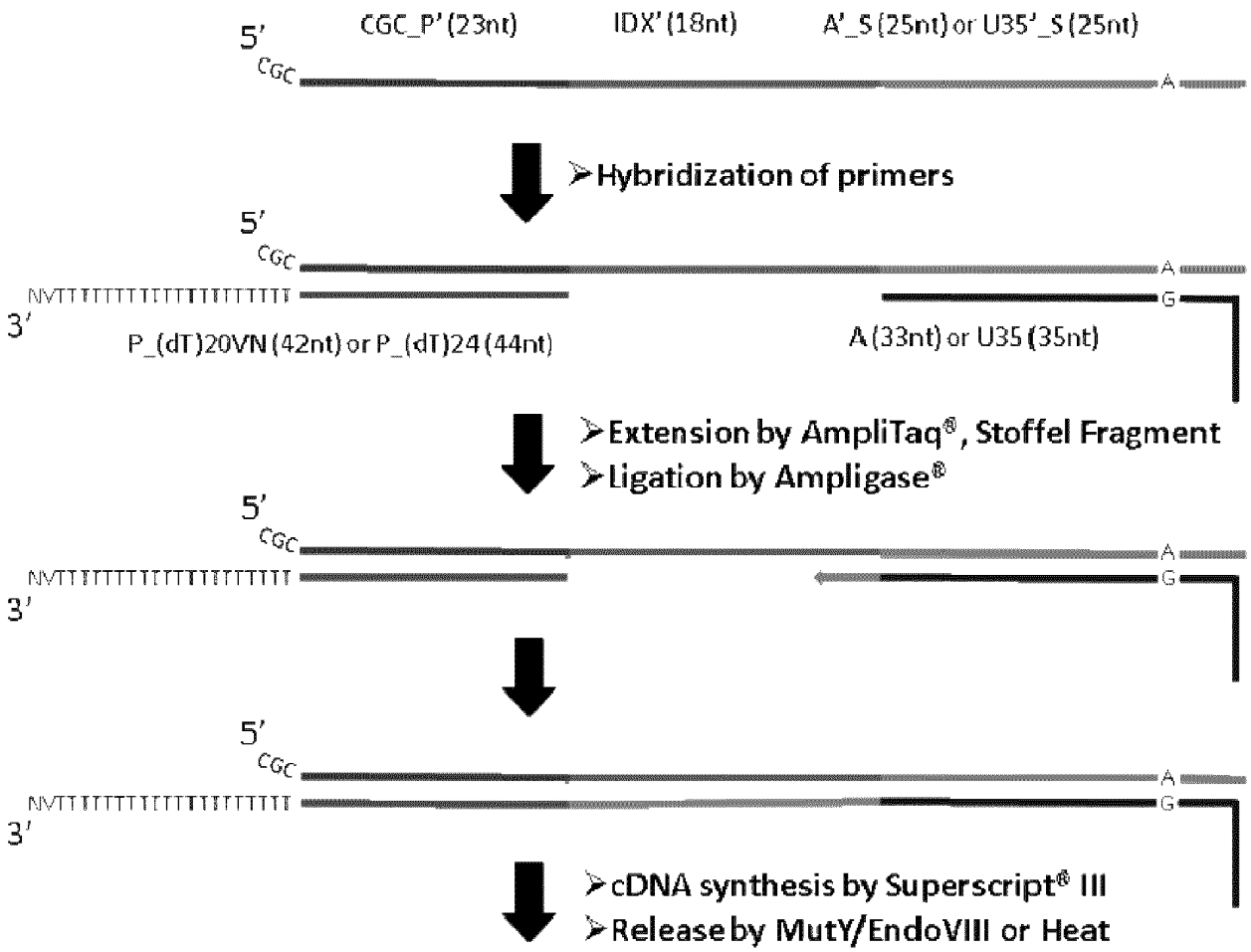

5'
CGC_P' (23nt)
IDX' (18nt)
A'_S (25nt) or U35'_S (25nt)
CGC
A
Hybridization of primers
5'
CGC
A
NVTTTTTTTTTTTTTTTTTTTT
G
3'
P_(dT)20VN (42nt) or P_(dT)24 (44nt)
A (33nt) or U35 (35nt)
Extension by AmpliTaq®, Stoffel Fragment
Ligation by Ampligase®
5'
CGC
A
NVTTTTTTTTTTTTTTTTTTTT
G
3'
5'
CGC
A
NVTTTTTTTTTTTTTTTTTTTT
G
3'
cDNA synthesis by Superscript® III
Release by MutY/EndoVIII or Heat

Figure 2

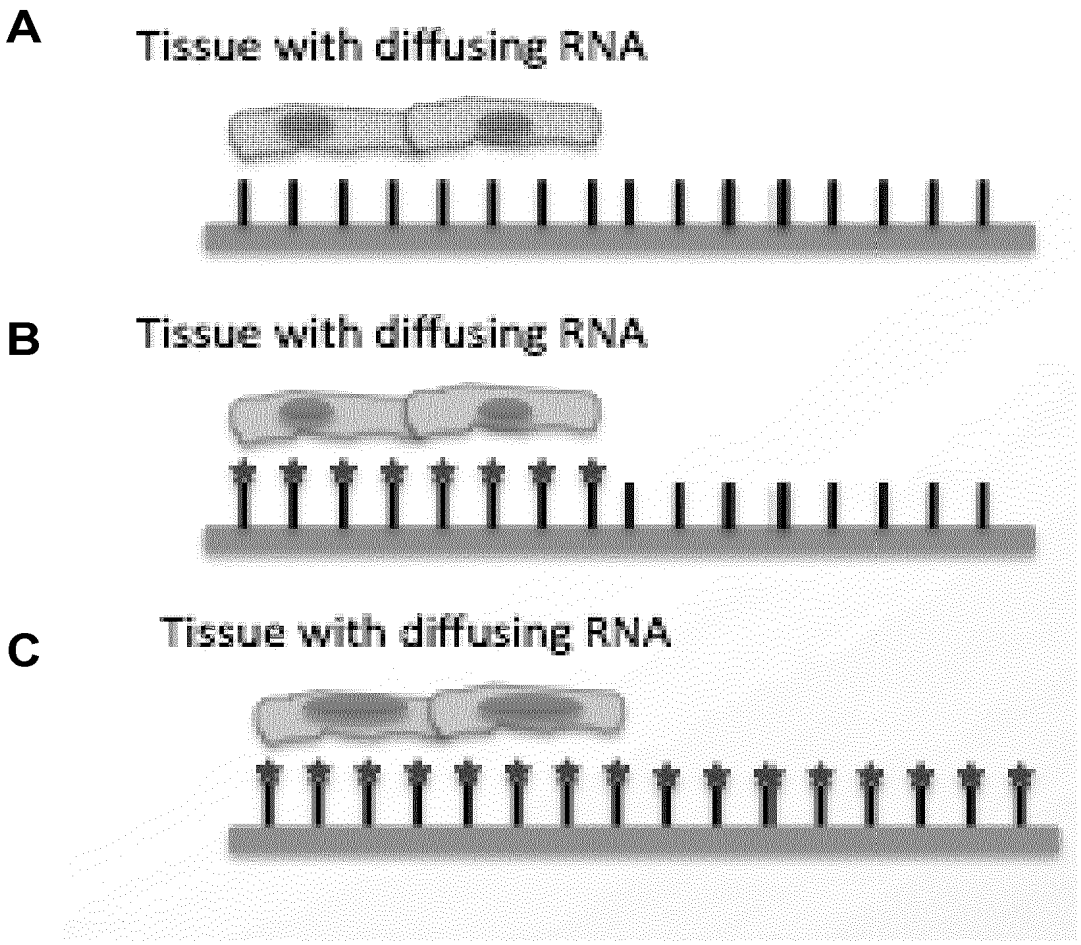

A
Tissue with diffusing RNA
B
Tissue with diffusing RNA
C
Tissue with diffusing RNA

Figure 3

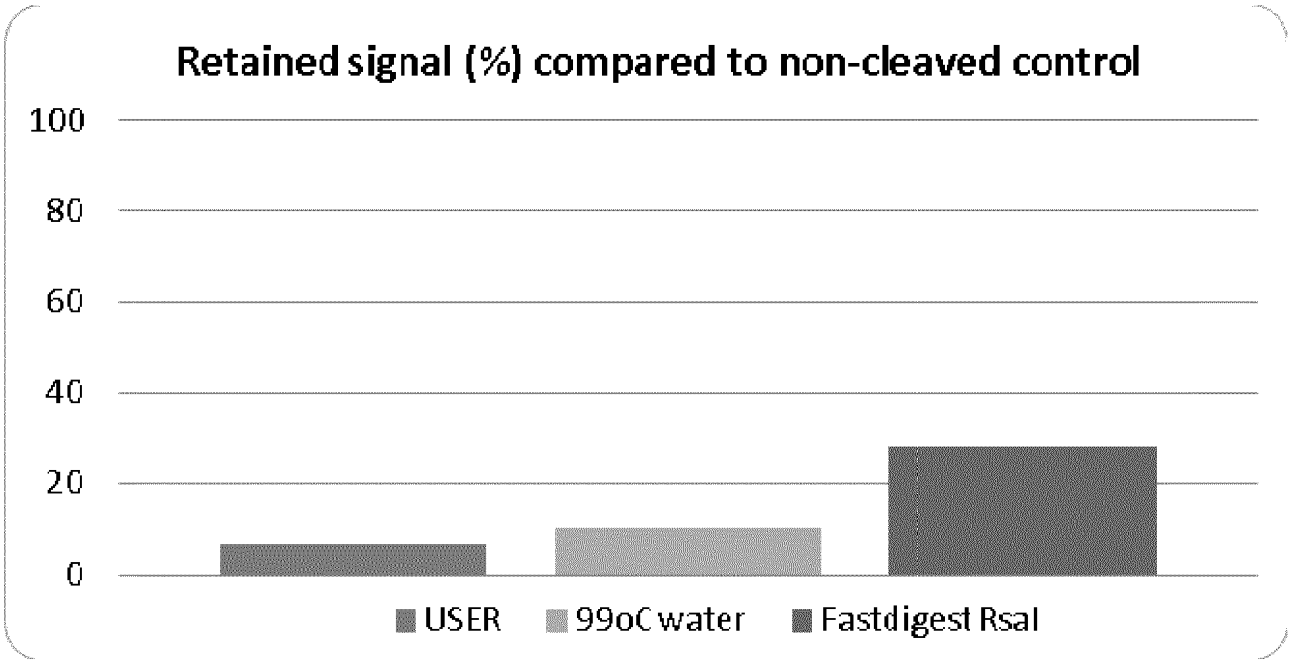

Retained signal (%) compared to non-cleaved control
100
80
60
40
20
0
USER
99oC water
Fastdigest Rsal

Figure 4

Lane 7 blast results	
E-value	1,00E-06
Barcode	Hits
Tag1	2165
Tag2	1223
Tag3	1338
Tag4	3605
Tag5	763
Tag6	1083
Tag7	1175
Tag8	1350
Tag9	616
Tag10	710
Tag11	1602
Tag12	1359
Tag13	1404
Tag14	1124
Tag15	1077
Tag16	1238
Tag17	1122
Tag18	1547
Tag19	1401
Tag20	1354

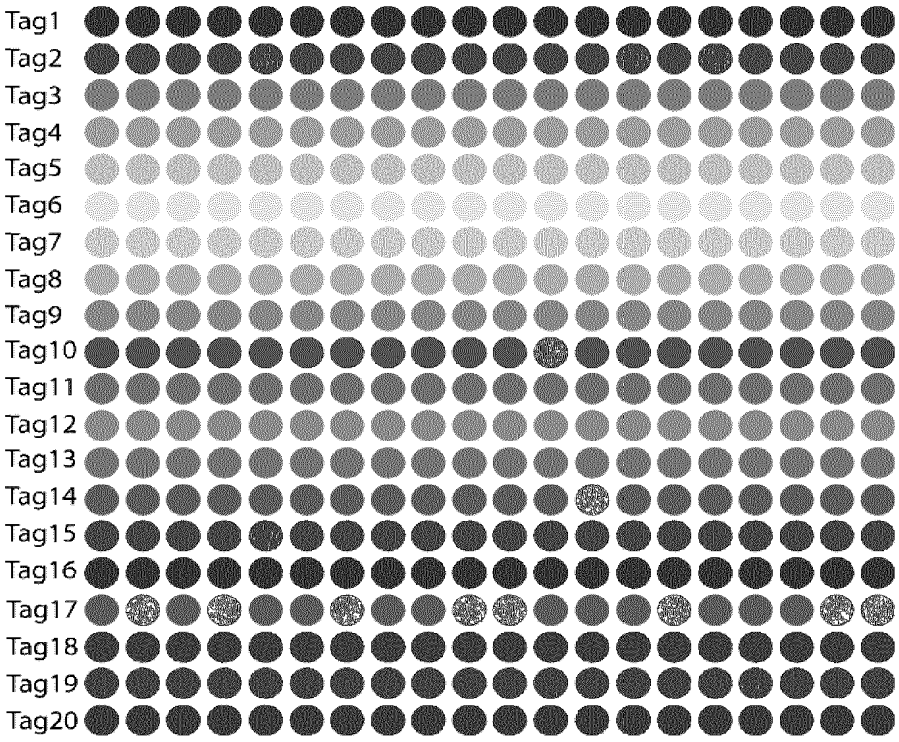


Figure 5

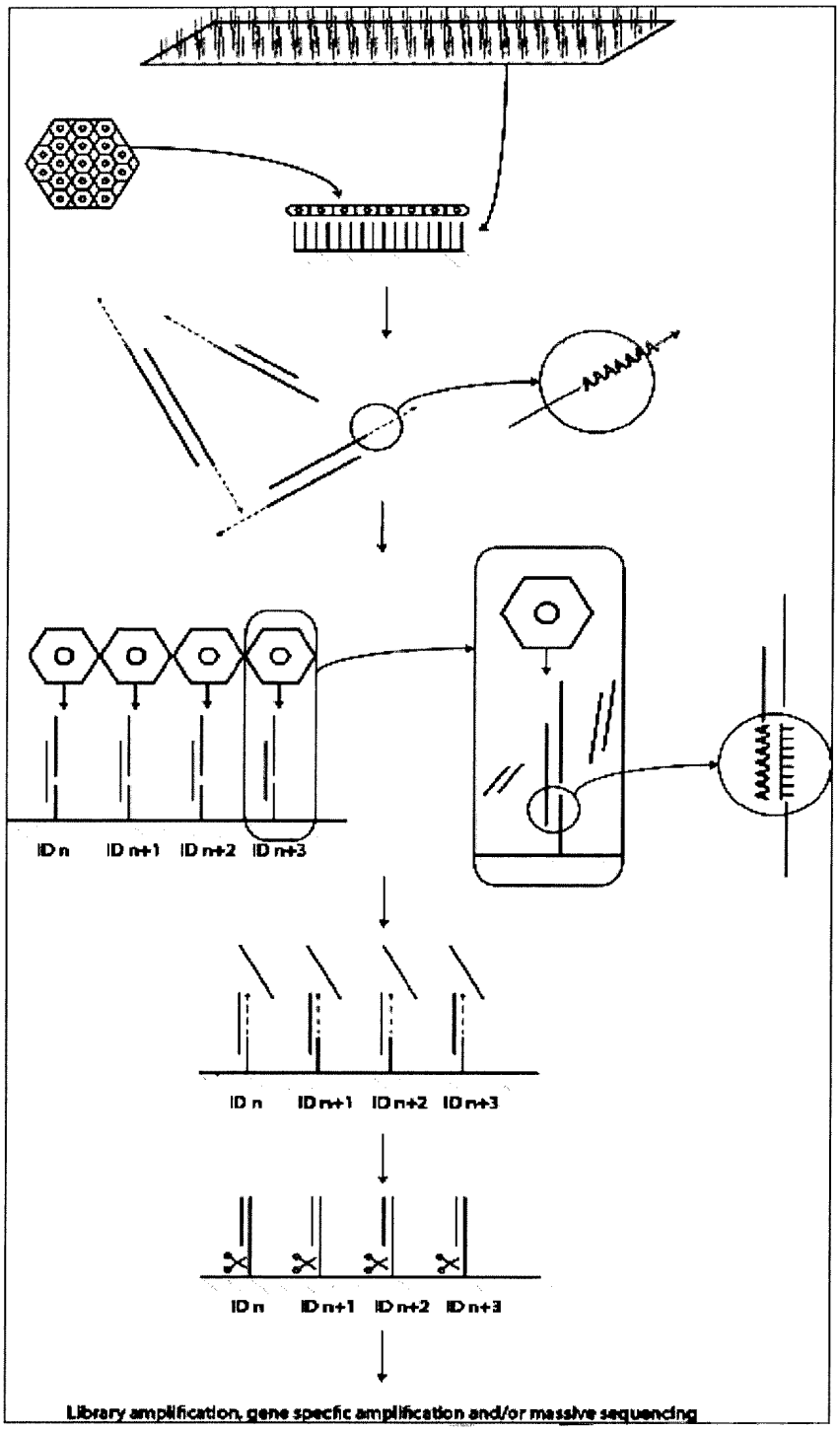

ID n
ID n+1
ID n+2
ID n+3
ID n
ID n+1
ID n+2
ID n+3
ID n
ID n+1
ID n+2
ID n+3
Library amplification, gene specific amplification and/or massive sequencing

Figure 6

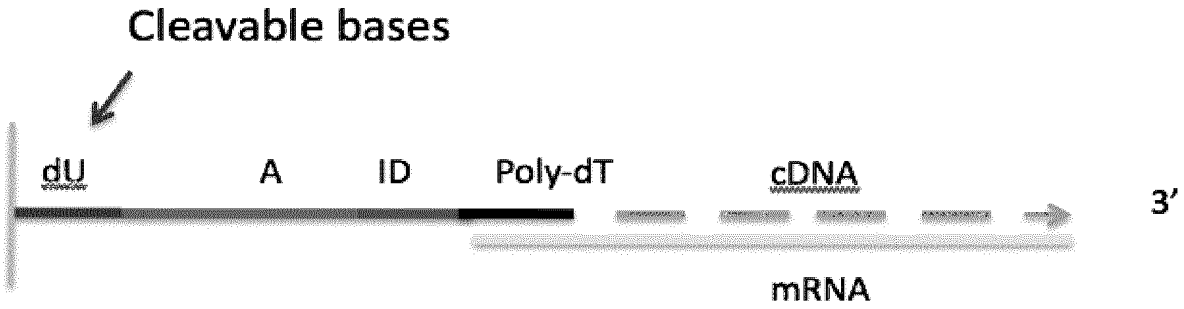


Figure 7

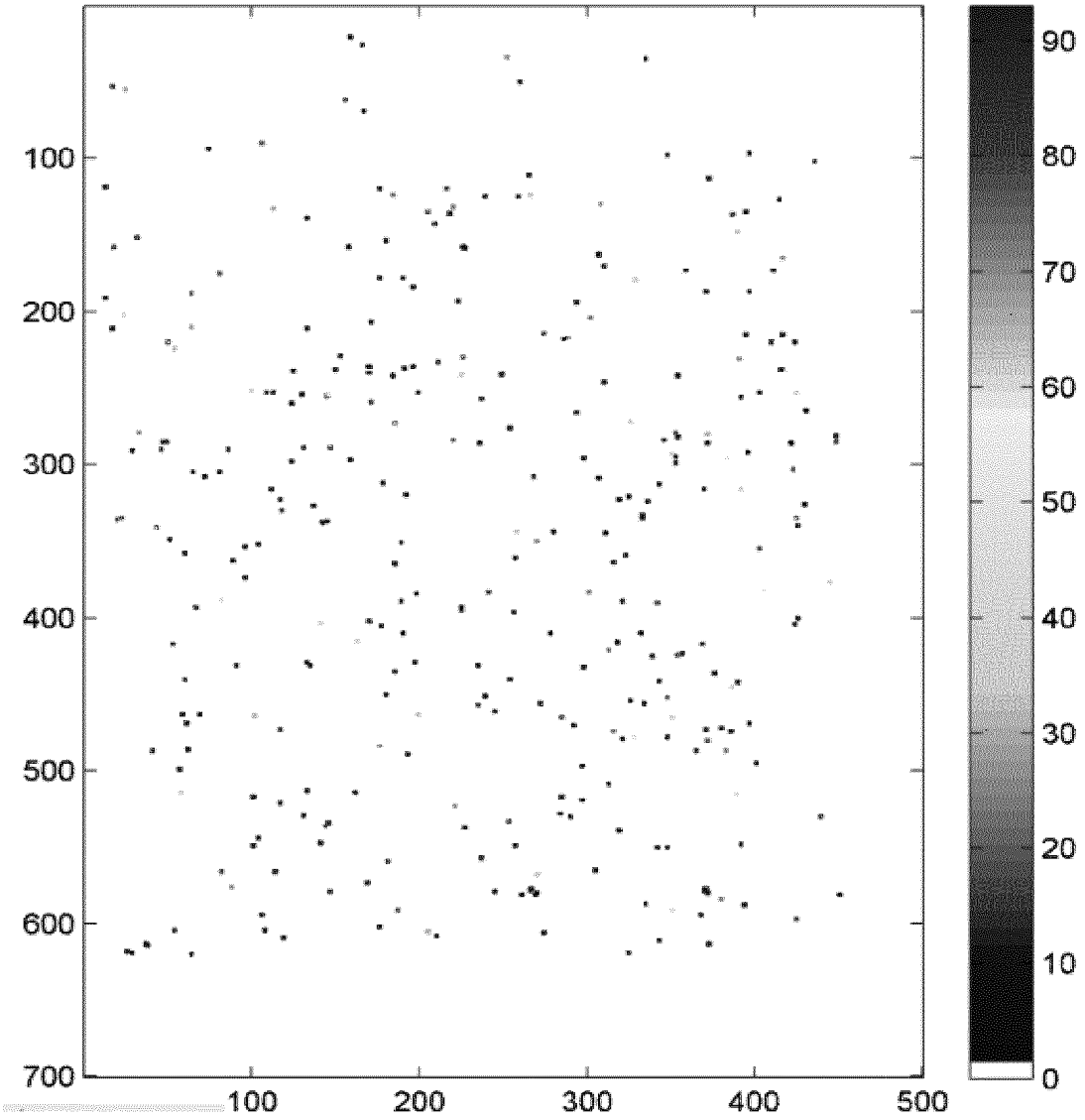

100
200
300
400
500
600
700
100
200
300
400
500
90
80
70
60
50
40
30
20
10
0

Figure 8

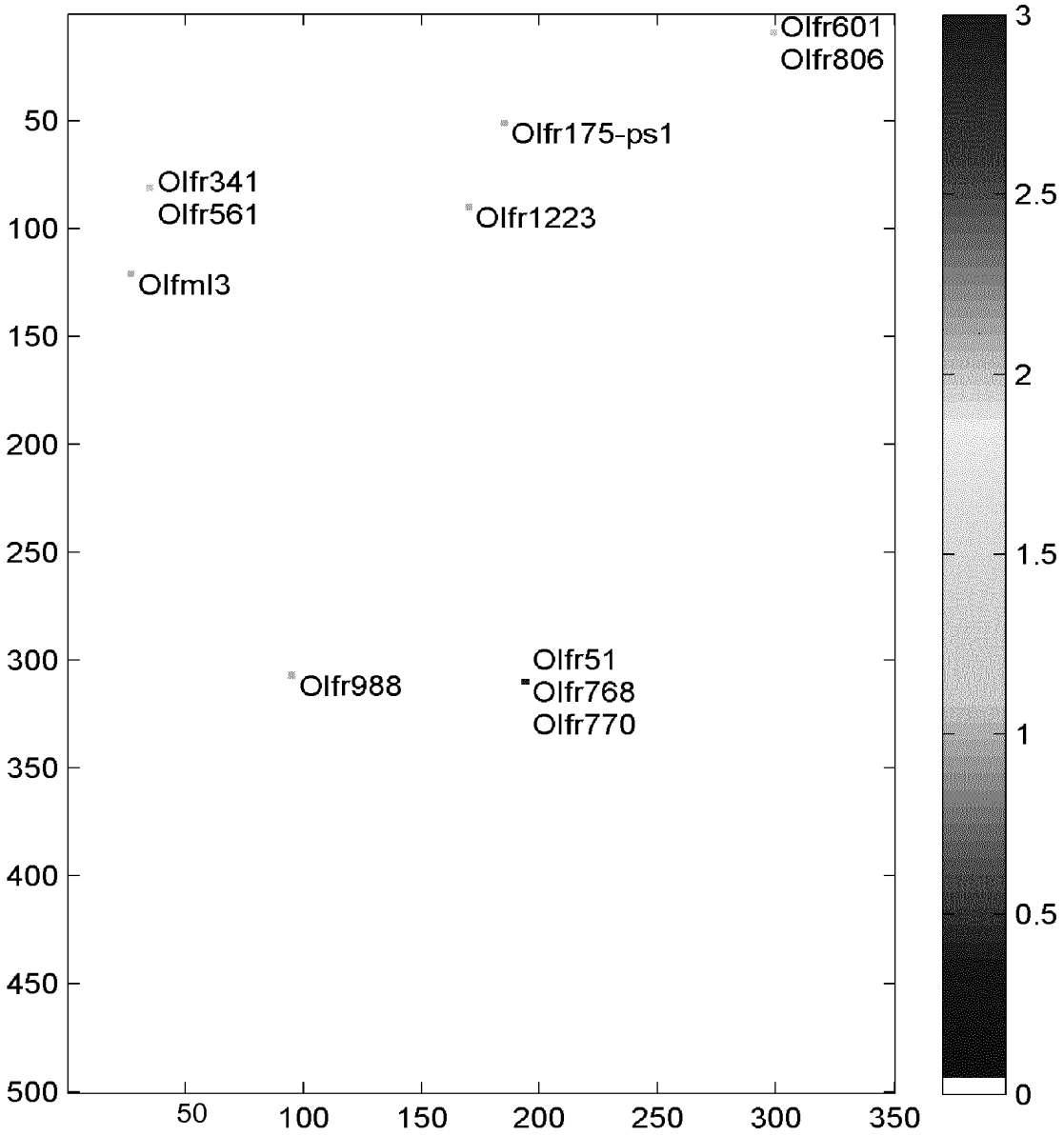

Olfr601
Olfr806
Olfr175-ps1
Olfr341
Olfr561
Olfr1223
Olfml3
Olfr988
Olfr51
Olfr768
Olfr770
50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
3
2.5
2
1.5
1
0.5
0

Figure 9

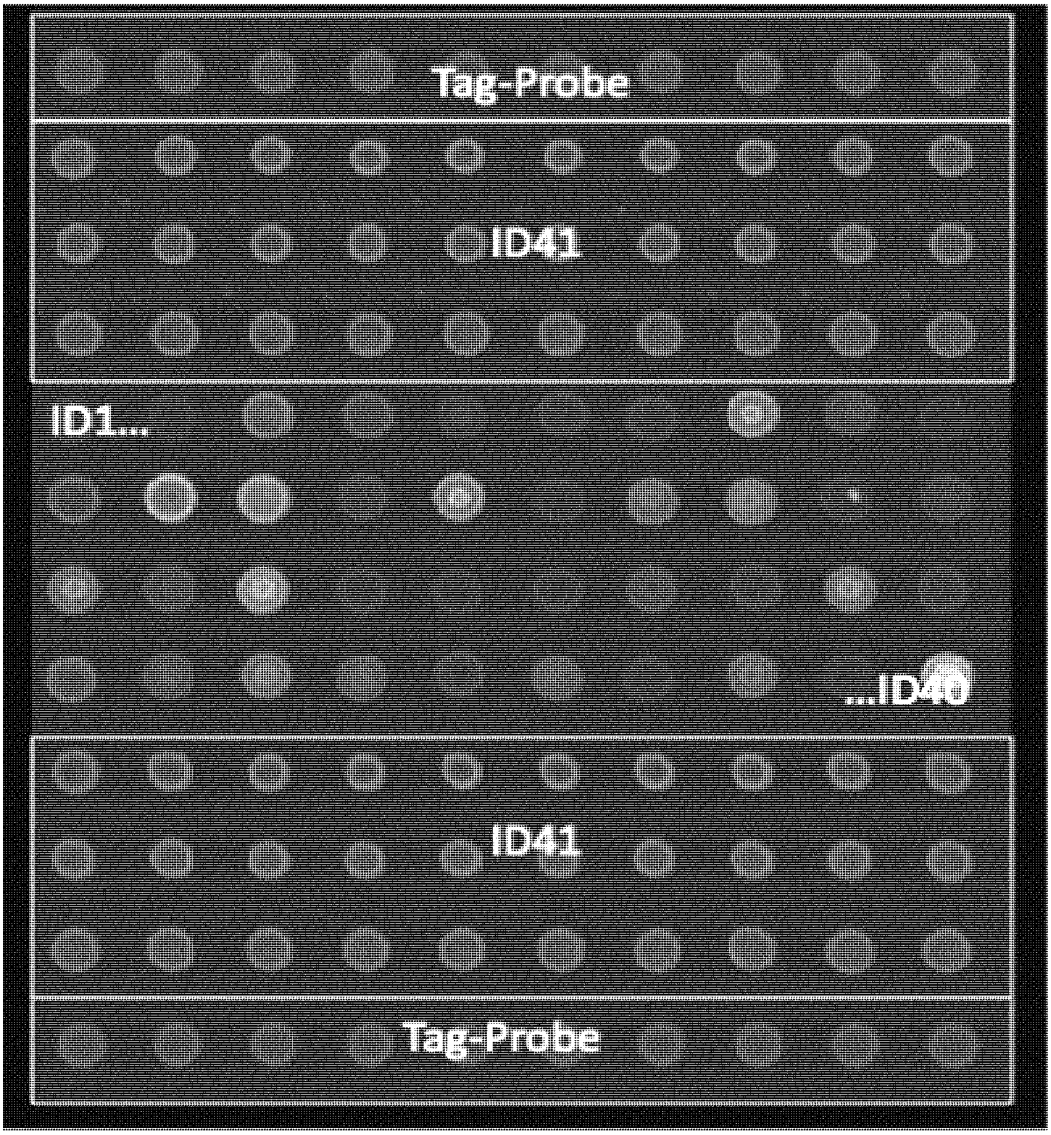


Figure 10

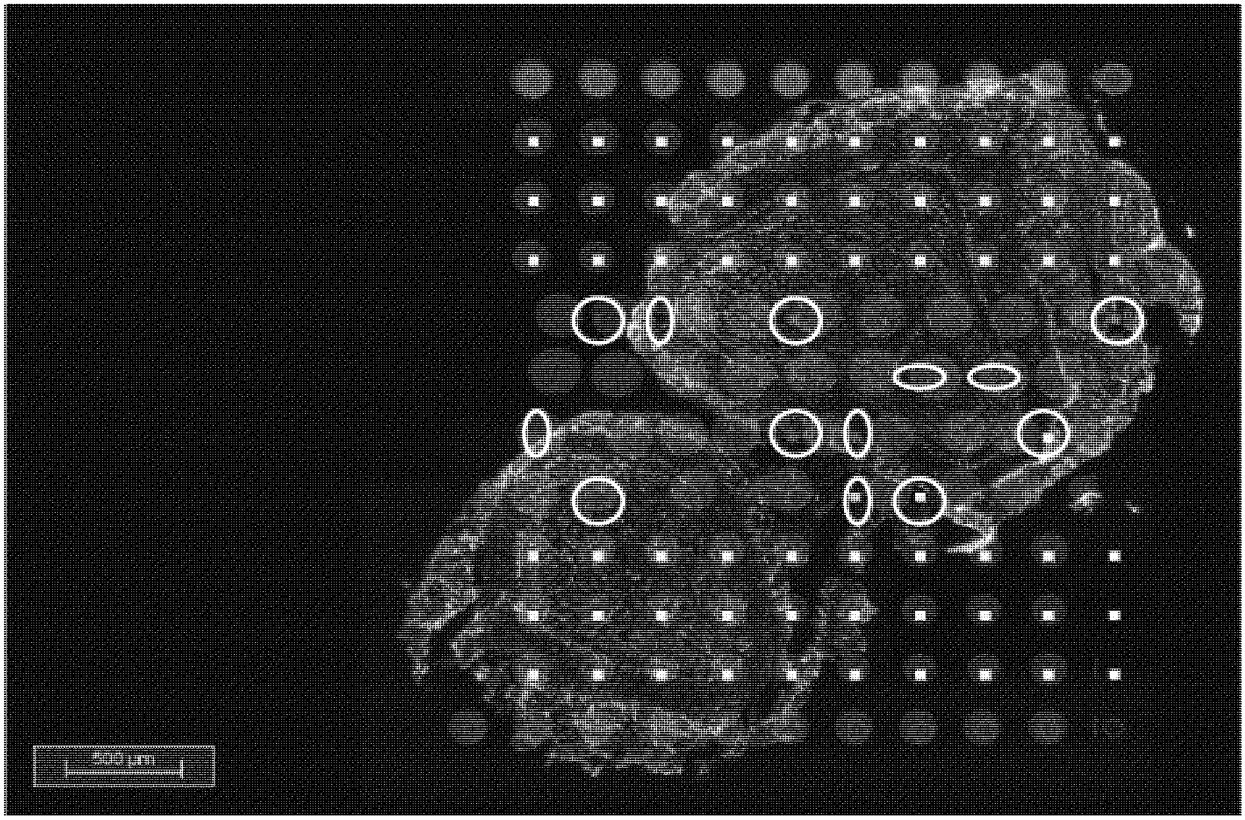


Figure 11

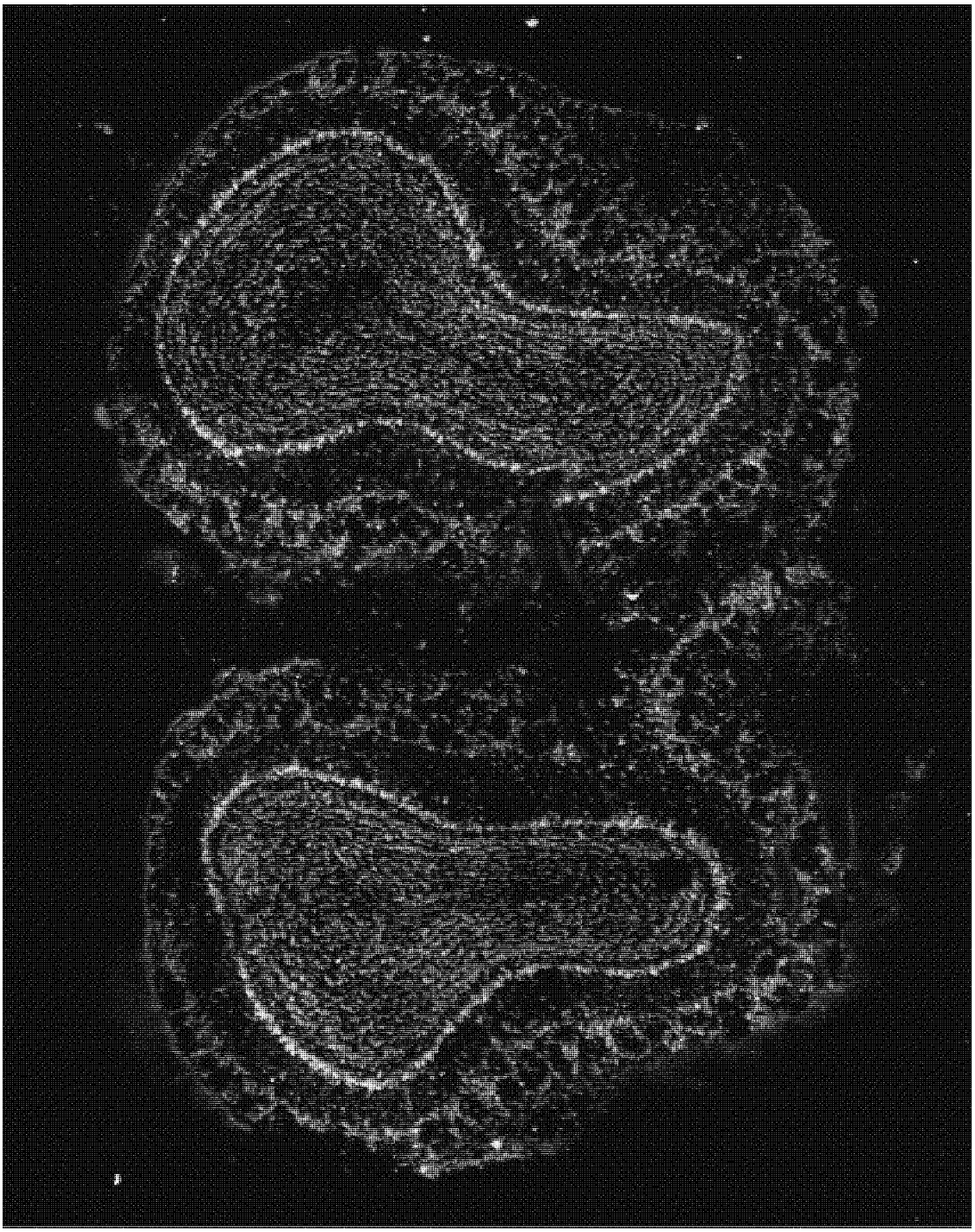


Figure 12

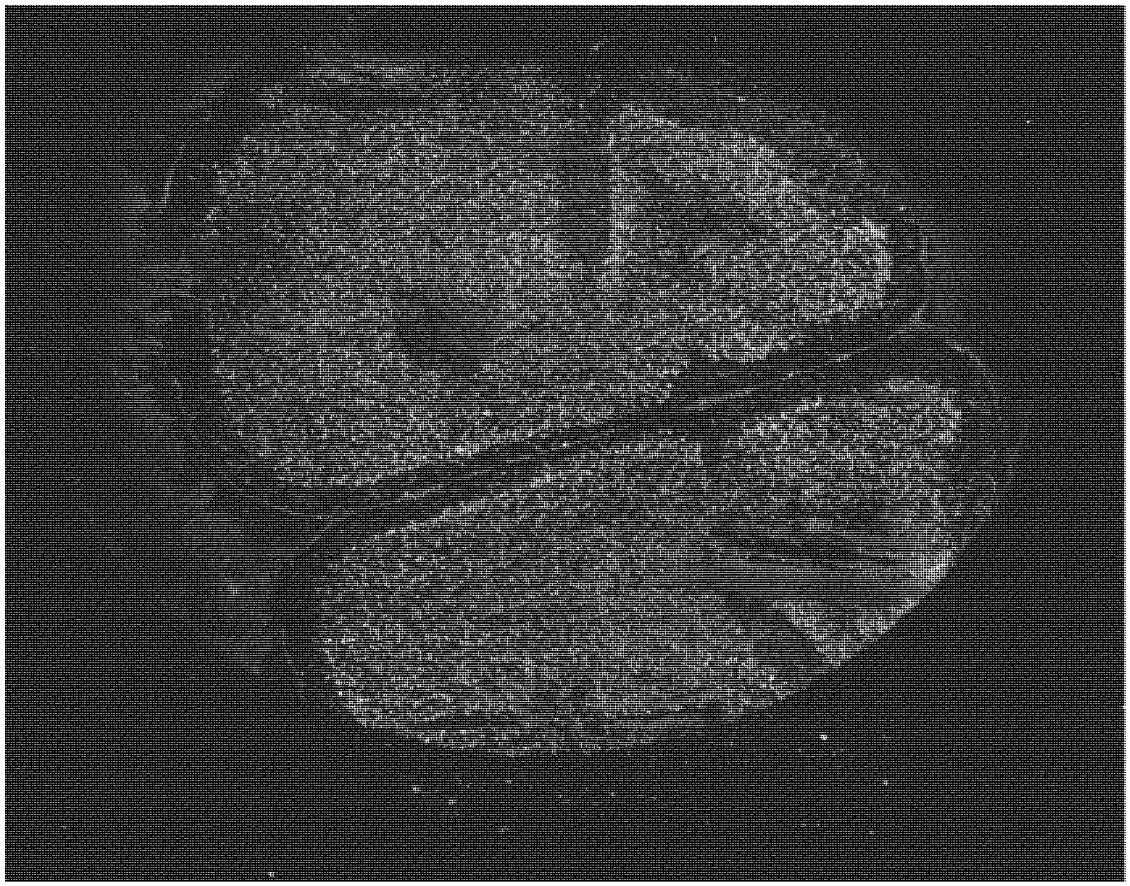


Figure 13

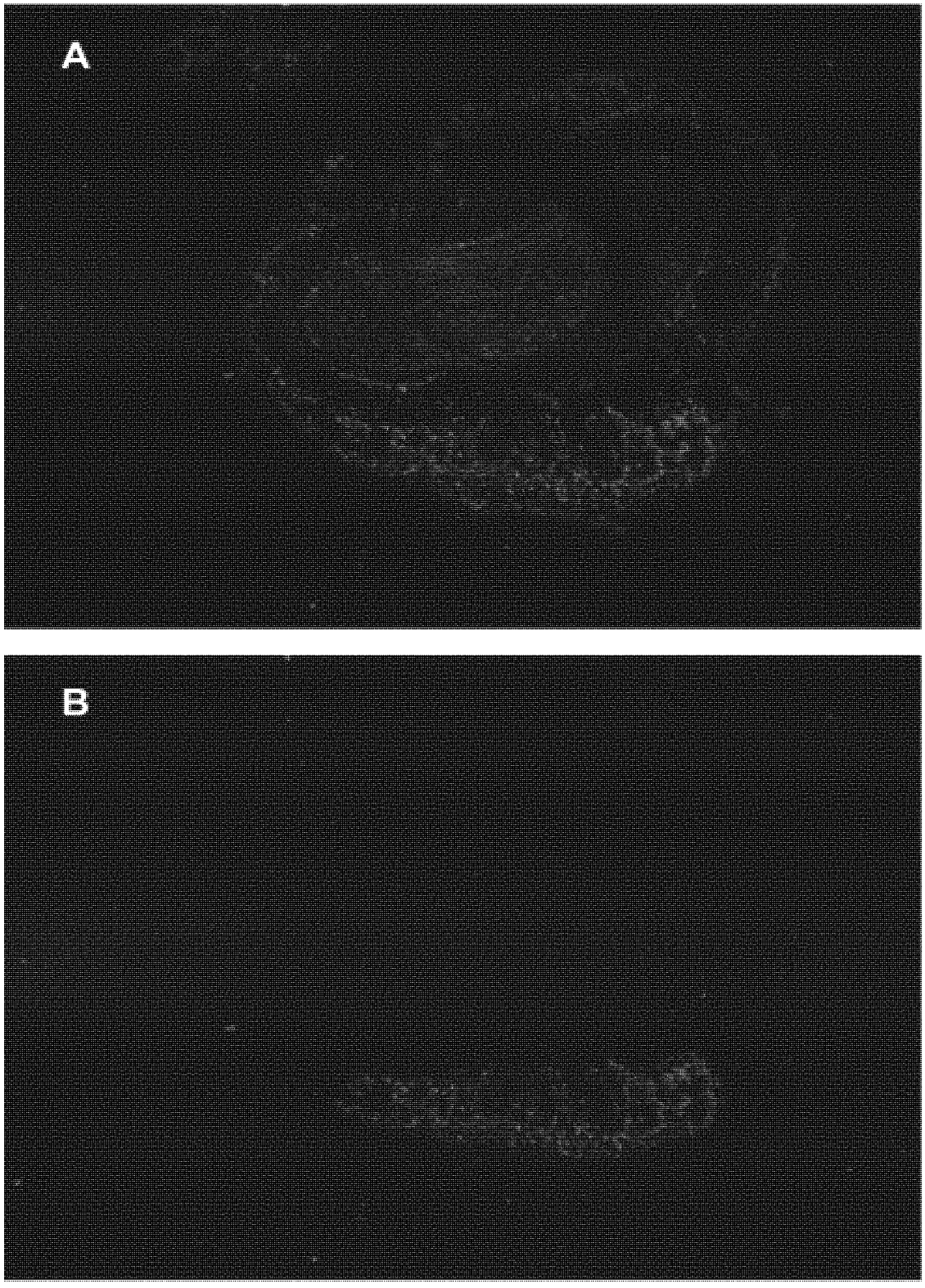


Figure 14

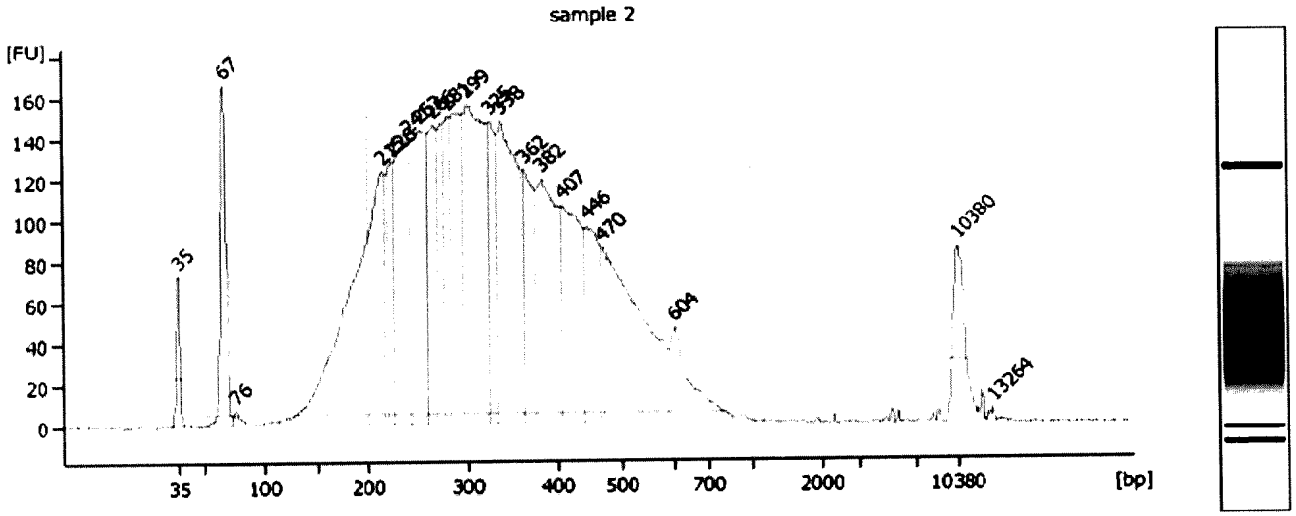

sample 2
[FU]
160
140
120
100
80
60
40
20
0
35
67
76
325
362
382
407
446
470
604
10380
13264
35
100
200
300
400
500
700
2000
10380
[bp]

Figure 15

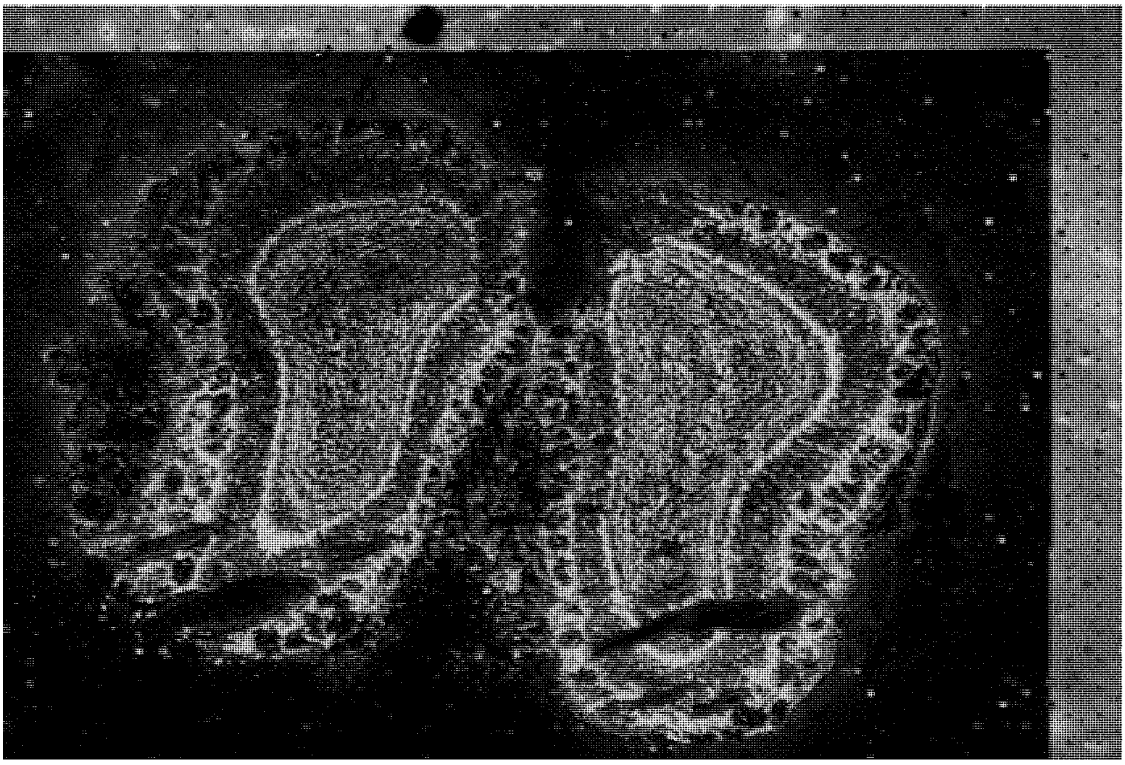


Figure 16

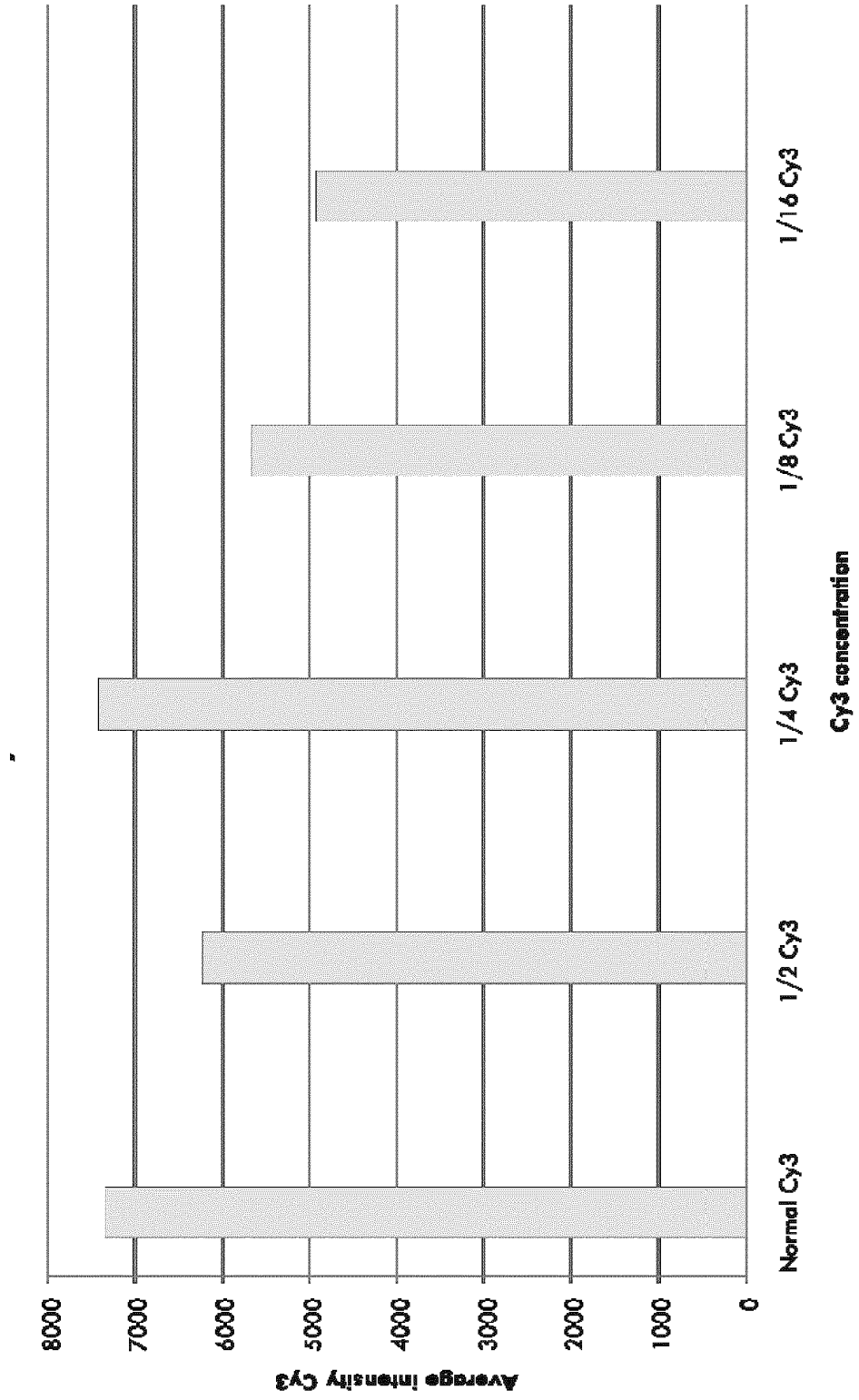
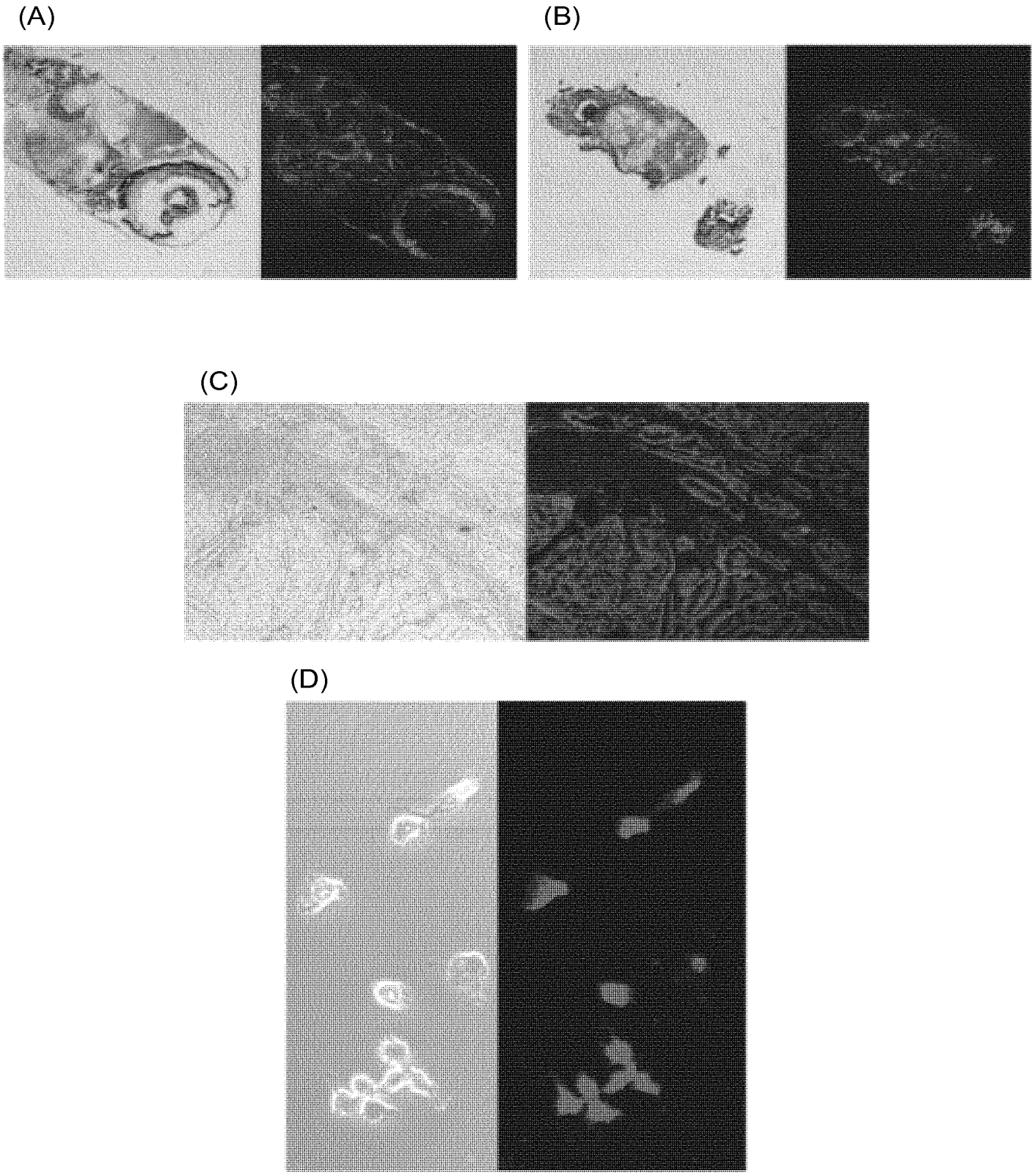

Average intensity Cy3
8000
7000
6000
5000
4000
3000
2000
1000
0
Normal Cy3
1/2 Cy3
1/4 Cy3
1/8 Cy3
1/16 Cy3
Cy3 concentration

Figure 17



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/071645

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, BIOSIS, Sequence Search, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	WO 2012/140224 A1 (FRISEN JONAS [SE]; STAAHL PATRIK [SE]; LUNDEBERG JOAKIM [SE]) 18 October 2012 (2012-10-18)	1-33
L	the whole document	1-33

X	WO 01/07915 A2 (US HEALTH [US]; EMMERT BUCK MICHAEL R [US]) 1 February 2001 (2001-02-01) claims 1,36	1

A	WO 02/24952 A1 (CLINOMICS LAB INC [US]; MURACA PATRICK J [US]) 28 March 2002 (2002-03-28) claims 1,2	1

A	US 2003/040035 A1 (SLAMON DENNIS J [US] ET AL) 27 February 2003 (2003-02-27) claim 1	1

	-/--	

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 November 2013	11/12/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Gabriels, Jan

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EGUILUZ C ET AL: "Multitissue array review: A chronological description of tissue array techniques, applications and procedures", PATHOLOGY RESEARCH AND PRACTICE, GUSTAV FISCHER, STUTTGART, DE, vol. 202, no. 8, 10 August 2006 (2006-08-10), pages 561-568, XP028031005, ISSN: 0344-0338, DOI: 10.1016/J.PRP.2006.04.003 [retrieved on 2006-08-10] the whole document	1
A	WO 00/24940 A1 (VYSIS INC [US]; US HEALTH [US]) 4 May 2000 (2000-05-04) the whole document	1

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2013/071645

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012140224 A1	18-10-2012	AU 2012241730 A1	31-10-2013
		WO 2012140224 A1	18-10-2012
WO 0107915 A2	01-02-2001	AT 328277 T	15-06-2006
		AU 771499 B2	25-03-2004
		AU 6609100 A	13-02-2001
		CA 2375034 A1	01-02-2001
		DE 60028393 T2	14-06-2007
		EP 1218743 A2	03-07-2002
		ES 2269171 T3	01-04-2007
		JP 4137444 B2	20-08-2008
		JP 2003505694 A	12-02-2003
		US 6602661 B1	05-08-2003
		US 2004081987 A1	29-04-2004
		WO 0107915 A2	01-02-2001
WO 0224952 A1	28-03-2002	AU 9099001 A	02-04-2002
		US 2002168639 A1	14-11-2002
		WO 0224952 A1	28-03-2002
US 2003040035 A1	27-02-2003	US 2003040035 A1	27-02-2003
		US 2004171101 A1	02-09-2004
WO 0024940 A1	04-05-2000	EP 1131470 A1	12-09-2001
		JP 5213202 B2	19-06-2013
		JP 2002528097 A	03-09-2002
		JP 2010162029 A	29-07-2010
		US 2002132246 A1	19-09-2002
		US 2005244880 A1	03-11-2005
		US 2009092993 A1	09-04-2009
		US 2013143770 A1	06-06-2013
		WO 0024940 A1	04-05-2000

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 October 2016 (13.10.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/162309 A1

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

(21) International Application Number:
PCT/EP2016/057355

(22) International Filing Date:
4 April 2016 (04.04.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/145,874 10 April 2015 (10.04.2015) US

(71) Applicants: **SPATIAL TRANSCRIPTOMICS AB** [SE/SE]; Tomtebodavägen 23a, 17165 Solna (SE). **ILLUMINA INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).

(72) Inventors: **FRISÉN, Jonas**; Bragevägen 5, 114 26 Stockholm (SE). **STÅHL, Patrik**; III-1-5 Fatburstrappan 18, T3:1, 118 26 Stockholm (SE). **LUNDEBERG, Joakim**; III-2-5 Torsvikssvän 20, 181 34 Lidingö (SE). **CANN, Gordon M.**; c/o Illumina Inc., 5200 Illumina Way, San Diego, CA 92122 (US). **BAZARGAN, Leila**; co/ Illumina Inc., 5200 Illumina Way, San Diego, CA 92122 (US). **ARAVANIS, Alex**; c/o Illumina Inc., 5200 Illumina Way, San Diego, CA 92122 (US).

(74) Agent: **WILKINS, Christopher**; St Bride's House, 10 Salisbury Square, London EC4Y 8JD (GB).

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

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

Published:

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

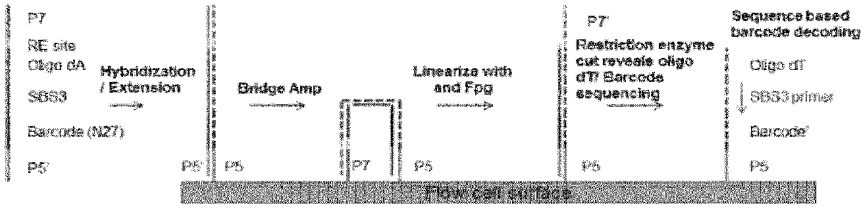
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2016/162309 A1

(54) **Title**: SPATIALLY DISTINGUISHED, MULTIPLEX NUCLEIC ACID ANALYSIS OF BIOLOGICAL SPECIMENS

Figure 1

A



(57) **Abstract**: A method for spatially tagging nucleic acids of a biological specimen, including steps of (a) providing a solid support comprising different nucleic acid probes that are randomly located on the solid support, wherein the different nucleic acid probes each includes a barcode sequence that differs from the barcode sequence of other randomly located probes on the solid support; (b) performing a nucleic acid detection reaction on the solid support to locate the barcode sequences on the solid support; (c) contacting a biological specimen with the solid support that has the randomly located probes; (d) hybridizing the randomly located probes to target nucleic acids from portions of the biological specimen; and (e) modifying the randomly located probes that are hybridized to the target nucleic acids, thereby producing modified probes that include the barcode sequences and a target specific modification, thereby spatially tagging the nucleic acids of the biological specimen.

SPATIALLY DISTINGUISHED, MULTIPLEX NUCLEIC ACID ANALYSIS OF BIOLOGICAL SPECIMENS

BACKGROUND

One of every four men will die of cancer. Further statistics from the American Cancer Society predict that one of every five women will suffer the same fate. Treatments are available for many cancers. However, success for most relies on early detection.

Cancer is now said to be a disease of the genome. Many oncologists and cancer researchers hope that advances in genomic analysis tools will provide early detection and a path to treatment. However, these tools are more prominent in research labs having not yet matured to the level of being readily available to the vast majority of oncologists. Improvements are needed.

It has been said that at the time of diagnosis, all cancer patients are mosaics. They are mosaics because they have at least two distinct genomes: the genome they were born with, and the genome that they unwillingly acquired via cancer. Furthermore, as tumors grow, distinct populations of cancer cells become apparent. Leading to even more complex mosaics within the tumor. This cancer cell heterogeneity often results in subpopulations of cells that respond differently to cancer therapies. The end result is often an initial positive response of one subpopulation of cells, resulting in the observation of the patient's tumor shrinking, only to be followed by regrowth of tumor tissue, and in some cases metastasis. Despite early detection of the tumor, an inability to identify the subpopulation of cells that are resistant to the treatment can result in loss of time needed to treat an aggressive cancer. This creates adverse consequences for the patient both emotionally and physically.

There is a need for genomic tools that can distinguish subpopulations of cancer cells in tumors. The present disclosure addresses this need and provides other advantages as well.

BRIEF SUMMARY

The present disclosure provides a method for spatially tagging nucleic acids of a biological specimen. The method can include steps of (a) providing a solid support comprising a plurality of different nucleic acid probes that are randomly located on the solid support, wherein the different nucleic acid probes each includes

a barcode sequence that is different from the barcode sequence of other randomly located probes on the solid support; (b) performing a nucleic acid detection reaction on the solid support to locate the barcode sequences on the solid support; (c) contacting a biological specimen with the solid support that has the randomly located probes; (d) hybridizing the randomly located probes to target nucleic acids from portions of the biological specimen that are proximal to the randomly located probes; and (e) modifying the randomly located probes that are hybridized to the target nucleic acids, thereby producing modified probes that include the barcode sequences and a target specific modification, thereby spatially tagging the nucleic acids of the biological specimen.

This disclosure further provides a method for spatially tagging nucleic acids of a biological specimen, the method including steps of (a) attaching different nucleic acid probes to a solid support to produce randomly located probes on the solid support, wherein the different nucleic acid probes each includes a barcode sequence, and wherein each of the randomly located probes includes different barcode sequences from other randomly located probes on the solid support; (b) performing a nucleic acid detection reaction on the solid support to determine the barcode sequences of the randomly located probes on the solid support; (c) contacting a biological specimen with the solid support that has the randomly located probes; (d) hybridizing the randomly located probes to target nucleic acids from portions of the biological specimen that are proximal to the randomly located probes; and (e) extending the randomly located probes to produce extended probes that include the barcode sequences and sequences from the target nucleic acids, thereby spatially tagging the nucleic acids of the biological specimen.

Also provided is a method for spatially tagging nucleic acids of a biological specimen that includes the steps of (a) providing a plurality of nucleic acid primers attached to a solid support, wherein the nucleic acid primers in the plurality include a universal primer sequence that is common to the nucleic acid primers in the plurality; (b) binding a population of nucleic acid probes to the plurality of nucleic acid primers, wherein the nucleic acid probes include a universal primer binding sequence that hybridizes to the universal primer sequence, a target capture sequence and a barcode sequence that differs from barcode sequences of other nucleic acid probes in the population, thereby attaching the different nucleic acid probes at randomly located positions on the solid support; (c) amplifying the different nucleic acid probes by extension of the nucleic acid primers, thereby

producing nucleic acid clusters having copies of the barcode sequence and target capture sequence at the randomly located positions on the solid support; (d) performing a sequencing reaction to determine the barcode sequences at the randomly located positions on the solid support; (e) contacting a biological specimen with the nucleic acid clusters on the solid support; (f) hybridizing the target capture sequences of the clusters to target nucleic acids from portions of the biological specimen that are proximal to the clusters; and (g) extending the target capture sequences to produce extended probes that include sequences from the target nucleic acids and the copies of the barcode sequences, thereby tagging the nucleic acids of the biological specimen.

This disclosure further provides a method for spatially tagging nucleic acids of a biological specimen, the method including steps of (a) providing an array of beads on a solid support, wherein different nucleic acid probes are attached to different beads in the array, wherein the different nucleic acid probes each include a barcode sequence, wherein each bead includes a different barcode sequence from other beads on the solid support, and wherein each of the different nucleic acid probes includes a target capture sequence; (b) performing a decoder probe hybridization reaction on the solid support to determine the barcode sequences at the randomly located probes on the solid support; (c) contacting a biological specimen with the array of beads; (d) hybridizing the different nucleic acid probes to target nucleic acids from portions of the biological specimen that are proximal to the beads; and (e) extending the different nucleic acid probes to produce extended probes that include sequences from the target nucleic acids and the barcode sequences, thereby tagging the nucleic acids of the biological specimen.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows a diagrammatic representation of steps and reagents that can be used to generate barcoded oligo dT probes on an Illumina flow cell, create extended barcoded probes having mRNA sequences and releasing the extended probes from the flow cell.

Fig. 2 shows data indicating the availability of oligo dT capture sequences on probes after bridge amplification of the probes and restriction enzyme digest with BspH1 to remove one of the primer binding sites used for bridge amplification.

Fig. 3 shows sequencing metrics of the flow cell described in Example 1 and shown in Fig. 2.

Fig. 4 the number of unique barcodes determined in 21 tiles of the flow cell described in Example 1 and shown in Fig. 2.

Fig. 5 shows an image of cells captured on a patterned flow cell (Panel A) and cell count data (Panel B).

Fig. 6 shows cells that remain adhered to a flow cell in different conditions.

Fig. 7 shows a diagrammatic representation of steps and reagents used to create probes attached to a gel (Panel A), a diagrammatic representation of steps and reagents used to capture target nucleic acids using the gel-attached probes and fluorescently label the probes (Panel B), and an image created by the fluorescently labeled target nucleic acids following capture by the probes and removal of the tissue from the gel.

Fig. 8 shows a diagrammatic representation of steps and reagents used to capture target nucleic acids using BeadArray™ -attached probes and fluorescently label the probes (Panel A), and an image created by the fluorescently labeled target nucleic acids following capture by the probes and removal of the tissue from the BeadArray™.

DETAILED DESCRIPTION

The present disclosure provides compositions, apparatus and methods for preserving spatial information when performing multiplex nucleic acid analyses of biological specimens. A variety of tools are available for multiplex nucleic acid analyses including, for example, nucleic acid microarrays and so-called “next generation” sequencing platforms. Such tools allow for parallel detection of very large and complex collections of nucleic acids, including for example, DNA collections that represent all or nearly all of the genetic material of an organism (i.e. the ‘genome’), RNA (or cDNA) collections that represent all or nearly all of the complement of expressed genes (i.e. the ‘transcriptome’) for an organism, and in some cases the collections can include several genomes and/or transcriptomes from several different organisms (e.g. a metabolome or biome from a community or ecosystem). Although these tools provide a vast amount of information about *what* nucleic acid sequences are present in a biological specimen being evaluated, they do not inherently distinguish *where* any particular nucleic acid resided in the biological specimen. Indeed the vast majority of samples applied to multiplex nucleic acid analysis tools are homogenates derived from mixtures of many different cells from a biological specimen. As a result, spatial information is lost and

the results obtained from these tools constitute an average transcriptome or average genome for the specimen, important differences between individual cells being lost.

In particular embodiments, the present disclosure provides new and useful modifications to existing multiplex nucleic acid analysis tools to allow for the preservation of spatial information for biological specimens from which the nucleic acids are obtained. For example, solid supports that are usually used for multiplex sequencing-by-synthesis (SBS) techniques can be modified for use in capturing and spatially tagging nucleic acids from a biological specimen. In an alternative example, arrays of beads, such as those used for genotyping or gene expression analysis, can be used for capturing and spatially tagging nucleic acids from a biological specimen. As set forth in examples below, the solid supports used for an SBS or BeadArray™ platform commercialized by Illumina (San Diego, CA) can be modified for spatial tagging. However, it will be understood that any of a variety of solid supports can be made and used in accordance with the teaching herein. The spatially tagged nucleic acids can be removed from the solid support, pooled together and attached to a second solid support for detection in any of a variety of multiplex nucleic acid analysis systems including, for example, a sequencing platform or microarray platform set forth herein.

The spatial information provided by a method, composition or apparatus herein can include, for example, the location of one or more cells in a tissue (or other specimen) that has a particular allele at one or more locus (e.g. a genotype), has a particular structural variation in the genome (e.g. fusion, insertion, deletion, rearrangement etc.), has a particular epigenetic signature (e.g. methylation), expresses a particular gene, expresses a particular allele of a gene, expresses a particular splice variant of a gene or the like. In addition to identifying nucleic acids according to their spatial location in a biological specimen, a method, composition or apparatus of the present disclosure can be used to quantify one or more nucleic acids according to spatial location. For example, the spatial information for one or more cells in a tissue (or other specimen) can include the amount of a particular allele or chromosomal region in a genome (e.g. ploidy); the amount of epigenetic modification of a genetic locus (e.g. methylation); expression level for a particular gene, allele or splice variant; or the like. The amounts can be absolute amounts or relative amounts in accordance with similar measurements obtained in the art for mixed or non-spatially tagged samples.

A method set forth herein can be used for localized detection of a nucleic acid in a biological specimen. In some embodiments, a method can be used for identifying or characterizing all of the transcriptome or genome of a biological specimen. Alternatively, a method can be used to identify or characterize only a part of a specimen's transcriptome or genome. A subset of transcripts or genes evaluated in a method herein can be related to a particular disease or condition.

A method set forth herein can be used for localized or spatial detection of nucleic acids, whether DNA or RNA, in a biological specimen. Thus one or more RNA or DNA molecules can be located with respect to its native position or location within a cell or tissue or other biological specimen. For example, one or more nucleic acids can be localized to a cell or group of adjacent cells, or type of cell, or to particular regions of areas within a tissue sample. The native location or position of individual RNA or DNA molecules can be determined using a method, apparatus or composition of the present disclosure.

Terms used herein will be understood to take on their ordinary meaning in the relevant art unless specified otherwise. Several terms used herein and their meanings are set forth below.

As used herein, the term "amplicon," when used in reference to a nucleic acid, means the product of copying the nucleic acid, wherein the product has a nucleotide sequence that is the same as or complementary to at least a portion of the nucleotide sequence of the nucleic acid. An amplicon can be produced by any of a variety of amplification methods that use the nucleic acid, or an amplicon thereof, as a template including, for example, polymerase extension, polymerase chain reaction (PCR), rolling circle amplification (RCA), multiple displacement amplification (MDA), ligation extension, or ligation chain reaction. An amplicon can be a nucleic acid molecule having a single copy of a particular nucleotide sequence (e.g. a PCR product) or multiple copies of the nucleotide sequence (e.g. a concatameric product of RCA). A first amplicon of a target nucleic acid is typically a complementary copy. Subsequent amplicons are copies that are created, after generation of the first amplicon, from the target nucleic acid or from the first amplicon. A subsequent amplicon can have a sequence that is substantially complementary to the target nucleic acid or substantially identical to the target nucleic acid.

As used herein, the term "array" refers to a population of features or sites that can be differentiated from each other according to relative location. Different

molecules that are at different sites of an array can be differentiated from each other according to the locations of the sites in the array. An individual site of an array can include one or more molecules of a particular type. For example, a site can include a single target nucleic acid molecule having a particular sequence or a site can include several nucleic acid molecules having the same sequence (and/or complementary sequence, thereof). The sites of an array can be different features located on the same substrate. Exemplary features include without limitation, wells in a substrate, beads (or other particles) in or on a substrate, projections from a substrate, ridges on a substrate or channels in a substrate. The sites of an array can be separate substrates each bearing a different molecule. Different molecules attached to separate substrates can be identified according to the locations of the substrates on a surface to which the substrates are associated or according to the locations of the substrates in a liquid or gel. Exemplary arrays in which separate substrates are located on a surface include, without limitation, those having beads in wells.

As used herein, the term “attached” refers to the state of two things being joined, fastened, adhered, connected or bound to each other. For example, an analyte, such as a nucleic acid, can be attached to a material, such as a gel or solid support, by a covalent or non-covalent bond. A covalent bond is characterized by the sharing of pairs of electrons between atoms. A non-covalent bond is a chemical bond that does not involve the sharing of pairs of electrons and can include, for example, hydrogen bonds, ionic bonds, van der Waals forces, hydrophilic interactions and hydrophobic interactions.

As used herein, the term “barcode sequence” is intended to mean a series of nucleotides in a nucleic acid that can be used to identify the nucleic acid, a characteristic of the nucleic acid, or a manipulation that has been carried out on the nucleic acid. The barcode sequence can be a naturally occurring sequence or a sequence that does not occur naturally in the organism from which the barcoded nucleic acid was obtained. A barcode sequence can be unique to a single nucleic acid species in a population or a barcode sequence can be shared by several different nucleic acid species in a population. For example, each nucleic acid probe in a population can include different barcode sequences from all other nucleic acid probes in the population. Alternatively, each nucleic acid probe in a population can include different barcode sequences from some or most other nucleic acid probes in a population. For example, each probe in a population can have a barcode that is

present for several different probes in the population even though the probes with the common barcode differ from each other at other sequence regions along their length. In particular embodiments, one or more barcode sequences that are used with a biological specimen are not present in the genome, transcriptome or other nucleic acids of the biological specimen. For example, barcode sequences can have less than 80%, 70%, 60%, 50% or 40% sequence identity to the nucleic acid sequences in a particular biological specimen.

As used herein, the term "biological specimen" is intended to mean one or more cell, tissue, organism or portion thereof. A biological specimen can be obtained from any of a variety of organisms. Exemplary organisms include, but are not limited to, a mammal such as a rodent, mouse, rat, rabbit, guinea pig, ungulate, horse, sheep, pig, goat, cow, cat, dog, primate (i.e. human or non-human primate); a plant such as *Arabidopsis thaliana*, corn, sorghum, oat, wheat, rice, canola, or soybean; an algae such as *Chlamydomonas reinhardtii*; a nematode such as *Caenorhabditis elegans*; an insect such as *Drosophila melanogaster*, mosquito, fruit fly, honey bee or spider; a fish such as zebrafish; a reptile; an amphibian such as a frog or *Xenopus laevis*; a *Dictyostelium discoideum*; a fungi such as *Pneumocystis carinii*, *Takifugu rubripes*, yeast, *Saccharamoyces cerevisiae* or *Schizosaccharomyces pombe*; or a Plasmodium falciparum. Target nucleic acids can also be derived from a prokaryote such as a bacterium, *Escherichia coli*, *Staphylococci* or *Mycoplasma pneumoniae*; an archae; a virus such as Hepatitis C virus or human immunodeficiency virus; or a viroid. Specimens can be derived from a homogeneous culture or population of the above organisms or alternatively from a collection of several different organisms, for example, in a community or ecosystem.

As used herein, the term "cleavage site" is intended to mean a location in a nucleic acid molecule that is susceptible to bond breakage. The location can be specific to a particular chemical, enzymatic or physical process that results in bond breakage. For example, the location can be a nucleotide that is abasic or a nucleotide that has a base that is susceptible to being removed to create an abasic site. Examples of nucleotides that are susceptible to being removed include uracil and 8-oxo-guanine as set forth in further detail herein below. The location can also be at or near a recognition sequence for a restriction endonuclease such as a nicking enzyme.

As used herein, the term “cluster,” when used in reference to nucleic acids, refers to a population of the nucleic acids that is attached to a solid support to form a feature or site. The nucleic acids are generally members of a single species, thereby forming a monoclonal cluster. A “monoclonal population" of nucleic acids is a population that is homogeneous with respect to a particular nucleotide sequence. Clusters need not be monoclonal. Rather, for some applications, a cluster can be predominantly populated with amplicons from a first nucleic acid and can also have a low level of contaminating amplicons from a second nucleic acid. For example, when an array of clusters is to be used in a detection application, an acceptable level of contamination would be a level that does not impact signal to noise or resolution of the detection technique in an unacceptable way. Accordingly, apparent clonality will generally be relevant to a particular use or application of an array made by the methods set forth herein. Exemplary levels of contamination that can be acceptable at an individual cluster include, but are not limited to, at most 0.1%, 0.5%, 1%, 5%, 10%, 5 25%, or 35% contaminating amplicons. The nucleic acids in a cluster are generally covalently attached to a solid support, for example, via their 5’ ends, but in some cases other attachment means are possible. The nucleic acids in a cluster can be single stranded or double stranded. In some but not all embodiments, clusters are made by a solid-phase amplification method known as bridge amplification. Exemplary configurations for clusters and methods for their production are set forth, for example, in U.S. Pat. No. 5,641,658; U.S. Patent Publ. No. 2002/0055100; U.S. Pat. No. 7,115,400; U.S. Patent Publ. No. 2004/0096853; U.S. Patent Publ. No. 2004/0002090; U.S. Patent Publ. No. 2007/0128624; and U.S. Patent Publ. No. 2008/0009420, each of which is incorporated herein by reference.

As used herein, the term “different", when used in reference to nucleic acids, means that the nucleic acids have nucleotide sequences that are not the same as each other. Two or more nucleic acids can have nucleotide sequences that are different along their entire length. Alternatively, two or more nucleic acids can have nucleotide sequences that are different along a substantial portion of their length. For example, two or more nucleic acids can have target nucleotide sequence portions that are different for the two or more molecules while also having a universal sequence portion that is the same on the two or more molecules. Two beads can be different from each other by virtue of being attached to different nucleic acids.

As used herein, the term "each," when used in reference to a collection of items, is intended to identify an individual item in the collection but does not necessarily refer to every item in the collection. Exceptions can occur if explicit disclosure or context clearly dictates otherwise.

As used herein, the term "extend," when used in reference to a nucleic acid, is intended to mean addition of at least one nucleotide or oligonucleotide to the nucleic acid. In particular embodiments one or more nucleotides can be added to the 3' end of a nucleic acid, for example, via polymerase catalysis (e.g. DNA polymerase, RNA polymerase or reverse transcriptase). Chemical or enzymatic methods can be used to add one or more nucleotide to the 3' or 5' end of a nucleic acid. One or more oligonucleotides can be added to the 3' or 5' end of a nucleic acid, for example, via chemical or enzymatic (e.g. ligase catalysis) methods. A nucleic acid can be extended in a template directed manner, whereby the product of extension is complementary to a template nucleic acid that is hybridized to the nucleic acid that is extended.

As used herein, the term "feature" means a location in an array for a particular species of molecule. A feature can contain only a single molecule or it can contain a population of several molecules of the same species. Features of an array are typically discrete. The discrete features can be contiguous or they can have spaces between each other. The size of the features and/or spacing between the features can vary such that arrays can be high density, medium density or lower density. High density arrays are characterized as having sites separated by less than about 15 µm. Medium density arrays have sites separated by about 15 to 30 µm, while low density arrays have sites separated by greater than 30 µm. An array useful herein can have, for example, sites that are separated by less than 100 µm, 50 µm, 10 µm, 5 µm, 1 µm, or 0.5 µm. An apparatus or method of the present disclosure can be used to detect an array at a resolution sufficient to distinguish sites at the above densities or density ranges.

As used herein, the term "fluidic mixture" is intended to mean two or more different items that are simultaneously present in a solution. Typically, the two or more items are freely diffusible in the solution. The two or more items can be different types of items (e.g. a nucleic acid and a protein which are different types of molecules) or they can be different species of the same type of items (e.g. two nucleic acid molecules having different sequences). Exemplary items that can be in a fluidic mixture include, but are not limited to, molecules, cells or beads.

As used herein, the term “flow cell” is intended to mean a vessel having a chamber where a reaction can be carried out, an inlet for delivering reagents to the chamber and an outlet for removing reagents from the chamber. In some embodiments the chamber is configured for detection of the reaction that occurs in the chamber. For example, the chamber can include one or more transparent surfaces allowing optical detection of biological specimens, optically labeled molecules, or the like in the chamber. Exemplary flow cells include, but are not limited to those used in a nucleic acid sequencing apparatus such as flow cells for the Genome Analyzer®, MiSeq®, NextSeq® or HiSeq® platforms commercialized by Illumina, Inc. (San Diego, CA); or for the SOLiD™ or Ion Torrent™ sequencing platform commercialized by Life Technologies (Carlsbad, CA). Exemplary flow cells and methods for their manufacture and use are also described, for example, in WO 2014/142841 A1; U.S. Pat. App. Pub. No. 2010/0111768 A1 and U.S. Pat. No. 8,951,781, each of which is incorporated herein by reference.

As used herein, the term “gel” is intended to mean a semi-rigid material that is permeable to liquids and gases. Typically, gel material can swell when liquid is taken up and can contract when liquid is removed by drying. Exemplary gels include, but are not limited to those having a colloidal structure, such as agarose; polymer mesh structure, such as gelatin; or cross-linked polymer structure, such as polyacrylamide, SFA (see, for example, US Pat. App. Pub. No. 2011/0059865 A1, which is incorporated herein by reference) or PAZAM (see, for example, US Pat. App. Publ. No. 2014/0079923 A1, which is incorporated herein by reference). Particularly useful gel material will conform to the shape of a well or other concave feature where it resides.

As used herein, the terms "nucleic acid" and “nucleotide” are intended to be consistent with their use in the art and to include naturally occurring species or functional analogs thereof. Particularly useful functional analogs of nucleic acids are capable of hybridizing to a nucleic acid in a sequence specific fashion or capable of being used as a template for replication of a particular nucleotide sequence. Naturally occurring nucleic acids generally have a backbone containing phosphodiester bonds. An analog structure can have an alternate backbone linkage including any of a variety of those known in the art. Naturally occurring nucleic acids generally have a deoxyribose sugar (e.g. found in deoxyribonucleic acid (DNA)) or a ribose sugar (e.g. found in ribonucleic acid (RNA)). A nucleic acid can contain nucleotides having any of a variety of analogs of these sugar moieties that are

known in the art. A nucleic acid can include native or non-native nucleotides. In this regard, a native deoxyribonucleic acid can have one or more bases selected from the group consisting of adenine, thymine, cytosine or guanine and a ribonucleic acid can have one or more bases selected from the group consisting of uracil, adenine, cytosine or guanine. Useful non-native bases that can be included in a nucleic acid or nucleotide are known in the art. The terms “probe” or "target," when used in reference to a nucleic acid or sequence of a nucleic acid, are intended as semantic identifiers for the nucleic acid or sequence in the context of a method or composition set forth herein and does not necessarily limit the structure or function of the nucleic acid or sequence beyond what is otherwise explicitly indicated. The terms “probe” and “target” can be similarly applied to other analytes such as proteins, small molecules, cells or the like.

As used herein, the term “pitch,” when used in reference to features of an array, is intended to refer to the center-to-center spacing for adjacent features. A pattern of features can be characterized in terms of average pitch. The pattern can be ordered such that the coefficient of variation around the average pitch is small or the pattern can be random in which case the coefficient of variation can be relatively large. In either case, the average pitch can be, for example, at least about 10 nm, 0.1 µm, 0.5 µm, 1 µm, 5 µm, 10 µm, 100 µm or more. Alternatively or additionally, the average pitch can be, for example, at most about 100 µm, 10 µm, 5 µm, 1 µm, 0.5 µm 0.1 µm or less. Of course, the average pitch for a particular pattern of features can be between one of the lower values and one of the upper values selected from the ranges above.

As used herein, the term “poly T or poly A,” when used in reference to a nucleic acid sequence, is intended to mean a series of two or more thiamine (T) or adenine (A) bases, respectively. A poly T or poly A can include at least about 2, 5, 8, 10, 12, 15, 18, 20 or more of the T or A bases, respectively. Alternatively or additionally, a poly T or poly A can include at most about, 30, 20, 18, 15, 12, 10, 8, 5 or 2 of the T or A bases, respectively.

As used herein, the term “random” can be used to refer to the spatial arrangement or composition of locations on a surface. For example, there are at least two types of order for an array described herein, the first relating to the spacing and relative location of features (also called “sites”) and the second relating to identity or predetermined knowledge of the particular species of molecule that is present at a particular feature. Accordingly, features of an array can be randomly

spaced such that nearest neighbor features have variable spacing between each other. Alternatively, the spacing between features can be ordered, for example, forming a regular pattern such as a rectilinear grid or hexagonal grid. In another respect, features of an array can be random with respect to the identity or predetermined knowledge of the species of analyte (e.g. nucleic acid of a particular sequence) that occupies each feature independent of whether spacing produces a random pattern or ordered pattern. An array set forth herein can be ordered in one respect and random in another. For example, in some embodiments set forth herein a surface is contacted with a population of nucleic acids under conditions where the nucleic acids attach at sites that are ordered with respect to their relative locations but 'randomly located' with respect to knowledge of the sequence for the nucleic acid species present at any particular site. Reference to "randomly distributing" nucleic acids at locations on a surface is intended to refer to the absence of knowledge or absence of predetermination regarding which nucleic acid will be captured at which location (regardless of whether the locations are arranged in an ordered pattern or not).

As used herein, the term "solid support" refers to a rigid substrate that is insoluble in aqueous liquid. The substrate can be non-porous or porous. The substrate can optionally be capable of taking up a liquid (e.g. due to porosity) but will typically be sufficiently rigid that the substrate does not swell substantially when taking up the liquid and does not contract substantially when the liquid is removed by drying. A nonporous solid support is generally impermeable to liquids or gases. Exemplary solid supports include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon™, cyclic olefins, polyimides etc.), nylon, ceramics, resins, Zeonor, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, optical fiber bundles, and polymers. Particularly useful solid supports for some embodiments are located within a flow cell apparatus. Exemplary flow cells are set forth in further detail herein.

As used herein, the term "spatial tag" is intended to mean a nucleic acid having a sequence that is indicative of a location. Typically, the nucleic acid is a synthetic molecule having a sequence that is not found in one or more biological specimen that will be used with the nucleic acid. However, in some embodiments the nucleic acid molecule can be naturally derived or the sequence of the nucleic

acid can be naturally occurring, for example, in a biological specimen that is used with the nucleic acid. The location indicated by a spatial tag can be a location in or on a biological specimen, in or on a solid support or a combination thereof. A barcode sequence can function as a spatial tag.

As used herein, the term “tissue” is intended to mean an aggregation of cells, and, optionally, intercellular matter. Typically the cells in a tissue are not free floating in solution and instead are attached to each other to form a multicellular structure. Exemplary tissue types include muscle, nerve, epidermal and connective tissues.

As used herein, the term "universal sequence" refers to a series of nucleotides that is common to two or more nucleic acid molecules even if the molecules also have regions of sequence that differ from each other. A universal sequence that is present in different members of a collection of molecules can allow capture of multiple different nucleic acids using a population of universal capture nucleic acids that are complementary to the universal sequence. Similarly, a universal sequence present in different members of a collection of molecules can allow the replication or amplification of multiple different nucleic acids using a population of universal primers that are complementary to the universal sequence. Thus, a universal capture nucleic acid or a universal primer includes a sequence that can hybridize specifically to a universal sequence. Target nucleic acid molecules may be modified to attach universal adapters, for example, at one or both ends of the different target sequences.

The embodiments set forth below and recited in the claims can be understood in view of the above definitions.

The present disclosure provides a method for spatially tagging nucleic acids of a biological specimen. The method can include the steps of (a) attaching different nucleic acid probes to a solid support to produce randomly located probes on the solid support, wherein the different nucleic acid probes each includes a barcode sequence, and wherein each of the randomly located probes includes different barcode sequences from other randomly located probes on the solid support; (b) performing a nucleic acid detection reaction on the solid support to determine the barcode sequences of the randomly located probes on the solid support; (c) contacting a biological specimen with the solid support that has the randomly located probes; (d) hybridizing the randomly located probes to target nucleic acids from portions of the biological specimen that are proximal to the randomly located

probes; and (e) extending the randomly located probes to produce extended probes that include the barcode sequences and sequences from the target nucleic acids, thereby spatially tagging the nucleic acids of the biological specimen.

Any of a variety of solid supports can be used in a method, composition or apparatus of the present disclosure. Particularly useful solid supports are those used for nucleic acid arrays. Examples include glass, modified glass, functionalized glass, inorganic glasses, microspheres (e.g. inert and/or magnetic particles), plastics, polysaccharides, nylon, nitrocellulose, ceramics, resins, silica, silica-based materials, carbon, metals, an optical fiber or optical fiber bundles, polymers and multiwell (e.g. microtiter) plates. Exemplary plastics include acrylics, polystyrene, copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes and Teflon™. Exemplary silica-based materials include silicon and various forms of modified silicon.

In particular embodiments, a solid support can be within or part of a vessel such as a well, tube, channel, cuvette, Petri plate, bottle or the like. A particularly useful vessel is a flow-cell, for example, as described in WO 2014/142841 A1; U.S. Pat. App. Pub. No. 2010/0111768 A1 and U.S. Pat. No. 8,951,781 or Bentley et al., *Nature* 456:53-59 (2008), each of which is incorporated herein by reference. Exemplary flow-cells are those that are commercially available from Illumina, Inc. (San Diego, CA) for use with a sequencing platform such as a Genome Analyzer®, MiSeq®, NextSeq® or HiSeq® platform. Another particularly useful vessel is a well in a multiwell plate or microtiter plate.

Optionally, a solid support can include a gel coating. Attachment of nucleic acids to a solid support via a gel is exemplified by flow cells available commercially from Illumina Inc. (San Diego, CA) or described in US Pat. App. Pub. Nos. 2011/0059865 A1, 2014/0079923 A1, or 2015/0005447 A1; or PCT Publ. No. WO 2008/093098, each of which is incorporated herein by reference. Exemplary gels that can be used in the methods and apparatus set forth herein include, but are not limited to, those having a colloidal structure, such as agarose; polymer mesh structure, such as gelatin; or cross-linked polymer structure, such as polyacrylamide, SFA (see, for example, US Pat. App. Pub. No. 2011/0059865 A1, which is incorporated herein by reference) or PAZAM (see, for example, US Pat. App. Publ. Nos. 2014/0079923 A1, or 2015/0005447 A1, each of which is incorporated herein by reference).

In some embodiments, a solid support can be configured as an array of features to which nucleic acids can be attached. The features can be present in any of a variety of desired formats. For example, the features can be wells, pits, channels, ridges, raised regions, pegs, posts or the like. In some embodiments, the features can contain beads. However, in particular embodiments the features need not contain a bead or particle. Exemplary features include wells that are present in substrates used for commercial sequencing platforms sold by 454 LifeSciences (a subsidiary of Roche, Basel Switzerland) or Ion Torrent (a subsidiary of Life Technologies, Carlsbad California). Other substrates having wells include, for example, etched fiber optics and other substrates described in US Pat Nos. 6,266,459; 6,355,431; 6,770,441; 6,859,570; 6,210,891; 6,258,568; 6,274,320; US Pat app. Publ. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; 2010/0282617 A1 or PCT Publication No. WO 00/63437, each of which is incorporated herein by reference. In some embodiments, wells of a substrate can include gel material (with or without beads) as set forth in US Pat. App. Publ. No. 2014/0243224 A1, which is incorporated herein by reference.

The features on a solid support can be metal features on a non-metallic surface such as glass, plastic or other materials exemplified above. A metal layer can be deposited on a surface using methods known in the art such as wet plasma etching, dry plasma etching, atomic layer deposition, ion beam etching, chemical vapor deposition, vacuum sputtering or the like. Any of a variety of commercial instruments can be used as appropriate including, for example, the FlexAL®, OpAL®, Ionfab 300plus®, or Optofab 3000® systems (Oxford Instruments, UK). A metal layer can also be deposited by e-beam evaporation or sputtering as set forth in Thornton, *Ann. Rev. Mater. Sci.* 7:239-60 (1977), which is incorporated herein by reference. Metal layer deposition techniques, such as those exemplified above, can be combined with photolithography techniques to create metal regions or patches on a surface. Exemplary methods for combining metal layer deposition techniques and photolithography techniques are provided in US Pat. No. 8,895,249 or US Pat App. Pub. No. 2014/0243224 A1, each of which is incorporated herein by reference.

Features can appear on a solid support as a grid of spots or patches. The features can be located in a repeating pattern or in an irregular, non-repeating pattern. Particularly useful repeating patterns are hexagonal patterns, rectilinear patterns, grid patterns, patterns having reflective symmetry, patterns having rotational symmetry, or the like. Asymmetric patterns can also be useful. The pitch

can be the same between different pairs of nearest neighbor features or the pitch can vary between different pairs of nearest neighbor features.

High density arrays are characterized as having average pitch of less than about 15 μm. Medium density arrays have average pitch of about 15 to 30 μm, while low density arrays have average pitch greater than 30 μm. An array useful in the invention can have average pitch that is less than 100 μm, 50 μm, 10 μm, 5 μm, 1 μm or 0.5 μm. The average pitch values and ranges set forth above or elsewhere herein are intended to be applicable to ordered arrays or random arrays.

In particular embodiments, features on a solid support can each have an area that is larger than about 100 nm^2, 250 nm^2, 500 nm^2, 1 μm^2, 2.5 μm^2, 5 μm^2, 10 μm^2, 100 μm^2, or 500 μm^2. Alternatively or additionally, features can each have an area that is smaller than about 1 mm^2, 500 μm^2, 100 μm^2, 25 μm2, 10 μm^2, 5 μm^2, 1 μm^2, 500 nm^2, or 100 nm^2. The above ranges can describe the apparent area of a bead or other particle on a solid support when viewed or imaged from above.

In particular embodiments, a solid support can include a collection of beads or other particles. The particles can be suspended in a solution or they can be located on the surface of a substrate. Examples of arrays having beads located on a surface include those wherein beads are located in wells such as a BeadChip array (Illumina Inc., San Diego CA), substrates used in sequencing platforms from 454 LifeSciences (a subsidiary of Roche, Basel Switzerland) or substrates used in sequencing platforms from Ion Torrent (a subsidiary of Life Technologies, Carlsbad California). Other solid supports having beads located on a surface are described in US Pat. Nos. 6,266,459; 6,355,431; 6,770,441; 6,859,570; 6,210,891; 6,258,568; or 6,274,320; US Pat. App. Publ. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; or 2010/0282617 A1 or PCT Publication No. WO 00/63437, each of which is incorporated herein by reference. Several of the above references describe methods for attaching nucleic acid probes to beads prior to loading the beads in or on a solid support. As such, the collection of beads can include different beads each having a unique probe attached. It will however, be understood that the beads can be made to include universal primers, and the beads can then be loaded onto an array, thereby forming universal arrays for use in a method set forth herein. As set forth previously herein, the solid supports typically used for bead arrays can be used without beads. For example, nucleic acids, such as probes or primers can be attached directly to the wells or to gel material in wells. Thus, the above

references are illustrative of materials, compositions or apparatus that can be modified for use in the methods and compositions set forth herein.

Accordingly, a solid support used in a method set forth herein can include an array of beads, wherein different nucleic acid probes are attached to different beads in the array. In this embodiment, each bead can be attached to a different nucleic acid probe and the beads can be randomly distributed on the solid support in order to effectively attach the different nucleic acid probes to the solid support. Optionally, the solid support can include wells having dimensions that accommodate no more than a single bead. In such a configuration, the beads may be attached to the wells due to forces resulting from the fit of the beads in the wells. It is also possible to use attachment chemistries or adhesives to hold the beads in the wells.

Nucleic acid probes that are attached to beads can include barcode sequences. A population of the beads can be configured such that each bead is attached to only one type of barcode and many different beads each with a different barcode are present in the population. In this embodiment, randomly distributing the beads to a solid support will result in randomly locating the nucleic acid probes (and their respective barcode sequences) on the solid support. In some cases there can be multiple beads with the same barcode sequence such that there is redundancy in the population. Randomly distributing a redundant population of beads on a solid support that has a capacity that is greater than the number of unique barcodes in the bead population will result in redundancy of barcodes on the solid support. Alternatively, the number of different barcodes in a population of beads can exceed the capacity of the solid support in order to produce an array that is not redundant with respect to the population of barcodes on the solid support. The capacity of the solid support will be determined in some embodiments by the number of features (e.g. single-bead occupancy wells) that attach or otherwise accommodate a bead.

A solid support can include, or can be made by the methods set forth herein to attach, a plurality of different nucleic acid probes. For example, a solid support can include at least 10, 100, 1×10^3, 1×10^4, 1×10^5, 1×10^6, 1×10^7, 1×10^8, 1×10^9 or more different probes. Alternatively or additionally, a solid support can include at most 1×10^9, 1×10^8, 1×10^7, 1×10^6, 1×10^5, 1×10^4, 1×10^3, 100, or fewer different probes. It will be understood that each of the different probes can be present in several copies, for example, when the probes have been amplified to form a cluster. Thus, the above ranges can describe the number of different nucleic

acid clusters on a solid support. It will also be understood that the above ranges can describe the number of different barcodes, target capture sequences, or other sequence elements set forth herein as being unique to particular nucleic acid probes. Alternatively or additionally, the ranges can describe the number of extended probes or modified probes created on a solid support using a method set forth herein.

Features, may be present on a solid support prior to contacting the solid support with nucleic acid probes. For example, in embodiments where probes are attached to a support via hybridization to primers, the primers can be attached at the features, whereas interstitial areas outside of the features substantially lack any of the primers. Nucleic acid probes can be captured at preformed features on a solid support, and optionally amplified on the solid support, using methods set forth in US Pat. No. 8,895,249, US Pat. No. 8,778,849, or US Pat App. Pub. No. 2014/0243224 A1, each of which is incorporated herein by reference. Alternatively, a solid support may have a lawn of primers or may otherwise lack features. In this case, a feature can be formed by virtue of attachment of a nucleic acid probe on the solid support. Optionally, the captured nucleic acid probe can be amplified on the solid support such that the resulting cluster becomes a feature. Although attachment is exemplified above as capture between a primer and a complementary portion of a probe, it will be understood that capture moieties other than primers can be present at pre-formed features or as a lawn. Other exemplary capture moieties include, but are not limited to, chemical moieties capable of reacting with a nucleic acid probe to create a covalent bond or receptors capable of biding non-covalently to a ligand on a nucleic acid probe.

A step of attaching nucleic acid probes to a solid support can be carried out by providing a fluid that contains a mixture of different nucleic acid probes and contacting this fluidic mixture with the solid support. The contact can result in the fluidic mixture being in contact with a surface to which many different nucleic acid probes from the fluidic mixture will attach. Thus, the probes have random access to the surface (whether the surface has pre-formed features configured to attach the probes or a uniform surface configured for attachment). Accordingly, the probes can be randomly located on the solid support.

The total number and variety of different probes that end up attached to a surface can be selected for a particular application or use. For example, in embodiments where a fluidic mixture of different nucleic acid probes is contacted

with a solid support for purposes of attaching the probes to the support, the number of different probe species can exceed the occupancy of the solid support for probes. Thus, the number and variety of different probes that attach to the solid support can be equivalent to the probe occupancy of the solid support. Alternatively, the number and variety of different probe species on the solid support can be less than the occupancy (i.e. there will be redundancy of probe species such that the solid support may contain multiple features having the same probe species). Such redundancy can be achieved, for example, by contacting the solid support with a fluidic mixture that contains a number and variety of probe species that is substantially lower than the probe occupancy of the solid support.

Attachment of the nucleic acid probes can be mediated by hybridization of the nucleic acid probes to complementary primers that are attached to the solid support, chemical bond formation between a reactive moiety on the nucleic acid probe and the solid support (examples are set forth in US Pat. No. 8,895,249, US Pat. No. 8,778,849, or US Pat App. Pub. No. 2014/0243224 A1, each of which is incorporated herein by reference), affinity interactions of a moiety on the nucleic acid probe with a solid support-bound moiety (e.g. between known receptor-ligand pairs such as streptavidinbiotin, antibody-epitope, lectin-carbohydrate and the like), physical interactions of the nucleic acid probes with the solid support (e.g. hydrogen bonding, ionic forces, van der Waals forces and the like), or other interactions known in the art to attach nucleic acids to surfaces.

In some embodiments, attachment of a nucleic acid probe is non-specific with regard to any sequence differences between the nucleic acid probe and other nucleic acid probes that are or will be attached to the solid support. For example, different probes can have a universal sequence that complements surface-attached primers or the different probes can have a common moiety that mediates attachment to the surface. Alternatively, each of the different probes (or a subpopulation of different probes) can have a unique sequence that complements a unique primer on the solid support or they can have a unique moiety that interacts with one or more different reactive moiety on the solid support. In such cases, the unique primers or unique moieties can, optionally, be attached at predefined locations in order to selectively capture particular probes, or particular types of probes, at the respective predefined locations.

One or more features on a solid support can each include a single molecule of a particular probe. The features can be configured, in some embodiments, to

accommodate no more than a single nucleic acid probe molecule. However, whether or not the feature can accommodate more than one nucleic acid probe molecule, the feature may nonetheless include no more than a single nucleic acid probe molecule. Alternatively, an individual feature can include a plurality of nucleic acid probe molecules, for example, an ensemble of nucleic acid probe molecules having the same sequence as each other. In particular embodiments, the ensemble can be produced by amplification from a single nucleic acid probe template to produce amplicons, for example, as a cluster attached to the surface.

A method set forth herein can use any of a variety of amplification techniques. Exemplary techniques that can be used include, but are not limited to, polymerase chain reaction (PCR), rolling circle amplification (RCA), multiple displacement amplification (MDA), or random prime amplification (RPA). In some embodiments the amplification can be carried out in solution, for example, when features of an array are capable of containing amplicons in a volume having a desired capacity. Preferably, an amplification technique used in a method of the present disclosure will be carried out on solid phase. For example, one or more primer species (e.g. universal primers for one or more universal primer binding site present in a nucleic acid probe) can be attached to a solid support. In PCR embodiments, one or both of the primers used for amplification can be attached to a solid support (e.g. via a gel). Formats that utilize two species of primers attached to a solid support are often referred to as bridge amplification because double stranded amplicons form a bridge-like structure between the two surface attached primers that flank the template sequence that has been copied. Exemplary reagents and conditions that can be used for bridge amplification are described, for example, in U.S. Pat. Nos. 5,641,658, 7,115,400, or 8,895,249; or U.S. Pat. Publ. Nos. 2002/0055100 A1, 2004/0096853 A1, 2004/0002090 A1, 2007/0128624 A1 or 2008/0009420 A1, each of which is incorporated herein by reference. Solid-phase PCR amplification can also be carried out with one of the amplification primers attached to a solid support and the second primer in solution. An exemplary format that uses a combination of a surface attached primer and soluble primer is the format used in emulsion PCR as described, for example, in Dressman et al., *Proc. Natl. Acad. Sci. USA* 100:8817-8822 (2003), WO 05/010145, or U.S. Pat. App. Publ. Nos. 2005/0130173 A1 or 2005/0064460 A1, each of which is incorporated herein by reference. Emulsion PCR is illustrative of the format and it will be

understood that for purposes of the methods set forth herein the use of an emulsion is optional and indeed for several embodiments an emulsion is not used.

RCA techniques can be modified for use in a method of the present disclosure. Exemplary components that can be used in an RCA reaction and principles by which RCA produces amplicons are described, for example, in Lizardi et al., *Nat. Genet.* 19:225-232 (1998) and US Pat. App. Publ. No. 2007/0099208 A1, each of which is incorporated herein by reference. Primers used for RCA can be in solution or attached to a solid support. The primers can be one or more of the universal primers described herein.

MDA techniques can be modified for use in a method of the present disclosure. Some basic principles and useful conditions for MDA are described, for example, in Dean et al., *Proc Natl. Acad. Sci. USA* 99:5261-66 (2002); Lage et al., *Genome Research* 13:294-307 (2003); Walker et al., *Molecular Methods for Virus Detection*, Academic Press, Inc., 1995; Walker et al., *Nucl. Acids Res.* 20:1691-96 (1992); US 5,455,166; US 5,130,238; and US 6,214,587, each of which is incorporated herein by reference. Primers used for MDA can be in solution or attached to a solid support at an amplification site. Again, the primers can be one or more of the universal primers described herein.

In particular embodiments a combination of the above-exemplified amplification techniques can be used. For example, RCA and MDA can be used in a combination wherein RCA is used to generate a concatameric amplicon in solution (e.g. using solution-phase primers). The amplicon can then be used as a template for MDA using primers that are attached to a solid support (e.g. universal primers). In this example, amplicons produced after the combined RCA and MDA steps will be attached to the solid support.

Nucleic acid probes that are used in a method set forth herein or present in an apparatus or composition of the present disclosure can include barcode sequences, and for embodiments that include a plurality of different nucleic acid probes, each of the probes can include a different barcode sequence from other probes in the plurality. Barcode sequences can be any of a variety of lengths. Longer sequences can generally accommodate a larger number and variety of barcodes for a population. Generally, all probes in a plurality will have the same length barcode (albeit with different sequences), but it is also possible to use different length barcodes for different probes. A barcode sequence can be at least 2, 4, 6, 8, 10, 12, 15, 20 or more nucleotides in length. Alternatively or additionally,

the length of the barcode sequence can be at most 20, 15, 12, 10, 8, 6, 4 or fewer nucleotides. Examples of barcode sequences that can be used are set forth, for example in, US Pat. App. Publ. No. 2014/0342921 A1 and US Pat. No. 8,460,865, each of which is incorporated herein by reference.

A method of the present disclosure can include a step of performing a nucleic acid detection reaction on a solid support to determine barcode sequences of nucleic acid probes that are located on the solid support. In many embodiments the probes are randomly located on the solid support and the nucleic acid detection reaction provides information to locate each of the different probes. Exemplary nucleic acid detection methods include, but are not limited to nucleic acid sequencing of a probe, hybridization of nucleic acids to a probe, ligation of nucleic acids that are hybridized to a probe, extension of nucleic acids that are hybridized to a probe, extension of a first nucleic acid that is hybridized to a probe followed by ligation of the extended nucleic acid to a second nucleic acid that is hybridized to the probe, or other methods known in the art such as those set forth in US Pat. No. 8,288,103 or 8,486,625, each of which is incorporated herein by reference.

Sequencing techniques, such as sequencing-by-synthesis (SBS) techniques, are a particularly useful method for determining barcode sequences. SBS can be carried out as follows. To initiate a first SBS cycle, one or more labeled nucleotides, DNA polymerase, SBS primers etc., can be contacted with one or more features on a solid support (e.g. feature(s) where nucleic acid probes are attached to the solid support). Those features where SBS primer extension causes a labeled nucleotide to be incorporated can be detected. Optionally, the nucleotides can include a reversible termination moiety that terminates further primer extension once a nucleotide has been added to the SBS primer. For example, a nucleotide analog having a reversible terminator moiety can be added to a primer such that subsequent extension cannot occur until a deblocking agent is delivered to remove the moiety. Thus, for embodiments that use reversible termination, a deblocking reagent can be delivered to the solid support (before or after detection occurs). Washes can be carried out between the various delivery steps. The cycle can then be repeated n times to extend the primer by n nucleotides, thereby detecting a sequence of length n. Exemplary SBS procedures, fluidic systems and detection platforms that can be readily adapted for use with a composition, apparatus or method of the present disclosure are described, for example, in Bentley et al., *Nature* 456:53-59 (2008), PCT Publ. Nos. WO 91/06678, WO 04/018497 or WO

07/123744; US Pat. Nos. 7,057,026, 7,329,492, 7,211,414, 7,315,019 or 7,405,281, and US Pat. App. Publ. No. 2008/0108082, each of which is incorporated herein by reference.

Other sequencing procedures that use cyclic reactions can be used, such as pyrosequencing. Pyrosequencing detects the release of inorganic pyrophosphate (PPi) as particular nucleotides are incorporated into a nascent nucleic acid strand (Ronaghi, et al., *Analytical Biochemistry* 242(1), 84-9 (1996); Ronaghi, *Genome Res.* 11(1), 3-11 (2001); Ronaghi et al. *Science* 281(5375), 363 (1998); or US Pat. Nos. 6,210,891, 6,258,568 or 6,274,320, each of which is incorporated herein by reference). In pyrosequencing, released PPi can be detected by being immediately converted to adenosine triphosphate (ATP) by ATP sulfurylase, and the level of ATP generated can be detected via luciferase-produced photons. Thus, the sequencing reaction can be monitored via a luminescence detection system. Excitation radiation sources used for fluorescence based detection systems are not necessary for pyrosequencing procedures. Useful fluidic systems, detectors and procedures that can be used for application of pyrosequencing to apparatus, compositions or methods of the present disclosure are described, for example, in PCT Pat. App. Publ. No. WO2012/058096, US Pat. App. Publ. No. 2005/0191698 A1, or US Pat. Nos. 7,595,883 or 7,244,559, each of which is incorporated herein by reference.

Sequencing-by-ligation reactions are also useful including, for example, those described in Shendure et al. *Science* 309:1728-1732 (2005); or US Pat. Nos. 5,599,675 or 5,750,341, each of which is incorporated herein by reference. Some embodiments can include sequencing-by-hybridization procedures as described, for example, in Bains et al., *Journal of Theoretical Biology* 135(3), 303-7 (1988); Drmanac et al., *Nature Biotechnology* 16, 54-58 (1998); Fodor et al., *Science* 251(4995), 767-773 (1995); or PCT Pat. App. Publ. No. WO 1989/10977, each of which is incorporated herein by reference. In both sequencing-by-ligation and sequencing-by-hybridization procedures, target nucleic acids (or amplicons thereof) that are present at sites of an array are subjected to repeated cycles of oligonucleotide delivery and detection. Compositions, apparatus or methods set forth herein or in references cited herein can be readily adapted for sequencing-by-ligation or sequencing-by-hybridization procedures. Typically, the oligonucleotides are fluorescently labeled and can be detected using fluorescence detectors similar

to those described with regard to SBS procedures herein or in references cited herein.

Some sequencing embodiments can utilize methods involving the real-time monitoring of DNA polymerase activity. For example, nucleotide incorporations can be detected through fluorescence resonance energy transfer (FRET) interactions between a fluorophore-bearing polymerase and γ-phosphate-labeled nucleotides, or with zeromode waveguides (ZMWs). Techniques and reagents for FRET-based sequencing are described, for example, in Levene et al. *Science* 299, 682–686 (2003); Lundquist et al. *Opt. Lett.* 33, 1026–1028 (2008); Korlach et al. *Proc. Natl. Acad. Sci. USA* 105, 1176–1181 (2008), each of which is incorporated herein by reference.

Some sequencing embodiments include detection of a proton released upon incorporation of a nucleotide into an extension product. For example, sequencing based on detection of released protons can use an electrical detector and associated techniques that are commercially available from Ion Torrent (Guilford, CT, a Life Technologies and Thermo Fisher subsidiary) or sequencing methods and systems described in US Pat app. Publ. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; or US 2010/0282617 A1, each of which is incorporated herein by reference.

Nucleic acid hybridization techniques are also useful method for determining barcode sequences. In some cases combinatorial hybridization methods can be used such as those used for decoding of multiplex bead arrays (see e.g. US Pat. No. 8,460,865, which is incorporated herein by reference). Such methods utilize labelled nucleic acid decoder probes that are complementary to at least a portion of a barcode sequence. A hybridization reaction can be carried out using decoder probes having known labels such that the location where the labels end up on the solid support identifies the nucleic acid probes according to rules of nucleic acid complementarity. In some cases, pools of many different probes with distinguishable labels are used, thereby allowing a multiplex decoding operation. The number of different barcodes determined in a decoding operation can exceed the number of labels used for the decoding operation. For example, decoding can be carried out in several stages where each stage constitutes hybridization with a different pool of decoder probes. The same decoder probes can be present in different pools but the label that is present on each decoder probe can differ from pool to pool (i.e. each decoder probe is in a different “state” when in different pools).

Various combinations of these states and stages can be used to expand the number of barcodes that can be decoded well beyond the number of distinct labels available for decoding. Such combinatorial methods are set forth in further detail in US Pat. No. 8,460,865 or Gunderson et al., *Genome Research* 14:870-877 (2004), each of which is incorporated herein by reference.

A method of the present disclosure can include a step of contacting a biological specimen with a solid support that has nucleic acid probes attached thereto. In some embodiments the nucleic acid probes are randomly located on the solid support. The identity and location of the nucleic acid probes may have been decoded prior to contacting the biological specimen with the solid support. Alternatively, the identity and location of the nucleic acid probes can be determined after contacting the solid support with the biological specimen.

In some embodiments the biological specimen is one or more cells. The cell(s) can be individual and free from any tissue or multicellular structure at the time contact is made with the solid support. For example, the cell(s) can be present in a fluid (e.g. when a plurality of different cells are present the fluid can be a fluidic mixture of the different cells) and the fluid can be contacted with the solid support to which the different probes are attached. Any of a variety of cells can be used including, for example, those from a prokaryote, archae or eukaryote. One or more cells used in a method, composition or apparatus of the present disclosure can be a single celled organisms or from a multicellular organism. Exemplary organisms from which one or more cell can be obtained include, but are not limited to a mammal, plant, algae, nematode, insect, fish, reptile, amphibian, fungi or Plasmodium falciparum. Exemplary species are set forth previously herein or known in the art.

Embodiments of the present disclosure can also use one or more subcellular components as a biological specimen. For example a fluidic mixture can include one or more nuclei, golgi apparatus, mitochondria, chloroplasts, membrane fractions, vesicles, endoplasmic reticulum, or other components known in the art. Other useful types of biological specimens are one or more viruses or a viroids.

It will be understood that a biological specimen can be a homogeneous culture or population of the above cells, subcellular components, viruses or viroids. Alternatively the biological specimen can be a non-homogenous collection of cells, subcellular components, viruses or viroids, for example, derived from several different organisms in a community or ecosystem. An exemplary community is the

collection of bacteria present in the digestive system, lung or other organ of a multicellular organism such as a mammal.

One or more cells, subcellular components, viruses or viroids that are contacted with a solid support in a method set forth herein can be attached to the solid support. Attachment can be achieved using methods known in the art such as those exemplified herein with respect to attachment of nucleic acids to a solid support. In some embodiments, attachment is selective for specific types of cells, subcellular components, viruses or viroids. For example, the solid support can include antibodies or other receptors that are selective for epitopes or ligands present on one or a subset of different cells, subcellular components, viruses or viroids present in a fluidic mixture. In other embodiments, the attachment of cells, subcellular components, viruses or viroids can be mediated by non-selective moieties such as chemical moieties that are broadly reactive.

In particular embodiments, one or more cells, subcellular components, viruses or viroids that have been contacted with a solid support can be lysed to release target nucleic acids. Lysis can be carried out using methods known in the art such as those that employ one or more of chemical treatment, enzymatic treatment, electroporation, heat, hypotonic treatment, sonication or the like. Exemplary lysis techniques are set forth in Sambrook et al., *Molecular Cloning: A Laboratory Manual, Third Ed.*, Cold Spring Harbor Laboratory, New York (2001) and in Ansubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, Md. (1999).

In some embodiments the biological specimen is a tissue section. The tissue can be derived from a multicellular organism such as those exemplified above in regard to cells. A tissue section can be contacted with a solid support, for example, by laying the tissue on the surface of the solid support. The tissue can be freshly excised from an organism or it may have been previously preserved for example by freezing, embedding in a material such as paraffin (e.g. formalin fixed paraffin embedded samples), formalin fixation, infiltration, dehydration or the like. Optionally, a tissue section can be attached to a solid support, for example, using techniques and compositions exemplified herein with regard to attaching nucleic acids, cells, viruses, beads or the like to a solid support. As a further option, a tissue can be permeabilized and the cells of the tissue lysed when the tissue is in contact with a solid support. Any of a variety of treatments can be used such as those set forth above in regard to lysing cells. Target nucleic acids that are released

from a tissue that is permeabilized can be captured by nucleic acid probes on the surface.

A tissue can be prepared in any convenient or desired way for its use in a method, composition or apparatus herein. Fresh, frozen, fixed or unfixed tissues can be used. A tissue can be fixed or embedded using methods described herein or known in the art.

A tissue sample for use herein, can be fixed by deep freezing at temperature suitable to maintain or preserve the integrity of the tissue structure, e.g. less than -20° C. In another example, a tissue can be prepared using formalin-fixation and paraffin embedding (FFPE) methods which are known in the art. Other fixatives and/or embedding materials can be used as desired. A fixed or embedded tissue sample can be sectioned, i.e. thinly sliced, using known methods. For example, a tissue sample can be sectioned using a chilled microtome or cryostat, set at a temperature suitable to maintain both the structural integrity of the tissue sample and the chemical properties of the nucleic acids in the sample.

In some embodiments, a tissue sample will be treated to remove embedding material (e.g. to remove paraffin or formalin) from the sample prior to release, capture or modification of nucleic acids. This can be achieved by contacting the sample with an appropriate solvent (e.g. xylene and ethanol washes). Treatment can occur prior to contacting the tissue sample with a solid support set forth herein or the treatment can occur while the tissue sample is on the solid support. Exemplary methods for manipulating tissues for use with solid supports to which nucleic acids are attached are set forth in US Pat. App. Publ. No. 2014/0066318 A1, which is incorporated herein by reference.

The thickness of a tissue sample or other biological specimen that is contacted with a solid support in a method, composition or apparatus set forth herein can be any suitable thickness desired. In representative embodiments, the thickness will be at least 0.1 µm, 0.25 µm, 0.5 µm, 0.75 µm, 1 µm, 5 µm, 10 µm, 50 µm, 100 µm or thicker. Alternatively or additionally, the thickness of a biological specimen that is contacted with a solid support will be no more than 100 µm, 50 µm, 10 µm, 5 µm, 1 µm, 0.5 µm, 0.25 µm, 0.1 µm or thinner.

A particularly relevant source for a biological specimen is a human being. The specimen can be derived from an organ, including for example, an organ of the musculoskeletal system such as muscle, bone, tendon or ligament; an organ of the digestive system such as salivary gland, pharynx, esophagus, stomach, small

intestine, large intestine, liver, gallbladder or pancreas; an organ of the respiratory system such as larynx, trachea, bronchi, lungs or diaphragm; an organ of the urinary system such as kidney, ureter, bladder or urethra; a reproductive organ such as ovary, fallopian tube, uterus, vagina, placenta, testicle, epididymis, vas deferens, seminal vesicle, prostate, penis or scrotum; an organ of the endocrine system such as pituitary gland, pineal gland, thyroid gland, parathyroid gland, or adrenal gland; an organ of the circulatory system such as heart, artery, vein or capillary; an organ of the lymphatic system such as lymphatic vessel, lymph node, bone marrow, thymus or spleen; an organ of the central nervous system such as brain, brainstem, cerebellum, spinal cord, cranial nerve, or spinal nerve; a sensory organ such as eye, ear, nose, or tongue; or an organ of the integument such as skin, subcutaneous tissue or mammary gland. In some embodiments, a biological specimen is obtained from a bodily fluid or excreta such as blood, lymph, tears, sweat, saliva, semen, vaginal secretion, ear wax, fecal matter or urine.

A specimen from a human can be considered (or suspected) healthy or diseased when used. In some cases, two specimens can be used: a first being considered diseased and a second being considered as healthy (e.g. for use as a healthy control). Any of a variety of conditions can be evaluated, including but not limited to, an autoimmune disease, cancer, cystic fibrosis, aneuploidy, pathogenic infection, psychological condition, hepatitis, diabetes, sexually transmitted disease, heart disease, stroke, cardiovascular disease, multiple sclerosis or muscular dystrophy. Particularly relevant conditions are genetic conditions or conditions associated with pathogens having identifiable genetic signatures.

As set forth above, a flow cell provides a convenient apparatus for use in a method set forth herein. For example, a flow cell is a convenient apparatus for housing a solid support that will be treated with multiple fluidic reagents such as the repeated fluidic deliveries used for some nucleic acid sequencing protocols or some nucleic acid hybridization protocols. In some embodiments, a biological specimen can be delivered to a solid support in a flow cell, for example, when a fluidic mixture of cells, subcellular components, viruses or viroids is delivered to the solid support. In some embodiments it may be preferable to open a flow cell to expose a solid support inside or to remove the solid support from the flow cell in order to allow convenient delivery of a biological specimen to the solid support. For example, opening the flow cell or removing the solid support can allow a user or robotic device to lay a tissue section on the solid support. The opening of a flow cell or

removal of a solid support from a flow cell can be temporary. Thus, the flow cell can subsequently be closed or the solid support returned to the flow cell to proceed with one or more subsequent steps of a method set forth herein.

In some embodiments, a flow cell can have a construction that allows it to be opened or taken apart. For example the flow cell can be in a closed state while performing a sequencing reaction, for example to decode barcodes. Then the flow cell can be taken apart so that tissue can be placed on the flow cell surface. The flow cell can be held together by adhesive such that one or more surface can be removed to open it. For example, a flow cell can have a spacer with adhesive surfaces on the top or bottom (akin to single-sided or double-sided sticky tape) and this spacer can occur between two solid supports. One or both of the solid supports can be configured to attach nucleic acids and support a biological specimen as set forth herein. The spacer can have open regions (e.g. created by laser cutting of the spacer material) that create fluidic channels bound by the two solid supports and the spacer. Thus, one or both of the solid supports can be non-permanently adhered to the spacer to allow one or both of them to be removed to allow access to the surface when placing a tissue or other specimen thereon.

A nucleic acid probe used in a composition, apparatus or method set forth herein can include a target capture moiety. In particular embodiments, the target capture moiety is a target capture sequence. The target capture sequence is generally complementary to a target sequence such that target capture occurs by formation of a probe-target hybrid complex. A target capture sequence can be any of a variety of lengths including, for example, lengths exemplified above in the context of barcode sequences.

In multiplex embodiments, a plurality of different nucleic acid probes can include different target capture sequences that hybridize to different target nucleic acid sequences from a biological specimen. Different target capture sequences can be used to selectively bind to one or more desired target nucleic acids from a biological specimen. In some cases, the different nucleic acid probes can include a target capture sequence that is common to all or a subset of the probes on a solid support. For example, the nucleic acid probes on a solid support can have a poly A or poly T sequence. Such probes or amplicons thereof can hybridize to mRNA molecules, cDNA molecules or amplicons thereof that have poly A or poly T tails. Although the mRNA or cDNA species will have different target sequences, capture will be mediated by the common poly A or poly T sequence regions.

Any of a variety of target nucleic acids can be captured and analyzed in a method set forth herein including, but not limited to, messenger RNA (mRNA), copy DNA (cDNA), genomic DNA (gDNA), ribosomal RNA (rRNA) or transfer RNA (tRNA). Particular target sequences can be selected from databases and appropriate capture sequences designed using techniques and databases known in the art.

Other target capture moieties that are useful include, for example, the moieties set forth herein as useful for attaching nucleic acid probes to a solid support.

A method set forth herein can include a step of hybridizing nucleic acid probes, that are on a solid support, to target nucleic acids that are from portions of the biological specimen that are proximal to the probes. Generally, a target nucleic acid will diffuse from a region of the biological specimen to an area of the solid support that is in proximity with that region of the specimen. Here the target nucleic acid will interact with nucleic acid probes that are proximal to the region of the specimen from which the target nucleic acid was released. A target-probe hybrid complex can form where the target nucleic acid encounters a complementary target capture sequence on a nucleic acid probe. The location of the target-probe hybrid complex will generally correlate with the region of the biological specimen from where the target nucleic acid was derived. In multiplex embodiments, the solid support will include a plurality of nucleic acid probes, the biological specimen will release a plurality of target nucleic acids and a plurality of target-probe hybrids will be formed on the solid support. The sequences of the target nucleic acids and their locations on the support will provide spatial information about the nucleic acid content of the biological specimen. Although the example above is described in the context of target nucleic acids that are released from a biological specimen, it will be understood that the target nucleic acids need not be released. Rather, the target nucleic acids may remain in contact with the biological specimen, for example, when they are attached to an exposed surface of the biological specimen in a way that the target nucleic acids can also bind to appropriate nucleic acid probes on the solid support.

A method of the present disclosure can include a step of extending solid support-attached probes to which target nucleic acids are hybridized. In embodiments where the probes include barcode sequences, the resulting extended probes will include the barcode sequences and sequences from the target

nucleic acids (albeit in complementary form). The extended probes are thus spatially tagged versions of the target nucleic acids from the biological specimen. The sequences of the extended probes identify what nucleic acids are in the biological specimen and where in the biological specimen the target nucleic acids are located. It will be understood that other sequence elements that are present in the nucleic acid probes can also be included in the extended probes. Such elements include, for example, primer binding sites, cleavage sites, other tag sequences (e.g. sample identification tags), capture sequences, recognition sites for nucleic acid binding proteins or nucleic acid enzymes, or the like.

Extension of probes can be carried out using methods exemplified herein or otherwise known in the art for amplification of nucleic acids or sequencing of nucleic acids. In particular embodiments one or more nucleotides can be added to the 3' end of a nucleic acid, for example, via polymerase catalysis (e.g. DNA polymerase, RNA polymerase or reverse transcriptase). Chemical or enzymatic methods can be used to add one or more nucleotide to the 3' or 5' end of a nucleic acid. One or more oligonucleotides can be added to the 3' or 5' end of a nucleic acid, for example, via chemical or enzymatic (e.g. ligase catalysis) methods. A nucleic acid can be extended in a template directed manner, whereby the product of extension is complementary to a template nucleic acid that is hybridized to the nucleic acid that is extended. In some embodiments, a DNA primer is extended by a reverse transcriptase using an RNA template, thereby producing a cDNA. Thus, an extended probe made in a method set forth herein can be a reverse transcribed DNA molecule. Exemplary methods for extending nucleic acids are set forth in US Pat. App. Publ. No. US 2005/0037393 A1 or US Pat. No. 8,288,103 or 8,486,625, each of which is incorporated herein by reference.

All or part of a target nucleic acid that is hybridized to a nucleic acid probe can be copied by extension. For example, an extended probe can include at least, 1, 2, 5,10, 25, 50, 100, 200, 500, 1000 or more nucleotides that are copied from a target nucleic acid. The length of the extension product can be controlled, for example, using reversibly terminated nucleotides in the extension reaction and running a limited number of extension cycles. The cycles can be run as exemplified for SBS techniques and the use of labeled nucleotides is not necessary. Accordingly, an extended probe produced in a method set forth herein can include no more than 1000, 500, 200, 100, 50, 25, 10, 5, 2 or 1 nucleotides that are copied

from a target nucleic acid. Of course extended probes can be any length within or outside of the ranges set forth above.

Although the methods of the present disclosure are exemplified by an embodiment where probes that are hybridized to target nucleic acids are extended to copy at least a portion of the target nucleic acid, it will be understood that the probes can be modified in alternative ways. The probes that are hybridized to target nucleic acids can be subjected to a reaction that creates a target specific modification of the probe. A target specific modification will result only when the probe interacts with a target nucleic acid, for example, via complementary based hybridization. In many embodiments, the target specific modification will be specific to the sequence of the particular target nucleic acid that interacts with the probe. Examples of useful target specific modifications, include but are not limited to, insertion or addition of a sequence by ligation or transposition (see, for example, US Pat. App. Publ. No. 2010/0120098 A1, incorporated herein by reference), chemical modifications such as psoralen crosslinking or addition of a detectable tag moiety, modifications by nucleic acid enzymes, ligation of a hairpin linker, or other modifications set forth in the nucleic acid assays of US Pat. App. Publ. No. US 2005/0037393 A1 or US Pat. No. 8,288,103 or 8,486,625, each of which is incorporated herein by reference.

It will be understood that probes used in a method, composition or apparatus set forth herein need not be nucleic acids. Other molecules can be used such as proteins, carbohydrates, small molecules, particles or the like. Probes can be a combination of a nucleic acid component (e.g. having a barcode, primer binding site, cleavage site and/or other sequence element set forth herein) and another moiety (e.g. a moiety that captures or modifies a target nucleic acid).

A method set forth herein can further include a step of acquiring an image of a biological specimen that is in contact with a solid support. The solid support can be in any of a variety of states set forth herein. For example, the solid support can include attached nucleic acid probes or clusters derived from attached nucleic acid probes. Alternatively, the solid support may not include nucleic acid probes, instead being in a state that precedes attachment of nucleic acid probes or in a state that follows removal of nucleic acid probes from the solid support. Accordingly, an image can be obtained at any of a variety of points in a method set forth herein.

An image can be obtained using detection devices known in the art. Examples include microscopes configured for light, bright field, dark field, phase

contrast, fluorescence, reflection, interference, or confocal imaging. A biological specimen can be stained prior to imaging to provide contrast between different regions or cells. In some embodiments, more than one stain can be used to image different aspects of the specimen (e.g. different regions of a tissue, different cells, specific subcellular components or the like). In other embodiments, a biological specimen can be imaged without staining.

In particular embodiments, a fluorescence microscope (e.g. a confocal fluorescent microscope) can be used to detect a biological specimen that is fluorescent, for example, by virtue of a fluorescent label. Fluorescent specimens can also be imaged using a nucleic acid sequencing device having optics for fluorescent detection such as a Genome Analyzer®, MiSeq®, NextSeq® or HiSeq® platform device commercialized by Illumina, Inc. (San Diego, CA); or a SOLiD™ sequencing platform commercialized by Life Technologies (Carlsbad, CA). Other imaging optics that can be used include those that are found in the detection devices described in Bentley et al., *Nature* 456:53-59 (2008), PCT Publ. Nos. WO 91/06678, WO 04/018497 or WO 07/123744; US Pat. Nos. 7,057,026, 7,329,492, 7,211,414, 7,315,019 or 7,405,281, and US Pat. App. Publ. No. 2008/0108082, each of which is incorporated herein by reference.

An image of a biological specimen can be obtained at a desired resolution, for example, to distinguish tissues, cells or subcellular components. Accordingly, the resolution can be sufficient to distinguish components of a biological specimen that are separated by at least 0.5 µm, 1 µm, 5 µm, 10 µm, 50 µm, 100 µm, 500 µm, 1 mm or more. Alternatively or additionally, the resolution can be set to distinguish components of a biological specimen that are separated by at least 1 mm, 500 µm, 100 µm, 50 µm, 10 µm, 5 µm, 1 µm, 0.5 µm or less.

A method set forth herein can include a step of correlating locations in an image of a biological specimen with barcode sequences of nucleic acid probes that are attached to a surface to which the biological specimen is, was or will be contacted. Accordingly, characteristics of the biological specimen that are identifiable in the image can be correlated with the nucleic acids that are found to be present in their proximity. Any of a variety of morphological characteristics can be used in such a correlation, including for example, cell shape, cell size, tissue shape, staining patterns, presence of particular proteins (e.g. as detected by immunohistochemical stains) or other characteristics that are routinely evaluated in pathology or research applications. Accordingly, the biological state of a tissue or its

components as determined by visual observation can be correlated with molecular biological characteristics as determined by spatially resolved nucleic acid analysis.

A solid support upon which a biological specimen is imaged can include fiducial markers to facilitate determination of the orientation of the specimen or the image thereof in relation to probes that are attached to the solid support. Exemplary fiducials include, but are not limited to beads (with or without fluorescent moieties or moieties such as nucleic acids to which labeled probes can be bound), fluorescent molecules attached at known or determinable features, or structures that combine morphological shapes with fluorescent moieties. Exemplary fiducials are set forth in US Pat. App. Publ. No. 2002/0150909 A1 or US Pat. App. Ser. No. 14/530,299, each of which is incorporated herein by reference. One or more fiducials are preferably visible while obtaining an image of a biological specimen. Preferably, the solid support includes at least 2, 3, 4, 5, 10, 25, 50, 100 or more fiducial markers. The fiducials can be provided in a pattern, for example, along an outer edge of a solid support or perimeter of a location where a biological specimen resides. In a preferred embodiment, one or more fiducials are detected using the same imaging conditions used to visualize a biological specimen. However if desired separate images can be obtained (e.g. one image of the biological specimen and another image of the fiducials) and the images can be aligned to each other.

Optionally, a biological specimen, can be removed from a solid support after an image has been obtained and after target nucleic acids have been captured by nucleic acid probes on the solid support. Thus, a method of the present disclosure can include a step of washing a solid support to remove cells, tissue or other materials from a biological specimen. Removal of the specimen can be performed using any suitable technique and will be dependent on the tissue sample. In some cases, the solid support can be washed with water. The water can contain various additives, such as surfactants (e.g. detergents), enzymes (e.g. proteases and collagenases), cleavage reagents, or the like, to facilitate removal of the specimen. In some embodiments, the solid support is treated with a solution comprising a proteinase enzyme. Alternatively or additionally, the solution can include cellulase, hemicelluase or chitinase enzymes (e.g. if desiring to remove a tissue sample from a plant or fungal source). In some cases, the temperature of a wash solution will be at least 30°C, 35°C, 50°C, 60°C or 90°C. Conditions can be selected for removal of a biological specimen while not denaturing hybrid complexes formed between target nucleic acids and solid support-attached nucleic acid probes.

A method of the present disclosure can further include a step of removing one or more extended probes from a solid support. In particular embodiments, the probes will have included a cleavage site such that the product of extending the probes will also include the cleavage site. Alternatively, a cleavage site can be introduced into a probe during a modification step. For example a cleavage site can be introduced into an extended probe during the extension step.

Exemplary cleavage sites include, but are not limited to, moieties that are susceptible to a chemical, enzymatic or physical process that results in bond breakage. For example, the location can be a nucleotide sequence that is recognized by an endonuclease. Suitable endonucleases and their recognition sequences are well known in the art and in many cases are even commercially available (e.g. from New England Biolabs, Beverley MA; ThermoFisher, Waltham, MA or Sigma Aldrich, St. Louis MO). A particularly useful endonuclease will break a bond in a nucleic acid strand at a site that is 3'-remote to its binding site in the nucleic acid, examples of which include Type II or Type IIs restriction endonucleases. In some embodiments an endonuclease will cut only one strand in a duplex nucleic acid (e.g. a nicking enzyme). Examples of endonucleases that cleave only one strand include Nt.BstNBI and Nt.AlwI.

In some embodiments, a cleavage site is an abasic site or a nucleotide that has a base that is susceptible to being removed to create an abasic site. Examples of nucleotides that are susceptible to being removed to form an abasic site include uracil and 8-oxo-guanine. Abasic sites can be created by hydrolysis of nucleotide residues using chemical or enzymatic reagents. Once formed, abasic sites may be cleaved (e.g. by treatment with an endonuclease or other single-stranded cleaving enzyme, exposure to heat or alkali), providing a means for site-specific cleavage of a nucleic acid. An abasic site may be created at a uracil nucleotide on one strand of a nucleic acid. The enzyme uracil DNA glycosylase (UDG) may be used to remove the uracil base, generating an abasic site on the strand. The nucleic acid strand that has the abasic site may then be cleaved at the abasic site by treatment with endonuclease (e.g. EndoIV endonuclease, AP lyase, FPG glycosylase/AP lyase, EndoVIII glycosylase/AP lyase), heat or alkali. In a particular embodiment, the USER™ reagent available from New England Biolabs is used for the creation of a single nucleotide gap at a uracil base in a nucleic acid.

Abasic sites may also be generated at non-natural/modified deoxyribonucleotides other than uracil and cleaved in an analogous manner by

treatment with endonuclease, heat or alkali. For example, 8-oxo-guanine can be converted to an abasic site by exposure to FPG glycosylase. Deoxyinosine can be converted to an abasic site by exposure to AlkA glycosylase. The abasic sites thus generated may then be cleaved, typically by treatment with a suitable endonuclease (e.g. EndoIV or AP lyase).

Other examples of cleavage sites and methods that can be used to cleave nucleic acids are set forth, for example, in US Pat. No. 7,960,120, which is incorporated herein by reference.

Modified nucleic acid probes (e.g. extended nucleic acid probes) that are released from a solid support can be pooled to form a fluidic mixture. The mixture can include, for example, at least 10, 100, 1 x 10^3, 1 x 10^4, 1 x 10^5, 1 x 10^6, 1 x 10^7, 1 x 10^8, 1 x 10^9 or more different modified probes. Alternatively or additionally, a fluidic mixture can include at most 1 x 10^9, 1 x 10^8, 1 x 10^7, 1 x 10^6, 1 x 10^5, 1 x 10^4, 1 x 10^3, 100, 10 or fewer different modified probes. The fluidic mixture can be manipulated to allow detection of the modified nucleic acid probes. For example, the modified nucleic acid probes can be separated spatially on a second solid support (i.e. different from the solid support from which the nucleic acid probes were released after having been contacted with a biological specimen and modified), or the probes can be separated temporally in a fluid stream.

Modified nucleic acid probes (e.g. extended nucleic acid probes) can be separated on a solid support in a capture or detection method commonly employed for microarray-based techniques or nucleic acid sequencing techniques such as those set forth previously herein. For example, modified probes can be attached to a microarray by hybridization to complementary nucleic acids. The modified probes can be attached to beads or to a flow cell surface and optionally amplified as is carried out in many nucleic acid sequencing platforms. Modified probes can be separated in a fluid stream using a microfluidic device, droplet manipulation device, or flow cytometer. Typically, detection is carried out on these separation devices, but detection is not necessary in all embodiments.

A particularly useful droplet manipulation device is a droplet actuator as described for example in US Pat. No. 8,637,242, US Pat. No. 6,911,132, entitled "Apparatus for Manipulating Droplets by Electrowetting-Based Techniques," issued on June 28, 2005; Pamula et al., U.S. Patent Pub. No. 20060194331, entitled "Apparatuses and Methods for Manipulating Droplets on a Printed Circuit Board," published on August 31, 2006; Pollack et al., International Patent Pub. No.

WO/2007/120241, entitled "Droplet-Based Biochemistry," published on October 25, 2007; Shenderov, U.S. Patent No. 6,773,566, entitled "Electrostatic Actuators for Microfluidics and Methods for Using Same," issued on August 10, 2004; Shenderov, U.S. Patent No. 25 6,565,727, entitled "Actuators for Microfluidics Without Moving Parts," issued on May 20, 2003; Kim et al., U.S. Patent Pub. No. 20030205632, entitled "Electrowettingdriven Micropumping," published on November 6, 2003; Kim et al., U.S. Patent Pub. No. 20060164490, entitled "Method and Apparatus for Promoting the Complete Transfer of Liquid Drops from a Nozzle," published on July 27, 2006; Kim et al., U.S. Patent Pub. No. 20070023292, entitled "Small Object Moving on Printed Circuit Board," published on February 1, 2007; Shah et al., U.S. Patent Pub. No. 20090283407, entitled "Method for Using Magnetic Particles in Droplet Microfluidics," published on November 19, 2009; Kim et al., U.S. Patent Pub. No. 20100096266, entitled "Method and Apparatus for Real-time Feedback Control of Electrical Manipulation of Droplets on Chip," published on April 22, 2010; Velev, U.S. Patent No. 7,547,380, entitled "Droplet Transportation Devices and Methods Having a Fluid Surface," issued on June 16, 2009; Sterling et al., U.S. Patent No. 7,163,612, entitled "Method, Apparatus and Article for Microfluidic Control via Electrowetting, for Chemical, Biochemical and Biological Assays and the Like," issued on January 16, 2007; Becker et al., U.S. Patent No. 7,641,779, entitled "Method and Apparatus for Programmable Fluidic Processing," issued on January 5, 2010; Becker et al., U.S. Patent No. 6,977,033, entitled "Method and Apparatus for Programmable Fluidic Processing," issued on December 20, 2005; Decre et al., U.S. Patent No. 7,328,979, entitled "System for Manipulation of a Body of Fluid," issued on February 12, 2008; Yamakawa et al., U.S. Patent Pub. No. 15 20060039823, entitled "Chemical Analysis Apparatus," published on February 23, 2006; Wu, U.S. Patent Pub. No. 20110048951, entitled "Digital Microfluidics Based Apparatus for Heat-exchanging Chemical Processes," published on March 3, 2011; Fouillet et al., U.S. Patent Pub. No. 20090192044, entitled "Electrode Addressing Method," published on July 30, 2009; Fouillet et al., U.S. Patent No. 7,052,244, entitled "Device for Displacement of Small Liquid Volumes Along a Micro-catenary Line by Electrostatic Forces," issued on May 30, 2006; Marchand et al., U.S. Patent Pub. No. 20080124252, entitled "Droplet Microreactor," published on May 29, 2008; Adachi et al., U.S. Patent Pub. No. 20090321262, entitled "Liquid Transfer Device," published on December 31, 2009; Roux et al., U.S. Patent Pub. No. 20050179746, entitled "Device for Controlling the

Displacement of a Drop Between Two or Several Solid Substrates," published on August 18, 2005; and Dhindsa et al., "Virtual Electrowetting Channels: Electronic Liquid Transport with Continuous Channel Functionality," Lab Chip,10:832–836 (2010), each of which is incorporated herein by reference.

Modified probes (e.g. extended nucleic acid probes) can be detected, for example, following separation from a fluidic mixture using methods set forth above or known in the art. In particular embodiments, modified probes that are separated on a second solid support (i.e. a solid support that is different from the first solid support where contact was made between probes and biological specimen) can be detected using microarray-based techniques or nucleic acid sequencing techniques such as those set forth previously herein. Probes that are separated in a fluid stream can be detected using optical, electrical or other detectors that are outfitted in known microfluidic devices, droplet manipulation devices, or flow cytometers. A detection method can be used to determine target nucleic acid sequences, barcode sequences or other sequence regions of extended probes.

Several embodiments have been exemplified with regard to removing modified probes from the solid support where the probes were produced. However, it will be understood that probes on a solid support can be contacted with a biological specimen, modified on the solid support in the presence of target nucleic acids from the specimen and then the modified probes can be detected on the solid support. In such an embodiment, the biological specimen can be removed from the solid support prior to the detection step.

In particular embodiments the present disclosure provides a method for spatially tagging nucleic acids of a biological specimen that includes the steps of (a) providing a plurality of nucleic acid primers attached to a solid support, wherein the nucleic acid primers in the plurality include a universal primer sequence that is common to the nucleic acid primers in the plurality; (b) binding a population of nucleic acid probes to the plurality of nucleic acid primers, wherein the nucleic acid probes include a universal primer binding sequence that hybridizes to the universal primer sequence, a target capture sequence and a barcode sequence that differs from barcode sequences of other nucleic acid probes in the population, thereby attaching the different nucleic acid probes at randomly located positions on the solid support; (c) amplifying the different nucleic acid probes by extension of the nucleic acid primers, thereby producing nucleic acid clusters having copies of the barcode sequence and target capture sequence at the randomly located positions on the

solid support; (d) performing a sequencing reaction to determine the barcode sequences at the randomly located positions on the solid support; (e) contacting a biological specimen with the nucleic acid clusters on the solid support; (f) hybridizing the target capture sequences of the clusters to target nucleic acids from portions of the biological specimen that are proximal to the clusters; and (g) extending the target capture sequences to produce extended probes that include sequences from the target nucleic acids and the copies of the barcode sequences, thereby tagging the nucleic acids of the biological specimen.

As exemplified previously herein, a plurality of nucleic acid primers can be attached to a solid support, wherein the nucleic acid primers in the plurality include a universal primer sequence that is common to the nucleic acid primers in the plurality. In this embodiment, a second plurality of nucleic acid primers can be attached to the solid support, and the nucleic acid primers in the second plurality can have a second universal primer sequence that is common to the nucleic acid primers in the second plurality. In this embodiment, a plurality of different nucleic acid probes that is contacted with the support can include a universal primer binding sequence that hybridizes to the universal primer on the solid support, as set forth above, and the different nucleic acid probes can also include a second universal primer binding sequence that hybridizes to the second universal primer sequence. This configuration of universal primers and universal primer binding sites can be particularly useful for amplifying the different nucleic acid probes via bridge amplification, wherein the nucleic acid primers in the first and second plurality are extended.

Typically, when a nucleic acid probe contains first and second universal primer binding sites, they will be located at the ends of the probe. In some embodiments it may be desirable to remove at least one of the primer binding sites from the nucleic acid probe or from amplicons produced from the probe. Accordingly, the nucleic acid probes can optionally include a cleavage site between the target capture sequence and one of the universal primer binding sequence. In this case, a cleavage reaction can be performed to separate the universal primer binding site from the target capture sequence. Generally, the portion of the probe (or its amplicons) that contains the target capture sequence will be attached to the solid support resulting in removal of the primer binding site from the solid support and retention of the target capture sequence. Thus, the cleaved probe can be used

for hybridizing target nucleic acids and the cleaved probe can be extended using method set forth previously herein.

In some embodiments, a nucleic acid probe will include two different cleavage sites. A first cleavage site will be located between a first primer binding site and one or more other sequence elements of the probe. A second cleavage site can be located between a second primer binding site and the one or more other sequence elements of the probe. The cleavage sites can be reactive to different cleavage reactions such that each one can be selectively cleaved without necessarily cleaving the other. Accordingly, the first cleavage site can be cleaved prior to modifying the probe (for example, prior to producing an extended probe), thereby separating the first primer binding site from the one or more other sequence elements that remain attached to a solid support. The second cleavage site can be cleaved after modifying the probe (for example, after producing the extended probe), thereby releasing the modified probe for subsequent detection.

Alternatively, a nucleic acid probe can include the first cleavage site and a primer that is used to capture or amplify the nucleic acid probe can include the second cleavage site. In this configuration, the first cleavage site can be located between a first primer binding site and one or more other sequence elements of the probe such that cleavage separates the first primer binding site from one or more other sequence elements of the probe that remain attached to a solid support. Again, this first cleavage step will typically be carried out prior to modifying the probe (for example, prior to producing an extended probe). A second cleavage step can be carried out to cleave the second cleavage site after modifying the probe (for example, after producing the extended probe), thereby releasing the modified probe for subsequent detection.

The two embodiments above exemplify a cleavage site located between a point of attachment of a nucleic acid probe (or modified nucleic acid probe) and one or more sequences of the probe (or modified probe) that contain information such as a spatial barcode or target sequence. Thus, this cleavage site is useful for release of modified probes (e.g. extended probes) to detect the sequence information and determine what sequences are present in a biological specimen and where the sequences are present in the specimen.

In some embodiments, one or more probes that are contacted with a solid support in a method set forth herein can include a sequencing primer binding site. Accordingly, a modified probe (e.g. extended probe) can be detected in a

sequencing technique that includes a step of hybridizing a sequencing primer to the sequencing primer binding site. The sequencing primer binding site can be located in the probe such that cleavage of a modified version of the probe (e.g. an extended probe) will yield a released probe that includes the sequencing primer binding site. The sequencing primer binding site can be a universal sequencing primer binding site such that a plurality of different probes (e.g. having different barcode and/or target sequences) will have the same sequencing primer binding site.

This disclosure further provides a method for spatially tagging nucleic acids of a biological specimen, the method including steps of (a) providing an array of beads on a solid support, wherein different nucleic acid probes are attached to different beads in the array, wherein the different nucleic acid probes each include a barcode sequence, wherein each bead includes a different barcode sequence from other beads on the solid support, and wherein each of the different nucleic acid probes includes a target capture sequence; (b) performing a decoder probe hybridization reaction on the solid support to determine the barcode sequences at the randomly located probes on the solid support; (c) contacting a biological specimen with the array of beads; (d) hybridizing the different nucleic acid probes to target nucleic acids from portions of the biological specimen that are proximal to the beads; and (e) extending the different nucleic acid probes to produce extended probes that include sequences from the target nucleic acids and the barcode sequences, thereby tagging the nucleic acids of the biological specimen.

It will be understood that manipulations of solid supports or of nucleic acids attached to solid supports can be carried out using beads as solid supports. The beads can be attached to a surface (e.g. an array of wells as in a BeadArray™ from Illumina) before or after such manipulations are carried out. For example, nucleic acid probes can be captured on beads before or after the beads are distributed on an array, nucleic acid probes can be amplified to create amplicons on beads before or after the beads are distributed on an array etc.

EXAMPLE I

Spatially Tagging mRNA from a Tissue Sample Using Illumina Flow Cells

A method for generating barcoded oligo-dT containing clusters, then revealing the barcoded oligo-dT with a restriction enzyme digest followed by sequencing is described in Fig. 1. A library of fragments containing a single stranded, barcoded oligo-dA, P5',P7, SBS3 sequencing primer binding site and a

BspHI restriction enzyme site (shown in the top panel of Fig. 1) were prepared by oligo synthesis (Integrated DNA Technologies). The barcodes were 27mers and were randomly generated during synthesis. The binding site for the SBS3 sequencing primer was included for decoding of the barcode by sequencing. An oligo-dA stretch was included to generate an oligo dT site upon clustering and linearization. Bridge amplification and clustering were performed according to standard cluster chemistry (Illumina TruSeq PE Cluster Kit v3 cBot P/N: 15037931) on an Illumina GA flow cell using manufacture's recommended protocol.

Following bridge amplification and clustering the clusters were linearized by cleavage of 8-oxo-G in P7 primer using Formamidopyrimidine DNA glycosylase (Fpg) enzyme provided in the TruSeq PE Cluster kit. This was followed by restriction enzyme digest with 200 Units/mL BspH1 (NEB Cat # R0517L at 37°C for 15min to remove P7' from the P5 adapter anchored strand of the cluster to unveil the oligo-dT stretch for subsequent extension in the presence of an mRNA. Enzyme concentrations in the range of 100-400U/mL have been tested for 15 or 30min. The de-coding of the barcode was initiated by the SBS3 sequencing primer.

As shown in the bottom panel of Fig. 1, oligo-dT sequences in the cluster were used to capture poly A+ RNA after decoding of the barcode. Barcoded cDNA was produced by extension of the oligo-dT strand of the cluster using TruSeq RNA Sample Prep Kit (Illumina P/N: 15012997) and MMLV Reverse Transcriptase 1st-Strand cDNA Synthesis Kit (Epicentre P/N: MM070150) according to the manufacturer's recommended conditions. The captured RNA was used as a template. Barcoded cDNA was released from the P5 sequence of the flow cell using Illumina's Uracil Specific Excision reagents (USER) (Illumina's TruSeq PE cluster kit) liberating a barcoded cDNA library that was used for sequencing on a second Illumina flow cell.

The availability of oligo dT capture sequence after the restriction enzyme digest with BspH1 was confirmed by hybridizing the linearized clusters with a Cy5 labeled poly A (24mer) as diagrammed in panel A of Fig. 2. Briefly, after the restriction enzyme digestion, the clusters were treated with 0.1N NaOH and washed with HT2 low salt buffer to remove the second strand on the flow cell. Then, 500nM of Cy5 oligo-dA (24mer) was flowed over the linearized and denatured clusters at 30µl / min rate and incubated at 40°C for 5min and then imaged. Hybridization of Cy5 labeled poly A to the oligo dT was detected in lanes 2-7 of the GA flow cell where the oligo dT containing BODT-1 libraries were present (see the image of the

flow cell shown in Fig. 2, Panel B). As evident from the flow cell image (Panel B), and the bargraph (Panel C), the control PhiX libraries (lanes 1 and 8 of the flow cell) were shown to have very low fluorescence in the Cy5 signal. These results demonstrated that an oligo-dT site can be created in the cluster that upon linearization can bind specifically to Cy5 poly A (24mer).

The sequencing metrics of the flow cell described above with 3.2pM of BODT-1 library is given in the table shown in Fig. 3. Millions of reads were detected in 21 tiles from GA sequencing. Following sequencing, the number of unique barcodes were determined as plotted in Fig. 4. This was done by assuming that every passing filter (PF) read was a barcode and determining the number of unique reads (barcodes) in each lane. Between 5 and 11 million unique barcoded clusters were detected after sequencing tiles compared to the PhiX control libraries. These results demonstrated that sequence decoding of a library of barcoded oligo-dT sequences is feasible and generates millions of unique barcodes.

EXAMPLE II

Cell Adhesion on Illumina Flow Cells

Single cells were captured on a patterned flow cell (HiSeq X10 flow cell, Illumina). All reagent flow steps were performed using a peristaltic pump or the cBOT cluster generation instrument (Illumina). Briefly, nuclease free water was flowed on all lanes of the patterned flow cell followed by 30-70K Poly D Lysine Solution (100 µg/ml and 20 µg/ml) at a flow rate of 100 µl/min for 8 min. Heat inactivated Fetal Bovine Serum (Life Technologies #10082-139) was also tested as an adhesive. The adhesives were incubated on the flow cell lanes for 1 hr, followed by a 1X PBS + 0.5% Pluronic F-68 (Life Technologies #24040-032) wash. Next, the cells were adhered to the coated flow cells by flowing 5 to 50 cells / µl or approximately 100 – 1000 cells per lane at a rate of 100 µl / min, followed by an incubation step for 60min to bind the cells. The flow cell was washed with 1X PBS/0.5% pluronic at a rate of 75 µl / min. If cells were fixed on the flow cell, 1% Paraformaldehyde (PFA) was flowed on the flow cell after flowing the cells as described above and incubated for 15min followed by the 1X PBS/0.5% pluronic was step. The flow cell was removed and the number of cells per lane counted using a microscope.

Fig. 5, Panel A shows an image of cells captured on the patterned flow cell. The cell count data shown in Fig. 5, Panel B confirmed that the poly D Lysine

coated flow cells aided cell adherence compared to the BSA coated or no adhesive treated control. As shown in Fig. 6, the adhered cells can be successfully fixed with 1% PFA.

EXAMPLE III

Spatially Localized Capture of Target mRNA by Probes Attached to a Gel Surface

This example describes creation of a lawn of poly T probes on a gel coated slide, placement of tissue slices on top of a lawn of poly T probes, release of RNA from the tissue sections, capture of the released mRNA by the poly T probes, reverse transcription to Cy 3 label the poly T probes, removal of the tissue and imaging of the slide.

Fig. 7, Panel A shows a diagrammatic representation of steps and reagents used to create probes attached to a gel. Briefly, a microscope slide was coated with silane free acrylamide (SFA), P5 and P7 primers were attached (see US Pat. App. Pub. No. 2011/0059865 A1, which is incorporated herein by reference), probes having a poly A sequence and either a P5 or P7 complementary sequence were hybridized to the P5 and P7 primers, respectively, and the P5 and P7 primers were extended to produce poly T sequence extensions. A quality control step was performed by hybridizing Cy5 labeled polyA oligonucleotides to the extended primers and imaging the surface using an Axon Imager.

As shown in Panel B of Fig. 7, a tissue section was placed on the gel having the polyT extended primers. The tissue was treated to release mRNA and poly A tails of the released mRNA were hybridized to poly T sequences of the extended primers. The poly T sequences were extended using the captured mRNAs as templates and the extended primers were selectively labeled with Cy3. The tissue was removed from the gel and the surface was imaged to detect Cy3 flourescence.

As shown in the image of Fig. 7, areas of the gel that were proximal to areas of the tissue that released mRNA species appeared fluorescent while areas that did not release mRNA appeared dark in the image. Thus, the captured mRNA created a fingerprint-like image of the tissue.

EXAMPLE IV

Spatially Localized Capture of Target mRNA by Probes Attached to a BeadArray™ Surface

This example describes placement of tissue slices on top of a BeadArray™ having poly T probes, release of RNA from the tissue sections, capture of the released mRNA by the poly T probes, reverse transcription to Cy5 label the poly T probes, removal of the tissue and imaging of the BeadArray™.

As shown in Panel A of Fig. 8, a mouse olfactory tissue section was placed on a BeadArray™ having polyT probes. The tissue was treated to release mRNA and poly A tails of the released mRNA were hybridized to poly T sequences of the probes. The poly T sequences were extended using the captured mRNAs as templates and the extended primers were selectively labeled with Cy5. The tissue was removed from the BeadArray™ and the BeadArray™ was imaged to detect Cy5 flourescence.

As shown in Panel B of Fig. 7, areas of the BeadArray™ that were proximal to areas of the tissue that released mRNA species appeared fluorescent while areas that did not release mRNA appeared dark in the image. Thus, the captured mRNA created a fingerprint-like image of the tissue.

Throughout this application various publications, patents or patent applications have been referenced. The disclosure of these publications in their entireties are hereby incorporated by reference in this application.

The term “comprising” is intended herein to be open-ended, including not only the recited elements, but further encompassing any additional elements.

A number of embodiments have been described. Nevertheless, it will be understood that various modifications may be made. Accordingly, other embodiments are within the scope of the following claims.

Claims

1. A method for spatially tagging nucleic acids of a biological specimen, comprising:

(a) attaching different nucleic acid probes to a solid support to produce randomly located probes on the solid support, wherein the different nucleic acid probes each comprise a barcode sequence, and wherein each of the randomly located probes comprise different barcode sequences from other randomly located probes on the solid support;

(b) performing a nucleic acid detection reaction on the solid support to determine the barcode sequences of the randomly located probes on the solid support;

(c) contacting a biological specimen with the solid support that comprises the randomly located probes;

(d) hybridizing the randomly located probes to target nucleic acids from portions of the biological specimen that are proximal to the randomly located probes; and

(e) extending the randomly located probes to produce extended probes that comprise the barcode sequences and sequences from the target nucleic acids, thereby spatially tagging the nucleic acids of the biological specimen.

2. The method of claim 1, wherein step (a) further comprises amplifying each of the randomly located probes on the solid support to form a cluster that comprises multiple amplicons of the nucleic acid probes on the solid support.

3. The method of claim 2, wherein the solid support comprises a plurality of the clusters and the average pitch for the plurality of clusters is less than 10 microns.

4. The method of claim 2 or 3, wherein the solid support comprises a plurality of the clusters and the average area of the clusters is less than 100 μm^2.

5. The method of any one of the preceding claims, wherein the nucleic acid detection reaction comprises a sequencing reaction.

6. The method of claim 1, wherein the nucleic acid detection reaction comprises a decoder probe hybridization reaction.

7. The method of any one of the preceding claims, wherein the solid support comprises an array of beads, wherein the different nucleic acid probes are attached to different beads in the array.

8. The method of claim 7, wherein step (a) comprises randomly distributing the different beads on the solid support.

9. The method of claim 7, wherein the solid support comprises wells having dimensions that accommodate no more than a single bead of said plurality of beads.

10. The method of claim 9, wherein the different beads are attached to different barcode sequences and wherein the number of different barcode sequences exceeds the number of wells.

11. The method of any one of claims 7 through 10, wherein the solid support comprises at least 1×10^6 of the different beads that comprise the different nucleic acid probes.

12. The method of any one of claims 7 through 11, wherein the array of beads comprises an average pitch of less than 10 microns

13. The method of any one of claims 7 through 12, wherein the beads comprise an average diameter of less than 10 microns.

14. The method of any one of the preceding claims, wherein the attaching of the different nucleic acid probes to the solid support comprises contacting a fluidic mixture of the different nucleic acid probes with the solid support.

15. The method of claim 14, wherein the solid support comprises a pattern of discrete features and wherein the different nucleic acid probes are randomly distributed to discrete features in the pattern.

16. The method of claim 15, wherein the number of different barcode sequences exceeds the number of features that are contacted by the fluidic mixture.

17. The method of any one of claims 1 or 14 through 16, wherein the solid support comprises at least 1 x 10^6 of the different nucleic acid probes.

18. The method of any one of the preceding claims, wherein the solid support comprises a gel coating.

19. The method of claim 18, wherein a plurality of nucleic acid primers is attached to the gel coating, wherein the nucleic acid primers in the plurality comprise a universal primer sequence that is common to the nucleic acid primers in the plurality.

20. The method of claim 19, wherein the different nucleic acid probes comprise a universal primer binding sequence that hybridizes to the universal primer sequence.

21. The method of claim 20, wherein step (a) further comprises amplifying the different nucleic acid probes by extension of the nucleic acid primers, thereby producing nucleic acid clusters that comprise multiple amplicons of the nucleic acid probes on the solid support.

22. The method of claim 21, wherein a second plurality of nucleic acid primers is attached to the gel coating, and wherein the nucleic acid primers in the second plurality comprise a second universal primer sequence that is common to the nucleic acid primers in the second plurality.

23. The method of claim 22, wherein the different nucleic acid probes comprise a second universal primer binding sequence that hybridizes to the second universal primer sequence and wherein the amplifying comprises bridge amplification.

24. The method of any one of the preceding claims, wherein the different nucleic acid probes comprise different target capture sequences that hybridize to different target nucleic acid sequences from the biological specimen.

25. The method of any one of claims 1 through 23, wherein the different nucleic acid probes comprise a target capture sequence that is common to the randomly located probes on the solid support.

26. The method of claim 25, wherein the target capture sequence comprises a poly T or poly A sequence.

27. The method of any one of the preceding claims, further comprising a step of acquiring an image of the biological specimen in contact with the solid support.

28. The method of claim 27, further comprising a step of correlating the barcode sequences determined at the randomly located probes with locations in the image of the biological specimen.

29. The method of claim 1, 27 or 28, further comprising removing the extended probes from the solid support.

30. The method of claim 29, further comprising pooling the extended probes to form a mixture of the extended probes that have been removed from the solid support.

31. The method of claim 29 or 30, further comprising determining the sequences from the target nucleic acids and the barcode sequences for the extended probes that have been removed from the solid support.

32. The method of claim 29, 30 or 31, further comprising attaching the extended probes that have been removed from the solid support to a second solid support.

33. The method of claim 31 or 32, wherein the sequences from the target nucleic acids and the barcode sequences are determined by a sequencing-by-synthesis technique.

34. The method of any one of the preceding claims, wherein the solid support is located in or on a flow cell during step (b).

35. The method of claim 34, wherein the solid support is removed from the flow cell during step (c).

36. The method of claim 34, wherein the flow cell is opened to expose the solid support during step (c).

37. The method of any one of the preceding claims, wherein the biological specimen that is contacted with the solid support is a mixture of cells.

38. The method of claim 37, wherein step (c) further comprises attaching the cells to the solid support.

39. The method of claim 37 or 38, wherein step (c) further comprises lysing the cells to release the target nucleic acids from the cells.

40. The method of any one of claims 1 through 36, wherein the biological specimen that is contacted with the solid support is a tissue.

41. The method of claim 40, wherein step (c) further comprises attaching the tissue to the solid support.

42. The method of claim 40 or 41, wherein step (c) further comprises permeabilizing the tissue to release the target nucleic acids from the tissue.

43. The method of any one of the preceding claims, wherein the target nucleic acid is selected from the group consisting of mRNA, gDNA, rRNA or tRNA.

44. A method for spatially tagging nucleic acids of a biological specimen, comprising:

(a) providing a plurality of nucleic acid primers attached to a solid support, wherein the nucleic acid primers in the plurality comprise a universal primer sequence that is common to the nucleic acid primers in the plurality;

(b) binding a population of nucleic acid probes to the plurality of nucleic acid primers, wherein the nucleic acid probes comprise a universal primer binding sequence that hybridizes to the universal primer sequence, a target capture sequence and a barcode sequence that differs from barcode sequences of other nucleic acid probes in the population, thereby attaching the different nucleic acid probes at randomly located positions on the solid support;

(c) amplifying the different nucleic acid probes by extension of the nucleic acid primers, thereby producing nucleic acid clusters having copies of the barcode sequence and target capture sequence at the randomly located positions on the solid support;

(d) performing a sequencing reaction to determine the barcode sequences at the randomly located positions on the solid support;

(e) contacting a biological specimen with the nucleic acid clusters on the solid support;

(f) hybridizing the target capture sequences of the clusters to target nucleic acids from portions of the biological specimen that are proximal to the clusters; and

(g) extending the target capture sequences to produce extended probes that comprise sequences from the target nucleic acids and the copies of the barcode sequences, thereby tagging the nucleic acids of the biological specimen.

45. The method of claim 44, wherein a second plurality of nucleic acid primers is attached to the solid support, wherein the nucleic acid primers in the second plurality comprise a second universal primer sequence that is common to the nucleic acid primers in the second plurality.

46. The method of claim 45, wherein the different nucleic acid probes comprise a second universal primer binding sequence that hybridizes to the second universal primer sequence.

47. The method of claim 46, wherein the amplifying of the different nucleic acid probes comprises bridge amplification wherein nucleic acid primers in the first and second plurality are extended.

48. The method of any one of claims 45 through 47, wherein the nucleic acid probes further comprise a cleavage site between the target capture sequence and the second universal primer binding sequence.

49. The method of claim 48, further comprising a step of cleaving the cleavage site between the target capture sequence and the second universal primer binding sequence prior to step (g).

50. The method of any one of claims 44 through 49, wherein the different nucleic acid probes comprise different target capture sequences that hybridize to different target nucleic acid sequences from the biological specimen.

51. The method of any one of claims 44 through 49, wherein the different nucleic acid probes comprise a target complement sequence that is common to the randomly located probes on the solid support.

52. The method of claim 51, wherein the target complement sequence comprises a poly T or poly A sequence.

53. The method of claim 51, wherein the target nucleic acids comprise mRNA molecules having polyA tails that hybridize to the target complement sequences in step (f).

54. The method of any one of claims 44 through 53, wherein each of the nucleic acid probes comprises a cleavage site located between the barcode sequence and a point of attachment of the nucleic acid probe to the solid support.

55. The method of claim 54, further comprising removing the extended probes from the solid support by cleaving at the cleavage site located between the barcode sequence and a point of attachment of the nucleic acid probe to the solid support.

56. The method of claim 55, further comprising pooling the extended probes to form a mixture of the extended probes that have been removed from the solid support.

57. The method of claim 55 or 56, further comprising determining the sequences from the target nucleic acids and the barcode sequences for the extended probes that have been removed from the solid support.

58. The method of claim 55, 56 or 57, further comprising attaching the extended probes that have been removed from the solid support to a second solid support.

59. The method of claim 57 or 58, wherein the sequences from the target nucleic acids and the barcode sequences are determined by a sequencing-by-synthesis technique.

60. The method of claim 59, wherein the different nucleic acid probes further comprise a sequencing primer binding site and the sequencing-by-synthesis technique comprises hybridizing a sequencing primer to the sequencing primer binding site.

61. The method of claim 60, wherein the sequence of the sequencing primer binding site is the same for the different nucleic acid probes.

62. The method of any one of claims 44 through 61, wherein the solid support comprises a gel coating.

63. The method of claim 62, wherein the nucleic acid primers are attached to the solid support via the gel.

64. The method of any one of claims 44 through 63, wherein the solid support comprises a plurality of the clusters and the average pitch for the plurality of clusters is less than 10 microns.

65. The method of any one of claims 44 through 64, wherein the solid support comprises a plurality of the clusters and the average diameter of the clusters is less than 100 μm^2.

66. The method of any one of claims 44 through 65, wherein step (b) comprises contacting a fluidic mixture of the different nucleic acid probes with the solid support.

67. The method of any one of claims 44 through 66, wherein the solid support comprises a pattern of discrete features and wherein the different nucleic acid probes are randomly distributed to discrete features in the pattern.

68. The method of claim 67, wherein the number of different barcode sequences exceeds the number of features that are contacted by the fluidic mixture.

69. The method of any one of claims 44 through 68, wherein the solid support comprises at least 1×10^6 of the different nucleic acid probes.

70. The method of any one of claims 44 through 66, further comprising a step of acquiring an image of the biological specimen in contact with the solid support.

71. The method of claim 70, further comprising a step of correlating the barcode sequences determined at the randomly located positions with locations in the image of the biological specimen.

72. The method of any one of claims 44 through 71, wherein the solid support is located in or on a flow cell during step (d).

73. The method of claim 72, wherein the solid support is removed from the flow cell during step (e).

74. The method of claim 72, wherein the flow cell is opened to expose the solid support during step (e).

75. The method of any one of claims 44 through 74, wherein the biological specimen that is contacted with the solid support is a mixture of cells.

76. The method of claim 75, wherein step (e) further comprises attaching the cells to the solid support.

77. The method of claim 75 or 76, wherein step (e) further comprises lysing the cells to release the target nucleic acids from the cells.

78. The method of any one of claims 44 through 77, wherein the biological specimen that is contacted with the solid support is a tissue.

79. The method of claim 78, wherein step (e) further comprises attaching the tissue to the solid support.

80. The method of claim 78 or 79, wherein step (e) further comprises permeabilizing the tissue to release the target nucleic acids from the tissue.

81. The method of any one of claims 44 through 80, wherein the target nucleic acid is selected from the group consisting of mRNA, gDNA, rRNA or tRNA.

82. A method for spatially tagging nucleic acids of a biological specimen, comprising:

(a) providing an array of beads on a solid support, wherein different nucleic acid probes are attached to different beads in the array, wherein the different nucleic acid probes each comprise a barcode sequence, wherein each bead comprises a different barcode sequences from other beads on the solid support, and wherein each of the different nucleic acid probes comprises a target capture sequence;

(b) performing a decoder probe hybridization reaction on the solid support to determine the barcode sequences at the randomly located probes on the solid support;

(c) contacting a biological specimen with the array of beads;

(d) hybridizing the different nucleic acid probes to target nucleic acids from portions of the biological specimen that are proximal to the beads; and

(e) extending the different nucleic acid probes to produce extended probes that comprise sequences from the target nucleic acids and the barcode sequences, thereby tagging the nucleic acids of the biological specimen.

83. The method of claim 82, wherein step (a) comprises randomly distributing the different beads on the solid support.

84. The method of claim 82 or 83, wherein the solid support comprises wells having dimensions that accommodate no more than a single bead of said plurality of beads.

85. The method of claim 84, wherein the number of different barcode sequences exceeds the number of wells.

86. The method of any one of claims 82 through 85, wherein the array of beads comprises at least 1×10^6 of the different beads that comprise the different nucleic acid probes.

87. The method of any one of claims 82 through 86, wherein the array of beads comprises an average pitch of less than 10 microns

88. The method of any one of claims 82 through 87, wherein the beads comprise an average diameter of less than 10 microns.

89. The method of any one of claims 82 through 88, wherein the different nucleic acid probes comprise different target capture sequences that hybridize to different target nucleic acid sequences from the biological specimen.

90. The method of any one of claims 82 through 88, wherein the different nucleic acid probes comprise a target capture sequence that is common to the different nucleic acid probes in the array of beads.

91. The method of claim 90, wherein the target capture sequence comprises a poly T or poly A sequence.

92. The method of any one of claims 82 through 91, further comprising a step of acquiring an image of the biological specimen in contact with the array of beads.

93. The method of claim 92, further comprising a step of correlating the barcode sequences of the beads with locations in the image of the biological specimen.

94. The method of claim 82, 92 or 93, further comprising removing the extended probes from the array of beads.

95. The method of claim 94, further comprising pooling the extended probes to form a mixture of the extended probes that have been removed from the array of beads.

96. The method of claim 94 or 95, further comprising determining the sequences from the target nucleic acids and the barcode sequences for the extended probes that have been removed from the solid support.

97. The method of claim 94, 95 or 96, further comprising attaching the extended probes that have been removed from the solid support to a second solid support.

98. The method of any one of claims 82 through 97, wherein the sequences from the target nucleic acids and the barcode sequences are determined by a sequencing-by-synthesis technique.

99. The method of any one of claims 82 through 98, wherein the biological specimen that is contacted with the solid support is a mixture of cells.

100. The method of claim 99, wherein step (c) further comprises attaching the cells to the solid support.

101. The method of claim 99 or 100, wherein step (c) further comprises lysing the cells to release the target nucleic acids from the cells.

102. The method of claim 82, wherein the biological specimen that is contacted with the solid support is a tissue.

103. The method of claim 102, wherein step (c) further comprises attaching the tissue to the solid support.

104. The method of claim 102 or 103, wherein step (c) further comprises permeabilizing the tissue to release the target nucleic acids from the tissue.

105. The method of any one of claims 82 through 104, wherein the target nucleic acid is selected from the group consisting of mRNA, gDNA, rRNA or tRNA.

Figure 1

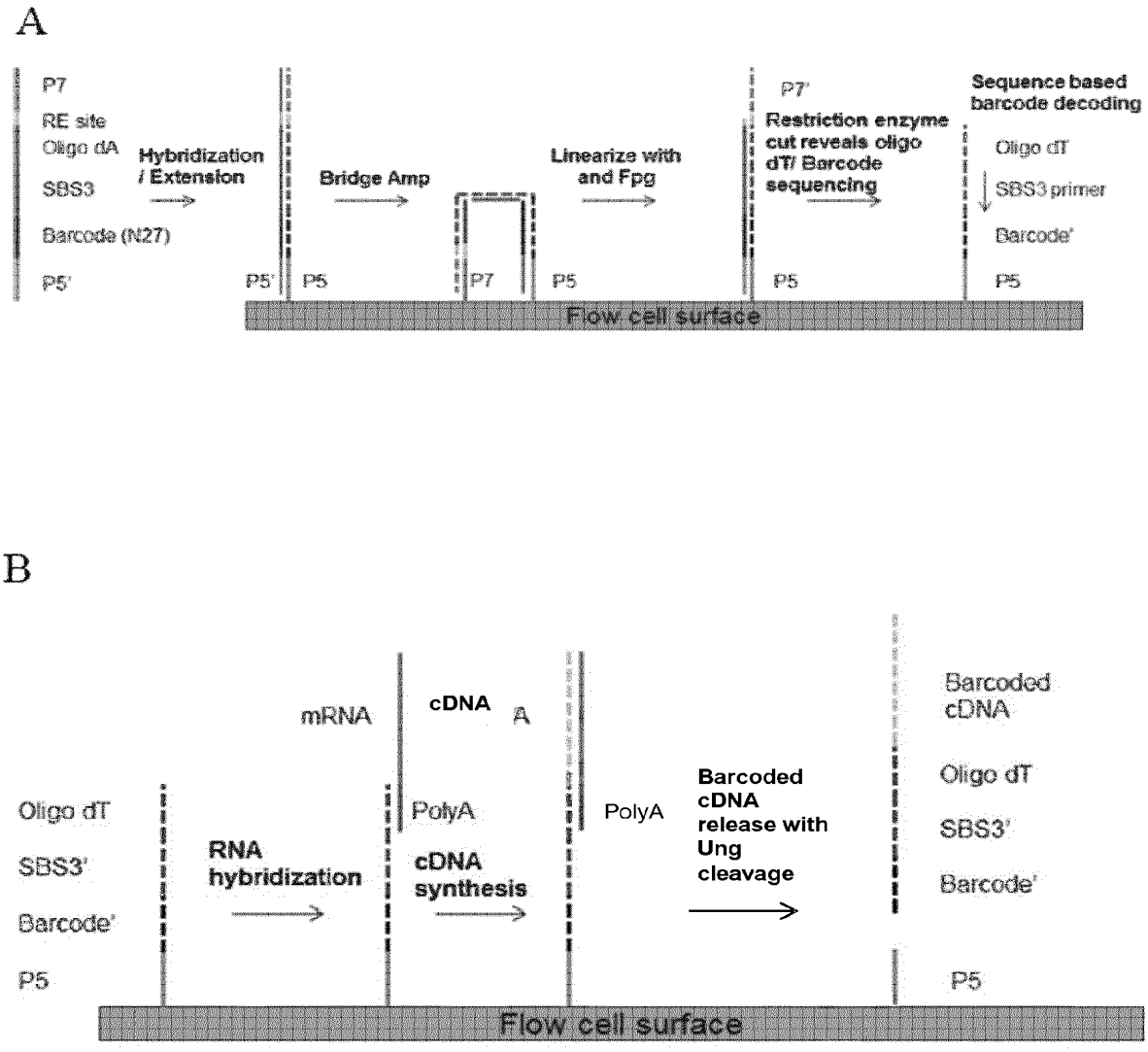
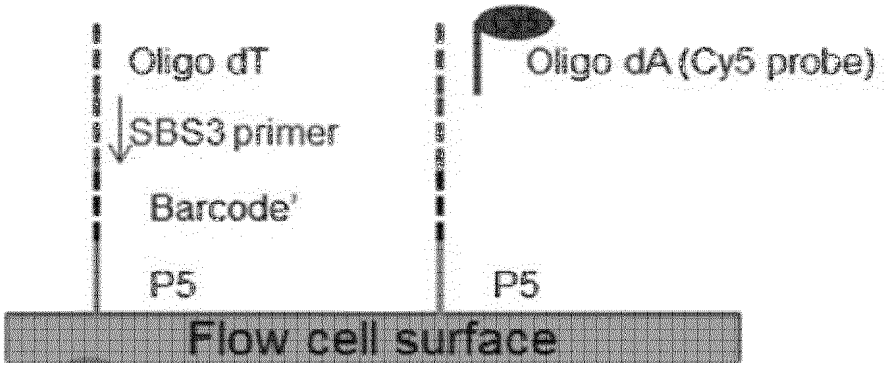
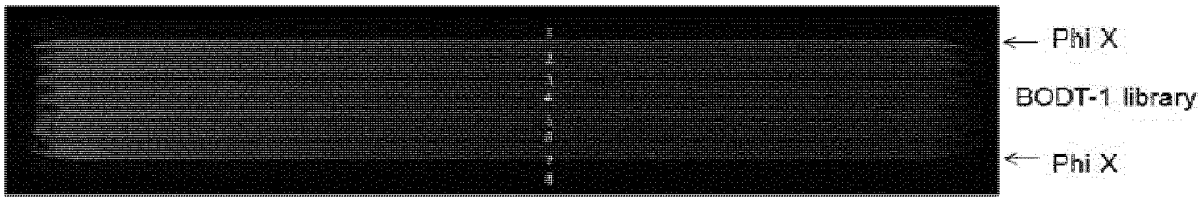

A
P7
RE site
Oligo dA
SBS3
Barcode (N27)
P5'
Hybridization / Extension
Bridge Amp
Linearize with and Fpg
P7'
Restriction enzyme cut reveals oligo dT/ Barcode sequencing
Sequence based barcode decoding
Oligo dT
SBS3 primer
Barcode'
P5
P7
Flow cell surface
B
mRNA
cDNA
A
Oligo dT
SBS3'
Barcode'
P5
RNA hybridization
PolyA
cDNA synthesis
Barcoded cDNA release with Ung cleavage
Barcoded cDNA
Flow cell surface

Figure 2

A


Oligo dT
SBS3 primer
Barcode'
P5
Oligo dA (Cy5 probe)
P5
Flow cell surface

B


Phi X
BODT-1 library
Phi X

C

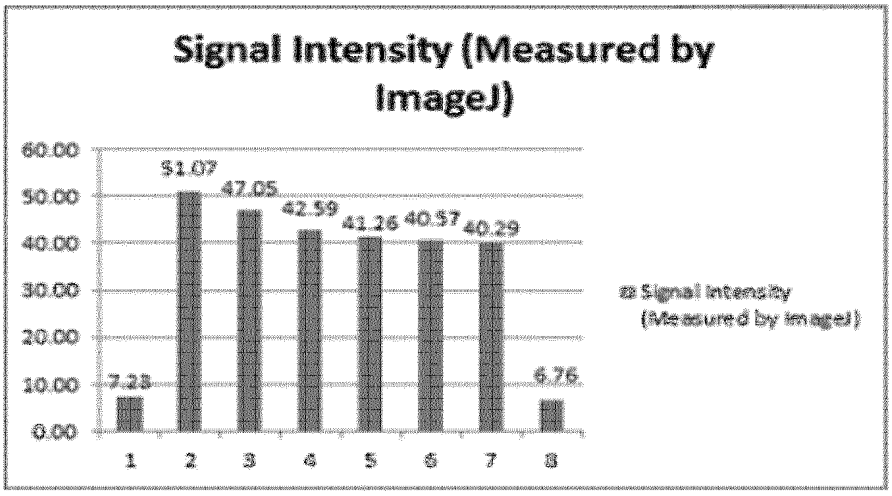

Signal Intensity (Measured by ImageJ)
60.00
50.00
40.00
30.00
20.00
10.00
0.00
7.28
51.07
47.05
42.59
41.26
40.57
40.29
6.76
1
2
3
4
5
6
7
8
Signal Intensity (Measured by ImageJ)

Figure 3

Lane	Tiles	Density (K/mm2)	Clusters PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30
1) Phi X	21	904 +/- 62	92.1 +/- 1.0	0.797 / 0.097	9.83	9.06	94.6
2) 3.2 nM BODT -1	21	1026 +/- 171	48.2 +/- 36.6	0.948 / 0.151	11.15	5.59	72.9
3) 3.2 nM BODT -1	21	975 +/- 186	41.9 +/- 43.2	0.964 / 0.154	10.6	5.03	70.8
4) 3.2 nM BODT -1	21	1144 +/- 24	85.3 +/- 14.3	0.969 / 0.163	12.45	10.62	76.4
5) 3.2 nM BODT -1	21	1136 +/- 37	90.6 +/- 0.8	0.977 / 0.173	12.35	11.19	80.5
6) 3.2 nM BODT -1	21	1128 +/- 28	90.5 +/- 0.7	0.953 / 0.182	12.27	11.1	80.6
7) 3.2 nM BODT -1	21	1131 +/- 28	90.3 +/- 0.8	0.955 / 0.170	12.3	11.1	82.6
8) Phi X	21	763 +/- 41	91.3 +/- 0.4	0.756 / 0.135	8.3	7.58	94

Figure 4

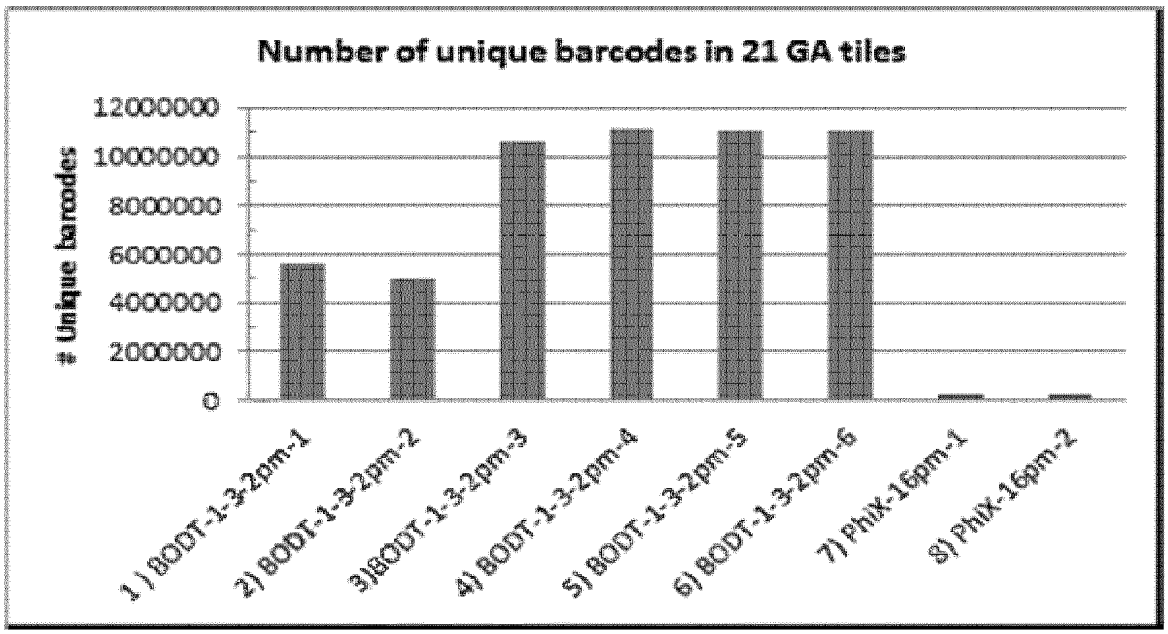
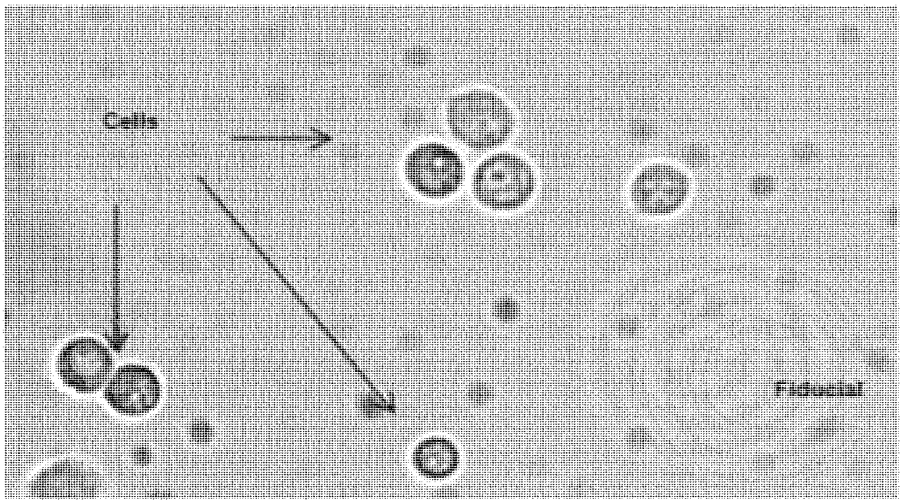

Number of unique barcodes in 21 GA tiles
Unique barcodes
12000000
10000000
8000000
6000000
4000000
2000000
0
1) BODT-1-3-2pm-1
2) BODT-1-3-2pm-2
3)BODT-1-3-2pm-3
4) BODT-1-3-2pm-4
5) BODT-1-3-2pm-5
6) BODT-1-3-2pm-6
7) PhiX-16pm-1
8) PhiX-16pm-2

Figure 5

A


Cells
Fiducial

B

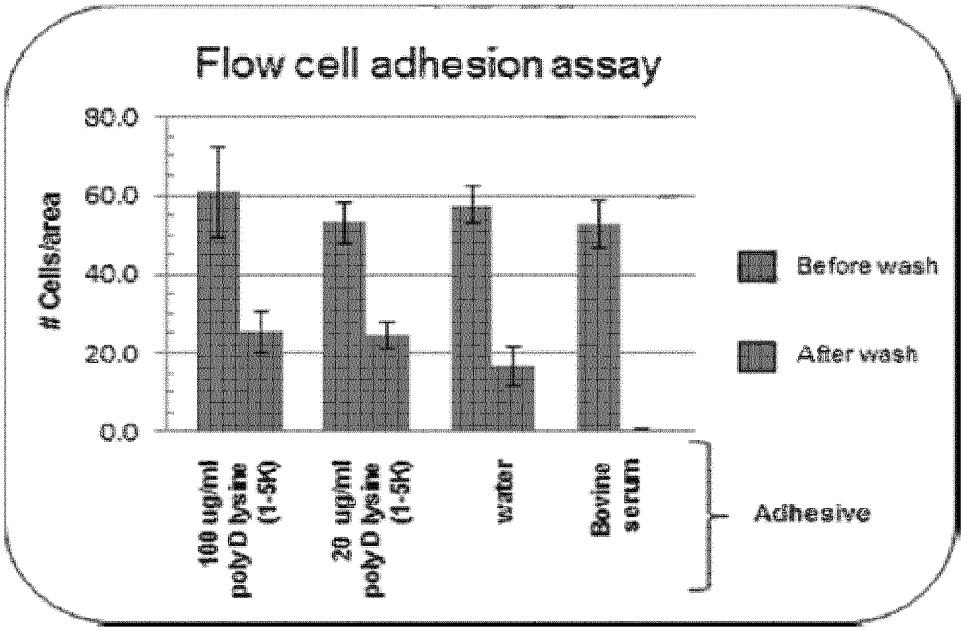

Flow cell adhesion assay
Cells/area
80.0
60.0
40.0
20.0
0.0
100 ug/ml polyD lysine (1-5K)
20 ug/ml polyD lysine (1-5K)
water
Bovine serum
Before wash
After wash
Adhesive

Figure 6

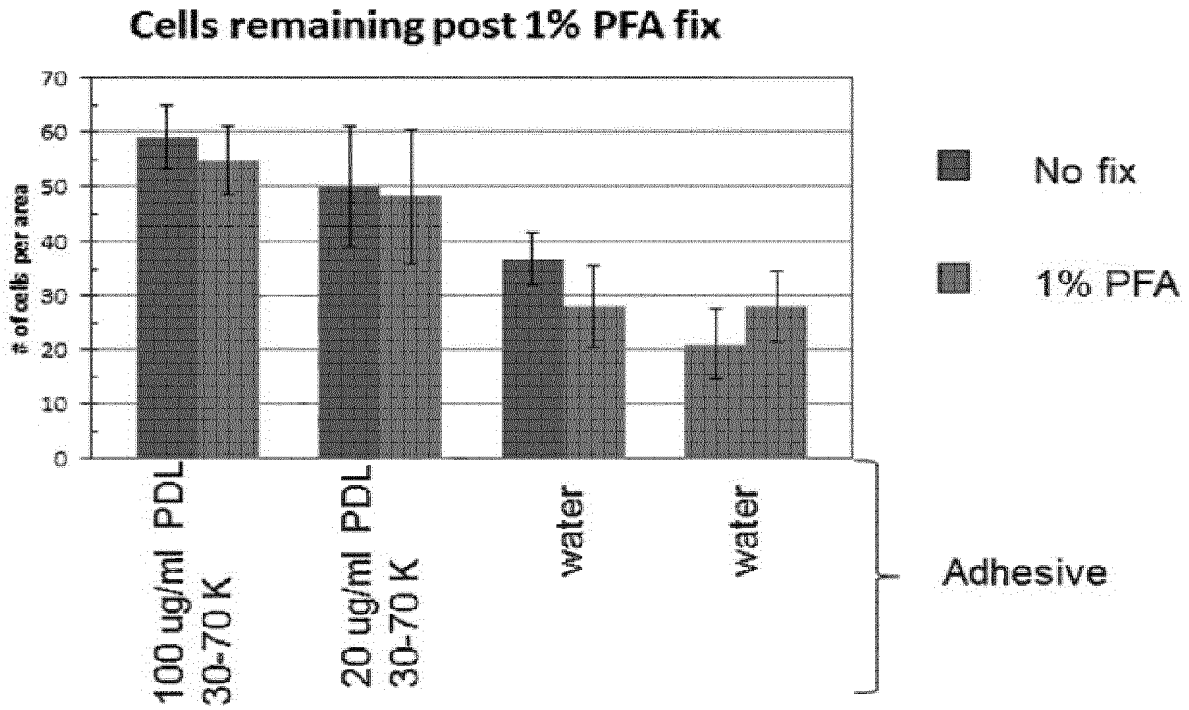
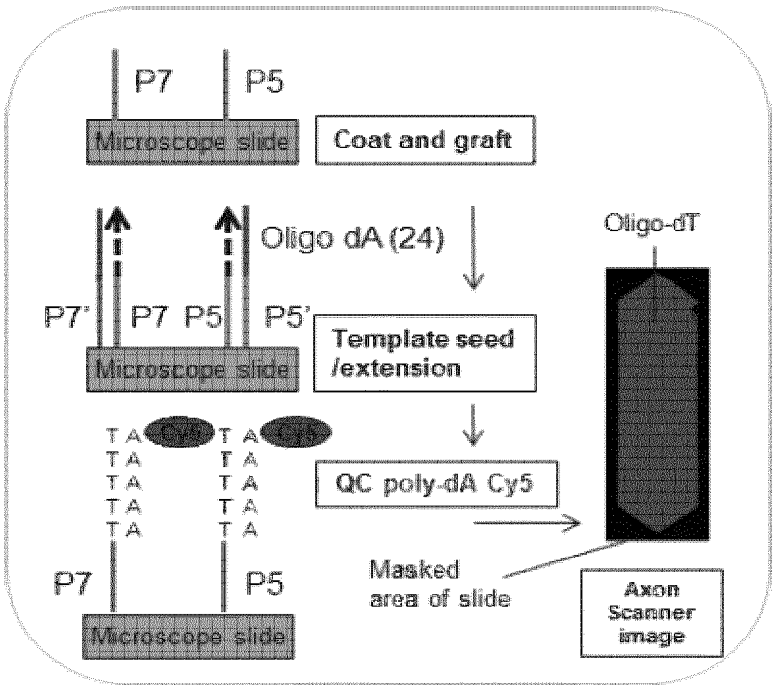

Cells remaining post 1% PFA fix
of cells per area
70
60
50
40
30
20
10
0
100 ug/ml PDL 30-70 K
20 ug/ml PDL 30-70 K
water
water
No fix
1% PFA
Adhesive

Figure 7

A


P7
P5
Microscope slide
Coat and graft
Oligo dA (24)
Oligo-dT
P7'
P7
P5
P5'
Microscope slide
Template seed /extension
Cy5
QC poly-dA Cy5
Masked area of slide
Axon Scanner image
P7
P5
Microscope slide

B

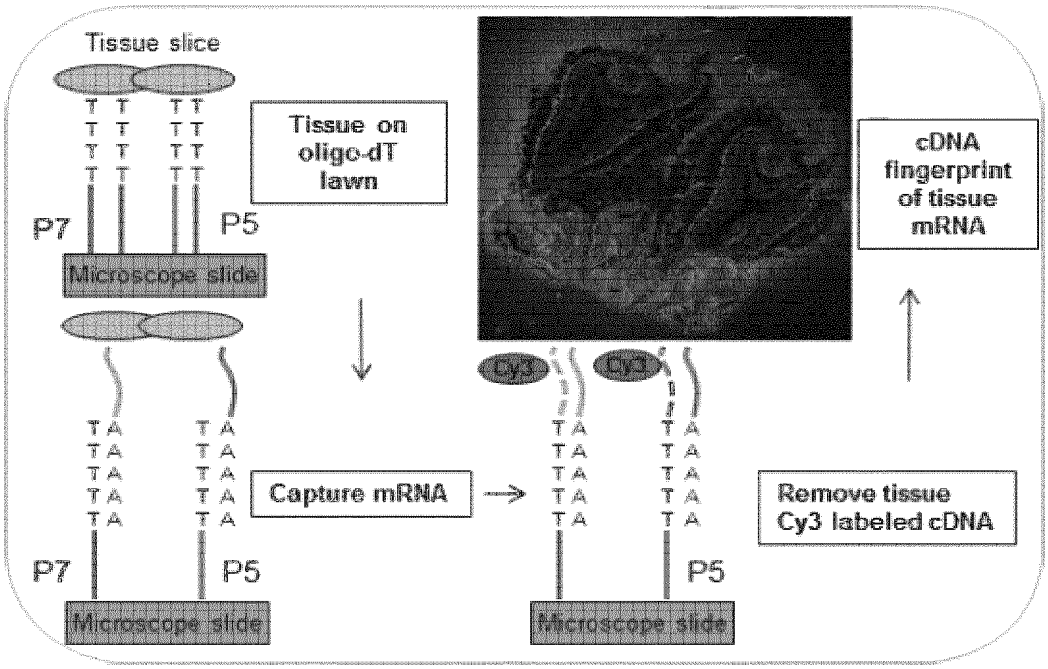
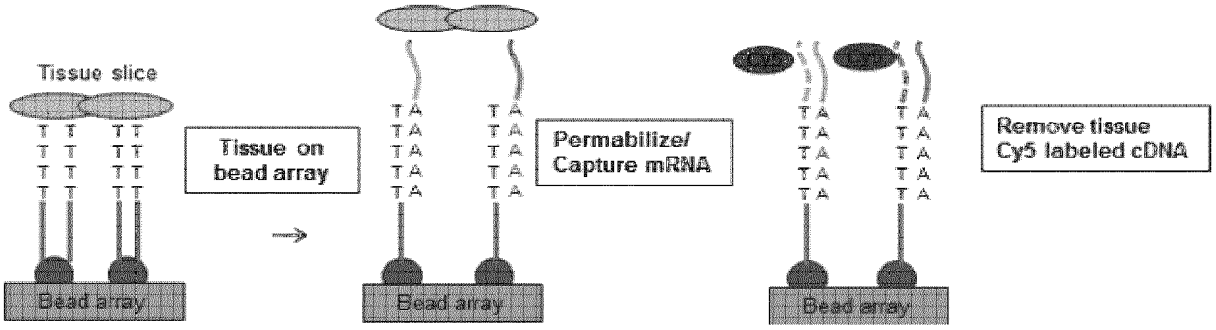
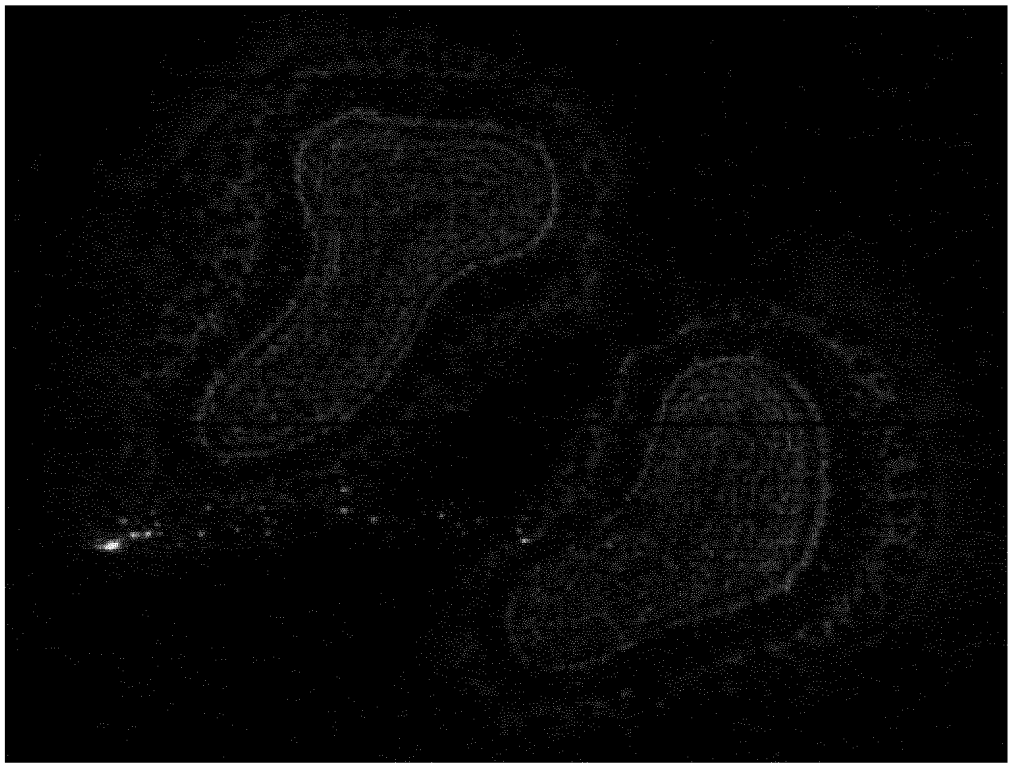

Tissue slice
Tissue on oligo-dT lawn
P7
P5
Microscope slide
cDNA fingerprint of tissue mRNA
Cy3
Capture mRNA
Remove tissue Cy3 labeled cDNA
P7
P5
Microscope slide
P5
Microscope slide

Figure 8

A


Tissue slice
Bead array
Tissue on bead array
Bead array
Permabilize/ Capture mRNA
Bead array
Remove tissue Cy5 labeled cDNA

B



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2016/057355

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2013/150083 A1 (MAX PLANCK GESELLSCHAFT [DE]) 10 October 2013 (2013-10-10) page 45	1-105
A	WO 2014/060483 A1 (SPATIAL TRANSCRIPTOMICS AB [SE]) 24 April 2014 (2014-04-24) page 30; example 6	1-105
A	Crosetto: "Spatially resolved transcriptomics and beyond", Nature Review Genetics, 1 January 2015 (2015-01-01), pages 57-66, XP055284555, Retrieved from the Internet: URL:http://www.nature.com/nrg/journal/v16/n1/pdf/nrg3832.pdf [retrieved on 2016-06-29]	1-105

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 July 2016	05/08/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Aslund, Fredrik

2

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2016/057355

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013150083 A1	10-10-2013	CA 2868689 A1	10-10-2013
		CA 2868691 A1	10-10-2013
		EP 2647426 A1	09-10-2013
		EP 2833997 A1	11-02-2015
		EP 2833998 A1	11-02-2015
		US 2015072867 A1	12-03-2015
		US 2015141269 A1	21-05-2015
		WO 2013150082 A1	10-10-2013
		WO 2013150083 A1	10-10-2013
WO 2014060483 A1	24-04-2014	CA 2886974 A1	24-04-2014
		EP 2909337 A1	26-08-2015
		US 2015344942 A1	03-12-2015
		WO 2014060483 A1	24-04-2014

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 May 2018 (24.05.2018)

WIPO | PCT

(10) International Publication Number
WO 2018/091676 A1

(51) International Patent Classification:
C12Q 1/6834 (2018.01) *C12Q 1/6874* (2018.01)
C12Q 1/6841 (2018.01)

(21) International Application Number:
PCT/EP2017/079650

(22) International Filing Date:
17 November 2017 (17.11.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1619458.1 17 November 2016 (17.11.2016) GB

(71) Applicant: **SPATIAL TRANSCRIPTOMICS AB** [SE/SE]; Södra Fiskartorpsvägen 15 C, SE - 11433 Stockholm (SE).

(72) Inventors: **FRISÉN, Jonas**; Bragevägen 5, 114 26 Stockholm (SE). **STÅHL, Patrik**; Ljustergränd 1, 116 68 Stockholm (SE). **LUNDEBERG, Joakim**; Torsvikssvän 20, 181 34 Lidingö (SE). **SALMÉN, Fredrik**; Tesselshadestraat 12 BIS, 3521 XV, 114 20 Utrecht (NL). **VICKOVIC, Sanja**; Ringvägen 87B, 118 61 Stockholm (SE).

(74) Agent: **WILKINS, Christopher**; St Bride's House, 10 Salisbury Square, London Greater London EC4Y 8JD (GB).

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

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

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

(54) Title: METHOD FOR SPATIAL TAGGING AND ANALYSING NUCLEIC ACIDS IN A BIOLOGICAL SPECIMEN

(57) Abstract: The present invention relates to methods for spatial tagging of nucleic acid molecules in a biological specimen and in particular to a method comprising: (a) providing a solid substrate on which multiple species of capture probes are immobilized such that each species occupies a distinct position on the solid substrate, wherein said probes are for a primer extension reaction and wherein each species of said capture probes comprise a nucleic acid molecule comprising: (i) a cleavage domain for releasing the capture probe from the surface of the solid substrate, (ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and (iii) a capture domain; (b) contacting said solid substrate with a biological specimen; and (c) releasing said capture probes from the surface of the solid substrate under conditions that allow nucleic acids of the biological specimen to hybridise to the capture domain in said capture probes and simultaneously and/or subsequently extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates to produce extended probes thereby spatially tagging the nucleic acids of the biological specimen, wherein step (c) comprises contacting said solid substrate with an aqueous reaction mixture comprising: (i) a polymerase enzyme capable of extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates; and (ii) means for releasing said capture probes from the surface of the solid substrate.

WO 2018/091676 A1

Method for spatial tagging and analysing nucleic acids in a biological specimen

The present invention relates generally to the spatial tagging of nucleic acid molecules in a biological specimen, e.g. the localised or spatial detection of nucleic acids in a biological specimen, such as a tissue sample (so-called “spatial genomics” and “spatial transcriptomics”). In particular, the present invention relates to an improvement in said methods to increase the number and diversity of unique nucleic acids captured and tagged from each biological specimen.

Spatial genomic and transcriptomic methods utilise nucleic acid probes immobilized to a solid substrate (e.g. an array) to capture nucleic acids (DNA or RNA) from a biological specimen and subsequently tag said nucleic acids (e.g. with a nucleic acid “barcode” sequence) based on their location within the biological specimen (see e.g. WO 2012/140224 herein incorporated by reference). The “tagged” nucleic acids are subsequently released from the solid substrate and analysed to generate information about the localisation, distribution and/or expression of genes, or indeed about the localisation or distribution of any genomic variation (not necessarily in a gene) in a biological specimen, such as a tissue sample.

By way of example, spatial transcriptomics utilises reverse transcription (RT) primers comprising unique positional tags (domains, such as barcode sequences), which are immobilized on a solid substrate, e.g. a glass slide, to generate an "array". The unique positional tags correspond to the location of the RT primers on the array (the features of the array). Biological specimens, such as tissue sections, are placed onto the array and a reverse transcription reaction is performed in the tissue section on the array. The RT primers, to which the RNA in the tissue sample binds (or hybridizes), are extended using the bound RNA as a template to obtain cDNA, which is therefore bound to the surface of the array. As consequence of the unique positional tags in the RT primers, each cDNA strand carries information about the position of the template RNA in the tissue section. The cDNA is then released from the surface of the array and analysed, e.g. sequenced, which results in a transcriptome with exact positional information. The sequence data can then be matched to a position in the tissue sample. For instance, the tissue section may be visualised or imaged, e.g. stained and photographed, before or after the cDNA synthesis step to facilitate the correlation of the positional tag in the cDNA molecule with a position within the tissue sample and the sequence data may be overlaid on

an image of the tissue specimen, e.g. using a computer, to display the expression pattern of any gene of interest across the tissue. However, the visualisation step is not essential as the tagged cDNA molecule may be correlated with a position in the tissue sample using other means, e.g. the unique profile of the molecules captured with the same positional tag, i.e. at the same location (feature) on the array, may enable the molecule to be correlated to a position within the tissue sample based on the known expression characteristics of cells or areas of cells within the tissue sample.

The sensitivity of spatial transcriptomics and genomics is limited by various factors including: (i) the efficiency of the hybridisation of the nucleic acid molecules to the capture probes; (ii) the efficiency of the subsequent extension reaction; (iii) the density of the capture probes on the array, i.e. the resolution of the array; and (iv) the non-specific extension of nucleic acid molecules that are not bound to capture probes, e.g. extension reactions primed by nucleic acids within the biological specimen, such as fragmented DNA and RNA caused by the preparation of the biological specimen, e.g. fixation of a tissue section. These limitations mean that some transcripts may be under-represented in the subsequent analysis. For instance, some transcripts may not be captured on low density arrays because the tissue in which they are expressed is does not make contact with a feature of the array. The limitations above can be particularly problematic for low abundance transcripts.

Methods for improving the density of capture probes on the array have been proposed in WO 2016/162309 (herein incorporated by reference), which utilises flow cell and bead array technology to generate “random” arrays with a high density of features, i.e. arrays on which the location of the capture probes is not pre-determined.

However, there is room to improve the sensitivity of spatial transcriptomic and genomic methods in order to overcome the limitations of the methods known in the art. In this respect, the inventors have surprisingly found that it is possible to increase the number and diversity of unique nucleic acid molecules captured by the capture probes by combining the step of releasing the probes from surface of the solid substrate, e.g. array, with the step of extending the probes using the captured nucleic acids as templates for extension, e.g. the release and extension steps are performed simultaneously.

In particular, the inventors have found that it is possible to combine the enzyme used to release the capture probes from the surface of the solid substrate (i.e. the cleavage enzyme) with the reaction mixture used in the nucleic acid extension step without a substantial loss of cleavage enzyme activity. Furthermore, and contrary to expectations, the combined (e.g. simultaneous) release and extension of the capture probes does not eliminate the localised capture of nucleic acid molecules, i.e. the spatial tagging of the nucleic acids in the biological specimen is retained in the methods of the invention. Moreover, as discussed in detail in the Examples below, the inventors have advantageously determined that the simultaneous release and extension of the capture probes increases the number and diversity of unique nucleic acid molecules captured by the methods of the invention, thereby improving the sensitivity of the spatial transcriptomic and genomic methods.

Whilst not wishing to be bound by theory it is hypothesised that the combination of the release and nucleic acid synthesis steps may result in the localised release of some or all of the capture probes prior to their extension, i.e. the probes may be released into the local vicinity of their original position on the array. It is thought that this may enhance the sensitivity of the methods by enabling capture probes (that have not hybridised to a nucleic acid from the biological specimen on contact between the specimen and the array) to diffuse into the specimen, thereby increasing the probability that the capture probe hybridises to a target sequence. Additionally or alternatively, the simultaneous release and extension of the capture probes may improve extension reaction, e.g. DNA synthesis reaction, because the capture probes are no longer physically constrained, i.e. immobilized on solid substrate, and/or because the microreactor environment formed between the biological specimen and the surface of the array is particularly suited for a templated nucleic acid extension reaction. Thus, the extension reaction may proceed more efficiently once capture probes, that hybridised to target sequences when they were still immobilised on the solid substrate, are released from the surface of the array.

Accordingly, in one embodiment the present invention provides a method for spatially tagging nucleic acids of a biological specimen comprising:

(a) providing a solid substrate on which multiple species of capture probes are immobilized such that each species occupies a distinct position on the solid

substrate, wherein said probes are for a primer extension reaction and wherein each species of said capture probe comprises a nucleic acid molecule comprising:

(i) a cleavage domain for releasing the capture probe from the surface of the solid substrate,

(ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and

(iii) a capture domain;

(b) contacting said solid substrate with a biological specimen; and

(c) releasing said capture probes from the surface of the solid substrate under conditions that allow nucleic acids of the biological specimen to hybridise to the capture domain in said capture probes and simultaneously and/or subsequently extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates to produce extended probes thereby spatially tagging the nucleic acids of the biological specimen.

The methods of the invention represent a significant improvement over other methods for spatial transcriptomics and genomics known in the art, e.g. WO 2012/140224, WO 2014/060483 and WO 2016/162309. The methods described herein result in an increase in the number and diversity of unique transcripts captured from the biological specimen, thereby providing a spatially tagged transcriptome that is more representative of the expression profile of the biological specimen. Moreover, as the capture probes may be released from the surface of the array prior to their extension the probes do not need to be attached to the array via their 5' ends in the methods of the present invention. In this respect, the capture probes may be attached to the array via their 3' ends and the release of the capture probes from the surface of the array will expose the 3' ends of the probes, enabling the probes to function as primers for nucleic acid extension reactions.

As discussed in more detail below, the method of the invention may comprise an additional step of analysing the extended probes. In this respect, it is evident that the combination of spatial tagging of the nucleic acids from a biological specimen and subsequent analysis of said tagged nucleic acids facilitates the localised detection of a nucleic acid in a biological specimen, e.g. tissue sample. Thus, in one embodiment, the method of the invention may be used for determining and/or analysing all of the transcriptome or genome of a biological specimen, e.g. the global transcriptome or genome of a biological specimen. However, the method is not limited to this and encompasses determining and/or analysing all or part of

the transcriptome or genome. Thus, the method may involve determining and/or analysing a part or subset of the transcriptome or genome, e.g. a transcriptome corresponding to a subset of genes, e.g. a set of particular genes, for example related to a particular disease or condition, tissue type etc.

Viewed from another aspect, the method steps set out above can be seen as providing a method of obtaining a spatially defined transcriptome or genome, and in particular the spatially defined global transcriptome or genome, of a biological specimen, e.g. tissue sample.

Alternatively viewed, the method of the invention may be seen as a method for localised or spatial detection of nucleic acid, whether DNA or RNA, in a biological specimen, e.g. tissue sample, or for localised or spatial determination and/or analysis of nucleic acid (DNA or RNA) in a tissue sample. In particular, the method may be used for the localised or spatial detection or determination and/or analysis of gene expression or genomic variation in a tissue sample. The localised/spatial detection/determination/analysis means that the RNA or DNA may be localised to its native position or location within a cell or tissue in the tissue sample. Thus for example, the RNA or DNA may be localised to a cell or group of cells, or type of cells in the sample, or to particular regions of areas within a tissue sample. The native location or position of the RNA or DNA (or in other words, the location or position of the RNA or DNA in the tissue sample), e.g. an expressed gene or genomic locus, may be determined.

Thus, in some embodiments, the invention provides a method for localised detection of nucleic acid in a biological specimen comprising:

(a) providing a solid substrate on which multiple species of capture probes are immobilized such that each species occupies a distinct position on the solid substrate, wherein said probes are for a primer extension reaction and wherein each species of said capture probe comprises a nucleic acid molecule comprising:

(i) a cleavage domain for releasing the capture probe from the surface of the solid substrate,

(ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and

(iii) a capture domain;

(b) contacting said solid substrate with a biological specimen;

(c) releasing said capture probes from the surface of the solid substrate under conditions that allow nucleic acids of the biological specimen to hybridise to

the capture domain in said capture probes and simultaneously and/or subsequently extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates to produce extended probes thereby spatially tagging the nucleic acids of the biological specimen; and

(d) analysing the extended probes of (c), i.e. analysing the spatially tagged nucleic acids of the biological specimen.

As described in more detail below, any method of nucleic acid analysis may be used in the analysis step (step (d)). Typically this may involve sequencing, i.e. analysing the sequence of the extended probes, but it is not necessary to perform an actual sequence determination. For example sequence-specific methods of analysis may be used. For example a sequence-specific amplification reaction may be performed, for example using primers which are specific for the positional domain and/or for a specific target sequence, e.g. a particular target DNA to be detected (i.e. corresponding to a particular cDNA/RNA or gene etc.). An exemplary analysis method is a sequence-specific PCR reaction.

The sequence analysis information obtained in step (d) may be used to obtain spatial information as to the RNA or DNA in the biological specimen, e.g. tissue sample. In other words the sequence analysis information may provide information as to the location of the RNA or DNA in the biological specimen, e.g. tissue sample. This spatial information may be derived from the nature of the sequence analysis information determined, for example it may reveal the presence of a particular RNA or DNA which may itself be spatially informative in the context of the biological specimen, e.g. tissue sample, used, and/or the spatial information (e.g. spatial localisation) may be derived from the position of the biological specimen, e.g. tissue sample, on the solid substrate, e.g. array, coupled with the sequencing information. Thus, the method may involve simply correlating the sequence analysis information to a position in the biological specimen, e.g. tissue sample, e.g. by virtue of the positional tag and its correlation to a position in the biological specimen, e.g. tissue sample. However, in some embodiments, spatial information may conveniently be obtained by correlating the sequence analysis data to an image of the biological specimen, e.g. tissue sample. Accordingly, in a preferred embodiment the method also includes a step of:

(e) correlating said sequence analysis information with an image of said biological specimen, wherein the biological specimen is imaged before or after step (c), i.e. after step (b).

It will be seen therefore that the array of the present invention may be used to capture RNA, e.g. mRNA, or DNA of a biological specimen, e.g. tissue sample, that is contacted with said solid substrate, e.g. array. The methods of the invention may thus be considered as methods of quantifying the spatial expression of one or more genes in a tissue sample. Expressed another way, the methods of the present invention may be used to detect the spatial expression of one or more genes in a biological specimen, e.g. tissue sample. In yet another way, the methods of the present invention may be used to determine simultaneously the expression of one or more genes at one or more positions within a biological specimen, e.g. tissue sample. Still further, the methods may be seen as methods for partial or global transcriptome or genome analysis of a biological specimen, e.g. tissue sample, with two-dimensional spatial resolution.

Step (c) in the method above of releasing said capture probes from the surface of the solid substrate and simultaneously and/or subsequently extending said capture probes will be seen as relating to combining the release and extension reactions, i.e. combining the means for achieving capture probe release and target nucleic acid templated extension of said capture probes. As discussed above, the improvement in the sensitivity of the method may be a consequence of the release of capture probes that have not bound to a target sequence in step (b), which may facilitate the binding of other target nucleic acid molecules in the biological specimen. Alternatively or additionally, the improvement in the sensitivity of the method may result from the reduction of steric hindrance on the capture probe-target nucleic acid complex or the formation of microenvironment that is particularly conducive to nucleic acid extension reactions, thereby enhancing the efficiency of the extension reaction.

Accordingly, the simultaneous release and extension of the capture probes does not mean that all capture probes will be released and extended at the same time, but rather that the means for releasing and extending the capture probes are applied to the solid substrate, e.g. array, simultaneously, i.e. at substantially the same time.

Notably, as mentioned above, the capture probes of the presented invention are not restricted to a particular orientation on the array, unlike the methods in WO 2012/140224, WO 2014/060483 and WO 2016/162309, which require the capture probes to be immobilized on the solid substrate such that they have a free 3' end capable of functioning as an extension primer. In this respect, the combination of

the release and extension steps eliminates the requirement for a particular orientation of the capture probes on the solid substrate. However, in some embodiments, it is preferred that the capture probes are immobilized on the solid substrate such that they have a free 3' end capable of functioning as an extension primer.

Furthermore, as the capture probes may be oriented on the solid substrate such that the capture domain is not free or available to interact with (i.e. bind or hybridise to) the nucleic acid molecules in the biological specimen (i.e. the capture probes may be immobilized via their 3' ends), it is not necessary to contact the biological specimen with the solid substrate under conditions that allow (i.e. are suitable for or facilitate) the nucleic acids of the biological specimen to hybridise to the capture domain in said capture probes. However, in some embodiments (e.g. where the capture probes are immobilized on the solid substrate such that they have a free 3' end capable of functioning as an extension primer, e.g. via their 5' ends) step (b) may be performed under conditions that allow the nucleic acids of the biological specimen to hybridise to the capture domain in said capture probes.

Thus, in embodiments where the capture probes are immobilized on the solid substrate such that they have a free 3' end capable of functioning as an extension primer (e.g. by their 5' end), some capture probes may be released from the solid substrate prior to their extension, i.e. some capture probes are released and subsequently extended. Moreover, some capture probes may be extended at the same time as they are released from the solid substrate, i.e. some capture probes are extended and released from the solid substrate simultaneously. Accordingly, in embodiments where the capture probes are immobilized on the solid substrate such that they have a free 3' end capable of functioning as an extension primer (e.g. by their 5' end), step (c) may comprise simultaneous release and extension reactions and sequential release and extension reactions, i.e. release and subsequent extension of the capture probes. The simultaneous reactions may occur particularly when step (b) is performed under conditions that allow the nucleic acids of the biological specimen to hybridise to the capture domain in said capture probes, i.e. such that some nucleic acids may hybridise to the capture domains of the capture probes prior to the combined release and extension reactions, i.e. prior to step (c).

Accordingly, in embodiments where the capture probes are immobilized on the solid substrate such that they do not have a free 3' end capable of functioning

as an extension primer (e.g. by their 3' end), all capture probes may be released from the solid substrate prior to their extension, i.e. all of capture probes are released and subsequently extended. Accordingly, in embodiments where the capture probes are immobilized on the solid substrate such that they do not have a free 3' end capable of functioning as an extension primer (e.g. by their 3' end), step (c) may comprise sequential release and extension reactions, i.e. release and subsequent extension of the capture probes, despite the presence of both components required for the release and extension reactions.

Thus, in some embodiments step (c) may be seen to comprise contacting said solid substrate, e.g. array, with:

(i) a polymerase enzyme capable of extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates; and

(ii) means for releasing said capture probes from the surface of the solid substrate,

at the same time.

The phrase "at the same time" means substantially the same time, i.e. one component may be contacted with the solid substrate before the other component, e.g. within seconds, (e.g. within 15, 30, 45, 60, 90, 120 or 180 seconds) or minutes (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12 or 15 minutes), but such that the reactions are allowed to proceed together. If one component is contacted with the solid substrate before the other component it is preferred that the means for releasing the capture probes from the surface of the solid substrate is contacted first and the polymerase enzyme is contacted within seconds or minutes as defined above. However, in some embodiments, it may be desirable to contact the polymerase enzyme first and contact the means for releasing the capture probes from the surface of the solid substrate within seconds or minutes as defined above.

In view of the fact that step (c) may comprise sequential release and extension reactions, i.e. release and subsequent extension of the capture probes, it will be evident that the sequential release and extension reactions may be achieved by contacting solid substrate with the means for releasing said capture probes from the surface of the solid substrate and the polymerase enzyme separately.

Thus, in some embodiments step (c) may be seen to comprise contacting said solid substrate, e.g. array, with means for releasing said capture probes from the surface of the solid substrate and subsequently contacting said solid substrate,

e.g. array, with a polymerase enzyme capable of extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates.

The term “subsequently” means that the polymerase enzyme is contacted with the solid substrate after the means for releasing said capture probes from the surface of the solid substrate is contacted with the solid substrate, such that the extension reaction does not occur immediately after the capture probes have hybridised to their target molecules. There is no particular limit on the amount of time that may be allowed to lapse between contacting the solid substrate with the means for releasing the capture probes and the polymerase enzyme. However, it is preferred that the polymerase enzyme is contacted with the solid substrate before the target nucleic acid molecules (e.g. RNA) have substantially degraded. Thus, in some embodiments, subsequently contacting the polymerase enzyme means that the polymerase enzyme is contacted with the solid substrate minutes or hours after the step of contacting the means for releasing said capture probes with the solid substrate. For instance, the polymerase enzyme may be contacted with the solid substrate at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50 or 60 minutes after the means for releasing said capture probes from the surface of the solid substrate, e.g. within 120, 90 or 60 minutes, i.e. between 1-120, 5-90, 10-60 minutes after the means for releasing said capture probes from the surface of the solid substrate. In some embodiments, the polymerase enzyme may be contacted with the solid substrate at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 18, 24, 36 or 48 hours after the means for releasing said capture probes from the surface of the solid substrate, e.g. within 72, 48 or 24 hours, i.e. between 1-72, 6-48, 12-24 hours after the means for releasing said capture probes from the surface of the solid substrate.

In a particularly preferred embodiment, the polymerase enzyme and means for releasing said capture probes from the surface of the solid substrate, e.g. cleavage enzyme, are combined in a single reaction mixture, which is contacted with the array. Thus, in some embodiments, step (c) comprises contacting said array with a reaction mixture (i.e. an aqueous reaction mixture) comprising:

(i) a polymerase enzyme capable of extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates; and

(ii) means for releasing said capture probes from the surface of the solid substrate.

Accordingly, the invention can be seen to provide the use of a reaction mixture (i.e. an aqueous reaction mixture) comprising:

(i) a polymerase enzyme capable of extending capture probes (as defined above) using the nucleic acid molecules hybridised to the capture probes as extension templates; and

(ii) means for releasing said capture probes from the surface of a solid substrate,

in a method for spatially tagging nucleic acids of a biological specimen, such as the methods defined herein.

The step of releasing the capture probes from the surface of the solid substrate may be achieved in a number of ways. In this respect, as mentioned above, the capture probes comprise a cleavage domain to facilitate the release of the capture probes. Accordingly, "means for releasing the capture probes from the surface of the solid substrate" includes any components that are suitable for this purpose, e.g. chemical or enzymatic means. As described below, the preferred method for releasing the capture probes from the surface of the solid substrate comprises enzymatic cleavage of the capture probe in the cleavage domain. In this respect, releasing the probe via the cleavage domain ensures that the released capture probes (including any extended or partially extended capture probes) comprise the positional domain, which contains information relating to the location of the capture probes on the array.

The term "cleavage domain" refers to a domain within the capture probe that can be cleaved specifically to release the capture probe (or part of the capture probe comprising the positional domain and the sequence 3' to the positional domain (if present), i.e. the capture domain and extension product (if present)) from the surface of the solid substrate.

As discussed below, the capture probes may be attached to the solid substrate by any suitable means, such as by chemical immobilization, e.g. by a chemical cross-linker. Thus, in some embodiments, the cleavage domain of the capture probe refers to a cleavable chemical cross-linker attaching the capture probe to the surface of the solid substrate. Thus, in some embodiments, the capture domain does not form part of the nucleotide sequence of the capture probe. For instance, the capture probe may comprise a nucleic acid positional domain and

capture domain and a non-nucleic acid cleavage domain (e.g. a cleavable chemical cross-linker) conjugated to the 5' or 3' end of the capture probe (depending on the orientation of the probe attached to the solid substrate). Thus, cleavage of said domain results in the release of the whole of the nucleic acid capture probe, i.e. comprising the positional domain and the sequence 3' to the positional domain from the surface of the solid substrate. Cleavable chemical cross-linkers capable of immobilizing nucleic acid molecules to a solid substrate, as defined herein, are known in the art.

In preferred embodiments, the cleavage domain of the capture probe refers to a nucleotide sequence within the capture probe that can be cleaved specifically, either chemically or preferably enzymatically. The location of the cleavage domain within the capture probe will depend on whether or not the capture probes are immobilized on the solid substrate such that they have a free 3' end capable of functioning as an extension primer (e.g. by their 5' or 3' end). For instance, if the capture probes are immobilized by their 5' end, the cleavage domain will be located 5' to the positional domain and cleavage of said domain results in the release of part of the capture probe comprising the positional domain and the sequence 3' to the positional domain, and optionally part of the cleavage domain, from the surface of the solid substrate. Alternatively, if the capture probes are immobilized by their 3' end, the cleavage domain will be located 3' to the capture domain (and positional domain) and cleavage of said domain results in the release of part of the capture probe comprising the positional domain and the sequence 3' to the positional domain from the surface of the solid substrate. Preferably, cleavage results in complete removal of the cleavage domain, particularly when the capture probes are immobilized via their 3' end as the presence of a part of the cleavage domain may interfere with the hybridisation of the capture domain and the target nucleic acid and/or its subsequent extension.

"Cleavage" is defined broadly herein to include any means of breaking a covalent bond. In some embodiments, cleavage involves cleavage of a covalent bond in a nucleotide chain (i.e. strand cleavage or strand scission), for example by cleavage of a phosphodiester bond.

Thus, the cleavage domain may comprise a sequence that is recognised by one or more enzymes capable of cleaving a nucleic acid molecule, i.e. capable of breaking the phosphodiester linkage between two or more nucleotides. For instance, the cleavage domain may comprise a restriction endonuclease (restriction

enzyme) recognition sequence. Restriction enzymes cut double-stranded or single stranded DNA at specific recognition nucleotide sequences known as restriction sites and suitable enzymes are well known in the art. For example, it is particularly advantageous to use rare-cutting restriction enzymes, i.e. enzymes with a long recognition site (at least 8 base pairs in length), to reduce the possibility of cleaving elsewhere in the nucleic acid, e.g. cDNA molecule. As mentioned above, it will be seen that releasing at least part of the capture probes from the surface of the solid substrate requires releasing a part comprising the positional domain and capture domain of the capture probe and, in some embodiments, any sequence produced by the templated extension of the capture domain. Thus, if the capture probe is extended partially or fully whilst still attached to (i.e. immobilized on) the solid substrate, cleavage of the cleavage domain will release the extended capture probe, comprising the positional domain, capture domain and extension product. Hence, in some preferred embodiments, i.e. when the capture probes are immobilized via their 5' end, cleavage of the capture probe should take place 5' to the positional domain.

By way of example, the cleavage domain may comprise a poly-U sequence which may be cleaved by a mixture of Uracil DNA glycosylase (UDG) and the DNA glycosylase-lyase Endonuclease VIII, commercially known as the USER™ enzyme.

A further example of a cleavage domain can be utilised in embodiments where the capture probe is immobilized to the solid substrate indirectly, i.e. via a surface probe defined below. The cleavage domain may comprise one or more mismatch nucleotides, i.e. when the complementary parts of the surface probe and the capture probe are not 100% complementary. Such a mismatch is recognised, e.g. by the MutY and T7 endonuclease I enzymes, which results in cleavage of the nucleic acid molecule at the position of the mismatch.

Another example of a cleavage domain that can be utilised in embodiments where the capture probe is immobilized to the solid substrate indirectly, i.e. via a surface probe, comprises a nickase recognition site or sequence. In this respect, nickase enzymes cleave only one strand in a nucleic acid duplex. Nickases are endonucleases which cleave only a single strand of a DNA duplex. Thus, the cleavage domain may comprise a nickase recognition site close to the 5' end of the surface probe (and/or the 5' end of the capture probe) such that cleavage of the surface probe or capture probe destabilises the duplex between the surface probe and capture probe thereby releasing the capture probe from the solid substrate.

Nickase enzymes could also be used in embodiments where the capture probe is immobilized to the solid substrate directly. For instance, the solid substrate may be contacted with a nucleic acid molecule that hybridises to the cleavage domain of the capture probe to provide or reconstitute a nickase recognition site, i.e. a cleavage helper probe. Thus, contact with a nickase enzyme will result in cleavage of the cleavage domain thereby releasing the capture probe from the solid substrate. It will be evident that such cleavage helper probes could be used to provide or reconstitute cleavage recognition sites for other cleavage enzymes, e.g. restriction enzymes, as discussed above.

Thus, in some embodiments, the method may also comprise a step of contacting the solid substrate with a cleavage helper probe, i.e. under conditions that allow the cleavage helper probe to hybridise to the cleavage domain of the capture probe to provide a cleavage recognition site, e.g. a nickase enzyme recognition site or a restriction enzyme recognition site.

Some nickases introduce single-stranded nicks only at particular sites on a DNA molecule, by binding to and recognizing a particular nucleotide recognition sequence. A number of naturally-occurring nickases have been discovered, of which at present the sequence recognition properties have been determined for at least four. Nickases are described in U.S. Patent No. 6,867,028, which is herein incorporated by reference in its entirety and any suitable nickase may be used in the methods of the invention.

In embodiments that utilise a nickase enzyme, the nickase enzyme may be removed from the assay or inactivated prior to step (d) to prevent unwanted cleavage of the extended probes or amplicons thereof.

Thus, in a preferred embodiment, “means for releasing said capture probes from the surface of a solid substrate” comprises a cleavage enzyme, such as an enzyme as defined above. It will be evident from the description above that a cleavage enzyme may refer to a combination or mixture of enzymes that have cleavage activity. For instance, in a preferred embodiment, the cleavage enzyme comprises a mixture of Uracil DNA glycosylase (UDG) and a DNA glycosylase-lyase, such as Endonuclease VIII. A commercially available mixture of these enzymes, known as the USER™ enzyme, is particularly preferred. Thus, in some embodiments, the cleavage domain comprises a poly-U sequence.

As mentioned above, the cleavage enzyme is contacted with the array at the same time as, or before, the polymerase enzyme, preferably in a single reaction

mixture. Accordingly, in some embodiments the cleavage enzyme must be functional in the same conditions as the polymerase enzyme, e.g. functional in the same buffer, salt, temperature conditions used for the polymerase enzyme.

By "functional" is meant that the cleavage enzyme may show some reduced activity in cleaving the cleavage domain of the capture probe relative to the activity of the cleavage enzyme on the cleavage domain in conditions that are optimum for the enzyme, e.g. in the buffer, salt and temperature conditions recommended by the manufacturer. Thus, the cleavage enzyme may be considered to be functional if it has at least 50%, e.g. at least 60, 70, 80, 85, 90, 95, 96, 97, 98, 99 or 100%, activity relative to the activity of the cleavage enzyme in conditions that are optimum for the enzyme.

Additionally or alternatively, in some embodiments the polymerase enzyme must be functional in the same conditions as the cleavage enzyme, e.g. functional in the same buffer, salt, temperature conditions used for the cleavage enzyme.

By "functional" is meant that the polymerase enzyme may show some reduced activity in target templated extension of the capture probe relative to the activity of the polymerase enzyme on the same target templated extension in conditions that are optimum for the enzyme, e.g. in the buffer, salt and temperature conditions recommended by the manufacturer. Thus, the polymerase enzyme may be considered to be functional if it has at least 50%, e.g. at least 60, 70, 80, 85, 90, 95, 96, 97, 98, 99 or 100%, activity relative to the activity of the polymerase enzyme in conditions that are optimum for the enzyme.

Whilst it will be evident from the discussion herein that the method of the invention may be used to capture DNA (e.g. genomic DNA) and RNA, for the sake of simplicity, the invention will be described in detail in the context of capturing RNA, which forms a preferred aspect of the invention. However, particular embodiments in relation to the capture of DNA are presented below.

The RNA may be any RNA molecule which may occur in a cell. Thus it may be mRNA, tRNA, rRNA, viral RNA, small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), microRNA (miRNA), small interfering RNA (siRNA), piwi-interacting RNA (piRNA), ribozymal RNA, antisense RNA or non-coding RNA. Preferably however it is mRNA.

Step (c) in the method above involves extending the capture probes using the nucleic acid molecules hybridised to the capture probes (i.e. "captured" by the capture probes) as extension templates to produce extended probes thereby

spatially tagging the nucleic acids of the biological specimen. Thus, in the context of RNA, this aspect of step (c) may be viewed as generating cDNA (particularly tagged cDNA) from the captured RNA, i.e. relating to the synthesis of the cDNA. This will involve a step of reverse transcription (RT) of the captured RNA, extending the capture probe, which functions as the RT primer, using the captured RNA as template. Such a step generates so-called first strand cDNA, i.e. an extended probe. As will be described in more detail below, second strand cDNA synthesis may take place in a separate step, prior to the step of analyzing the extended probes (e.g. the sequence of the extended probes) or may take place as part of the analysis step. Thus, for instance, second strand synthesis may occur in the first step of amplification of a released first strand cDNA molecule. In some embodiments, second strand synthesis may occur contemporaneously with the first strand synthesis or may be performed immediately following the first strand synthesis reaction. For instance, second strand synthesis may occur contemporaneously with the first strand synthesis reaction when a template switching reaction is used for second strand synthesis. Template switching reactions are described in detail below.

Thus, in some embodiments (i.e. when the method is used to capture RNA), the extension reaction is performed using a reverse transcriptase enzyme, i.e. the polymerase enzyme is a reverse transcriptase enzyme. The desired reverse transcriptase activity may be provided by one or more distinct reverse transcriptase enzymes, wherein suitable examples are: M-MLV, MuLV, AMV, HIV, ArrayScript™, MultiScribe™, ThermoScript™, and SuperScript® I, II, and III enzymes. As used herein, the term "reverse transcriptase" includes not only naturally occurring enzymes but also all such modified derivatives, including also derivatives of naturally occurring reverse transcriptase enzymes.

Particularly preferred reverse transcriptase enzymes for use in the methods of the present application include M-MLV, MuLV, AMV and HIV reverse transcriptase enzymes and derivatives, e.g. sequence-modified derivatives, or mutants thereof.

Sequence-modified derivatives or mutants of M-MLV, MuLV, AMV and HIV reverse transcriptase enzymes include mutants that retain at least some of the functional, e.g. reverse transcriptase, activity of the wild-type sequence. Mutations may affect the activity profile of the enzymes, e.g. enhance or reduce the rate of polymerisation, under different reaction conditions, e.g. temperature, template

concentration, primer concentration etc. Mutations or sequence-modifications may also affect the RNase activity and/or thermostability of the enzyme. The reverse transcriptase enzyme may be provided as part of a composition which comprises other components, e.g. stabilizing components, that enhance or improve the activity of the reverse transcriptase enzyme, such as RNase inhibitor(s), inhibitors of DNA-dependent DNA synthesis, e.g. actinomycin D. Many sequence-modified derivatives or mutants of reverse transcriptase enzymes, e.g. M-MLV, and compositions comprising unmodified and modified enzymes are known in the art and are commercially available, e.g. ArrayScript™, MultiScribe™, ThermoScript™, and SuperScript® I, II, III and IV enzymes, and all such enzymes are considered to be useful in the methods of the invention.

Hence, in one embodiment the preferred reverse transcriptase enzyme of the invention is encoded by a nucleotide sequence comprising SEQ ID NO: 1 (M-MLV reverse transcriptase) or a nucleotide sequence that has at least 80, 85 or 90% sequence identity thereto. Alternatively or additionally, the preferred reverse transcriptase enzyme of the invention comprises the polypeptide sequence of SEQ ID NO: 2 or a polypeptide sequence that has at least 80, 85 or 90% sequence identity thereto. Thus, the reverse transcriptase enzyme of the invention may be a natural variant or derivative of the M-MLV reverse transcriptase enzyme, e.g. from another strain of M-MLV.

Preferably said nucleotide or polypeptide sequence is at least 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99% identical to the sequence to which it is compared.

Sequence identity of nucleotide molecules may be determined by, e.g. FASTA Search using GCG packages, with default values and a variable pamfactor, and gap creation penalty set at 12.0 and gap extension penalty set at 4.0 with a window of 6 nucleotides.

Preferably such sequence identity related nucleic acid molecules are functionally equivalent to the nucleic acid molecule which is set forth in SEQ ID NO: 1. Such functionally equivalent nucleic acid molecules may take the form of derivatives and are considered functionally equivalent if they encode polypeptides which would be considered functional equivalents according to the reverse transcriptase activity tests known in the art. Preferred functional equivalents are those which encode the preferred polypeptides as set out above.

Sequence identity of polypeptide molecules may be determined by, e.g. using the SWISS-PROT protein sequence databank using FASTA pep-cmp with a

variable pamfactor, and gap creation penalty set at 12.0 and gap extension penalty set at 4.0, and a window of 2 amino acids. Preferably said comparison is made over the full length of the sequence, but may be made over a smaller window of comparison, e.g. less than 600, 500, 400, 300, 200, 100 or 50 contiguous amino acids.

Preferably such sequence identity related polypeptides are functionally equivalent to the polypeptide which is set forth in SEQ ID NO: 2. As such, the polypeptide with a sequence as set forth in SEQ ID NO: 2 may be modified without affecting the sequence of the polypeptide.

Modifications that do not affect the sequence of the polypeptide include, e.g. chemical modification, including by deglycosylation or glycosylation. Such polypeptides may be prepared by post-synthesis/isolation modification of the polypeptide without affecting functionality, e.g. certain glycosylation, methylation etc. of particular residues.

As referred to herein, to achieve "functional equivalence" the polypeptide may show some increased or reduced efficacy in reverse transcriptase activity relative to the parent molecule (i.e. the molecule from which it was derived, e.g. by amino acid substitution), but preferably is as efficient or is more efficient. In some embodiments, part of the native enzyme activity may be reduced or inhibited in the functionally equivalent polypeptide, e.g. RNase activity. Thus, functional equivalence relates to a polypeptide which has reverse transcriptase activity capable of extending the capture domains of the capture probes in a RNA templated extension reaction. This may be tested by comparison of the reverse transcriptase activity of the derivative polypeptide relative to the polypeptide from which it is derived in a quantitative manner, as described above. The derivative is preferably at least 30, 50, 70 or 90% as effective as the parent polypeptide in the methods of the invention.

Functionally-equivalent proteins which are related to or derived from the naturally-occurring protein, may be obtained by modifying the native amino acid sequence by single or multiple amino acid substitution, addition and/or deletion (providing they satisfy the above-mentioned sequence identity requirements), but without destroying the molecule's function. Preferably the native sequence has less than 20 substitutions, additions or deletions, e.g. less than 10, 5, 4, 3, 2, or 1 such modifications. Such proteins are encoded by "functionally-equivalent nucleic acid

molecules" which are generated by appropriate substitution, addition and/or deletion of one or more bases.

As used herein the term "multiple" means two or more, or at least two, e.g. 3, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 400, 500, 1000, 2000, 5000, 10,000, 100,000, 1000,000, 10,000,000, 15,000,000 or more etc. Thus for example, the number of capture probes may be any integer in any range between any two of the aforementioned numbers.

As discussed above, the capture probes are immobilised on a solid substrate and thus the solid substrate may therefore be viewed as an array or comprising one or more arrays, e.g. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more arrays. For instance, a solid substrate may comprise 20, 25, 30, 35, 40 or more arrays. Arrays for use in the context of nucleic acid analysis in general, and DNA analysis in particular, are discussed and described below. Specific details and embodiments described herein in relation to arrays and capture probes for use in the context of RNA, apply equally (where appropriate) to all such arrays, including those for use with DNA. In this respect, it will be appreciated however that it is envisaged that conventional-type arrays comprising capture probes as defined herein may be used in the methods of the invention.

Thus, the methods outlined herein may utilise high density nucleic acid arrays comprising "capture probes" as defined herein for capturing and tagging transcripts from all of the single cells within a biological specimen, e.g. tissue sample, such as a thin tissue sample slice, or "section". Tissue samples or sections for analysis may be produced in a highly parallelized fashion, such that the spatial information in the section is retained. The captured RNA (preferably mRNA) molecules for each cell, or "transcriptomes", are transcribed into cDNA and the resultant cDNA molecules are analysed, for example by high throughput sequencing. The resultant data may be correlated to images of the biological specimen, e.g. tissue samples through so-called positional domains, e.g. unique nucleotide sequences within the capture probes, e.g. barcode sequences or ID tags).

High density nucleic acid arrays or microarrays are a core component of the spatial transcriptome and genome labelling and tagging methods described herein. A microarray is a multiplex technology used in molecular biology. A typical microarray consists of an arrayed series of microscopic spots (features or sites) of oligonucleotides (millions of spots can be incorporated on a single array and

several arrays can be presented on a single solid substrate). The distinct (discrete) position of each nucleic acid (oligonucleotide) spot (each species of oligonucleotide/nucleic acid molecule) is known as a "feature" or “site” (and hence in the methods set out above each species of capture probe may be viewed as a specific feature of the array; each feature occupies a distinct (discrete) position on the array), and typically each separate feature contains in the region of picomoles (10^{-12} moles) of a specific DNA sequence (a "species"), which are known as "probes" (or "reporters"). Typically, these can be a short section of a gene or other nucleic acid element to which a cDNA or cRNA sample (or "target") can hybridize under high-stringency hybridization conditions. However, as described below, the probes of the present invention differ from the probes of standard microarrays.

Thus, the term "array" refers to a population of nucleic acid features or sites on a solid substrate that can be differentiated from each other according to relative location. Different nucleic acid molecules that are at different sites or features of an array can be differentiated from each other according to the locations of the sites or features in the array. An individual site or feature of an array can include one or more molecules of a particular type (e.g. species of capture probe). For example, a site or feature can include a single nucleic acid molecule having a particular sequence or a site can include several nucleic acid molecules having the same sequence.

The capture probes may be attached to the solid substrate, e.g. array, of the invention by any suitable means. As used herein, the term “attached” refers to the state of two things being joined, fastened, adhered, connected or bound to each other. For example, a nucleic acid can be attached to a material, such as a gel or solid support, by a covalent or non-covalent bond. A covalent bond is characterized by the sharing of pairs of electrons between atoms. A non-covalent bond is a chemical bond that does not involve the sharing of pairs of electrons and can include, for example, hydrogen bonds, ionic bonds, van der Waals forces, hydrophilic interactions and hydrophobic interactions.

In a preferred embodiment the probes are immobilized to the solid substrate, e.g. array, by chemical immobilization, i.e. by a covalent bond. This may be an interaction between the solid substrate and the probe based on a chemical reaction. Such a chemical reaction typically does not rely on the input of energy via heat or light, but can be enhanced by either applying heat, e.g. a certain optimal temperature for a chemical reaction, or light of certain wavelength. For example, a

chemical immobilization may take place between functional groups on the substrate and corresponding functional elements on the probes. Such corresponding functional elements in the probes may either be an inherent chemical group of the probe, e.g. a hydroxyl group or be additionally introduced. An example of a functional group introduced to the capture probe is an amine group. Typically, the probe to be immobilized comprises a functional amine group or is chemically modified in order to comprise a functional amine group. As discussed above, in some embodiments the probe may comprise a cleavable chemical linker which attaches the probe to the solid substrate, such that it forms a cleavage domain of the capture probe. Means and methods for such a chemical modification are well known.

The localization of said functional group within the probe to be immobilized may be used in order to control and shape the binding behaviour and/or orientation of the probe, e.g. the functional group may be placed at the 5' or 3' end of the probe or within sequence of the probe. A typical substrate on which a probe may be immobilized comprises moieties which are capable of binding to such probes, e.g. to amine-functionalized nucleic acids. Examples of such substrates are carboxy, aldehyde or epoxy substrates. Such materials are known to the person skilled in the art. Functional groups, which impart a connecting reaction between probes which are chemically reactive by the introduction of an amine group, and array substrates are known to the person skilled in the art.

Alternative substrates on which probes may be immobilized may have to be chemically activated, e.g. by the activation of functional groups, available on the solid substrate. The term "activated substrate" relates to a material in which interacting or reactive chemical functional groups were established or enabled by chemical modification procedures as known to the person skilled in the art. For example, a substrate comprising carboxyl groups has to be activated before use. Furthermore, there are substrates available that contain functional groups that can react with specific moieties already present in the nucleic acid probes.

In some embodiments, the probes may be synthesized directly on the solid substrate. Suitable methods for such an approach are known to the person skilled in the art. For instance, features (i.e. clusters of probes) may be made by a solid-phase amplification method known as bridge amplification, as discussed in more detail below. Other examples including manufacture techniques developed by Agilent Inc., Affymetrix Inc., Roche Nimblegen Inc. or Flexgen BV. Typically, lasers

and a set of mirrors that specifically activate the spots where nucleotide additions are to take place are used. Such an approach may provide, for example, spot sizes (i.e. features) of around 30 µm or larger.

The solid substrate therefore may be any suitable substrate known to the person skilled in the art, i.e. suitable for use as an array. The substrate may have any suitable form or format, e.g. it may be flat, curved, e.g. convexly or concavely curved towards the area where the interaction between the biological specimen, e.g. tissue sample, and the substrate takes place. Particularly preferred is the where the substrate is a flat, i.e. planar, chip or slide. In some embodiments, the substrate may comprise various structures such as wells, beads (or other particles) in or on a substrate, projections from the substrate, ridges on the substrate or channels in a substrate. In some embodiments, the structures of a substrate, e.g. wells or beads, can each bear a different capture probe. Different capture probes attached to each structure can be identified according to the locations of the structures in or on the surface of the substrate. Exemplary arrays in which separate structures are located on the substrate include, without limitation, those having beads in wells.

The term “solid substrate” typically refers to a rigid support that is insoluble in aqueous liquid and which allows for an accurate and traceable positioning of the probes on the substrate. An example of a substrate is a solid material or a substrate comprising functional chemical groups, e.g. amine groups or amine-functionalized groups. The substrate can be non-porous or porous. Thus, the substrate can optionally be capable of taking up a liquid (e.g. due to porosity) but will typically be sufficiently rigid that the substrate does not swell substantially when taking up the liquid and does not contract substantially when the liquid is removed by drying. A non-porous solid support is generally impermeable to liquids or gases. Exemplary solid supports include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon™, cyclic olefins, polyimides etc.), nylon, ceramics, resins, Zeonor, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, optical fiber bundles, and polymers, such as polystyrene, cyclic olefin copolymers (COCs), cyclic olefin polymers (COPs), polypropylene, polyethylene and polycarbonate. A preferred substrate envisaged by the present invention is a non-

porous substrate. In some embodiments, the solid substrate is located within a flow cell apparatus.

Any suitable material known to the person skilled in the art may be used. Typically, glass or polystyrene is used. Polystyrene is a hydrophobic material suitable for binding negatively charged macromolecules because it normally contains few hydrophilic groups. For nucleic acids immobilized on glass slides, it is furthermore known that by increasing the hydrophobicity of the glass surface the nucleic acid immobilization may be increased. Such an enhancement may permit a relatively more densely packed formation. In addition to a coating or surface treatment with poly-L-lysine, the substrate, in particular glass, may be treated by silanation, e.g. with epoxy-silane or amino-silane or by silynation or by a treatment with polyacrylamide.

A number of standard arrays are commercially available and both the number and size of the features may be varied. In the present invention, the arrangement of the features may be altered to correspond to the size and/or density of the cells present in different biological specimens, e.g. tissues or organisms. For instance, animal cells typically have a cross-section in the region of 1-100µm, whereas the cross-section of plant cells typically may range from 1-10000µm. Hence, Nimblegen® arrays, which are available with up to 2.1 million features, or 4.2 million features, and feature sizes of 13 micrometers, may be preferred for tissue samples from an animal or fungus, whereas other formats, e.g. with 8x130k features, may be sufficient for plant tissue samples. Commercial arrays are also available or known for use in the context of sequence analysis and in particular in the context of NGS technologies. Such arrays may also be used as the array surface in the context of the present invention e.g. an Illumina bead array or flow cell, on which the probes are produced using bridge amplification. In addition to commercially available arrays, which can themselves be customized, it is possible to make custom or non-standard "in-house" arrays and methods for generating arrays are well-established. The methods of the invention may utilise both standard and non-standard arrays that comprise probes as defined below.

The term “flow cell” is intended to mean a vessel having a chamber where a reaction can be carried out, an inlet for delivering reagents to the chamber and an outlet for removing reagents from the chamber. In some embodiments the chamber is configured for detection of the reaction that occurs in the chamber. For example, the chamber can include one or more transparent surfaces allowing optical

detection of biological specimens, optically labelled molecules, or the like in the chamber. Exemplary flow cells include, but are not limited to those used in a nucleic acid sequencing apparatus such as flow cells for the Genome Analyzer®, MiSeq®, NextSeq® or HiSeq® platforms commercialized by Illumina, Inc. (San Diego, CA); or for the SOLiD™ or Ion Torrent™ sequencing platform commercialized by Life Technologies (Carlsbad, CA). Exemplary flow cells and methods for their manufacture and use are also described, for example, in WO 2016/162309, WO 2014/142841, U.S. Pat. App. Pub. No. 2010/0111768 and U.S. Pat. No. 8,951,781, each of which is incorporated herein by reference.

The term "bead array" refers to a solid support comprising a collection of beads or other particles. The particles can be suspended in a solution or they can be located on the surface of a solid substrate. Examples of arrays having beads located on a surface include those wherein beads are located in wells such as a BeadChip array (Illumina Inc., San Diego CA), substrates used in sequencing platforms from 454 LifeSciences (a subsidiary of Roche, Basel Switzerland) or substrates used in sequencing platforms from Ion Torrent (a subsidiary of Life Technologies, Carlsbad California). Other solid supports having beads located on a surface are described in US Pat. Nos. 6,266,459; 6,355,431; 6,770,441; 6,859,570; 6,210,891; 6,258,568; or 6,274,320; US Pat. App. Publ. Nos. 2009/0026082 A1; 2009/0127589 A1; 2010/0137143 A1; or 2010/0282617 A1 or PCT Publication No. WO 00/63437 and WO 2016/162309, each of which is incorporated herein by reference. Several of the above references describe methods for attaching nucleic acid probes to beads prior to loading the beads in or on a solid support. As such, the collection of beads can include different beads each having a unique capture probe attached. It will however, be understood that the beads can be made to include universal primers, and the beads can then be loaded onto an array, thereby enabling the synthesis of capture probes on the array, e.g. using bridge amplification methods. As set forth previously herein, the solid supports typically used for bead arrays can be used without beads. For example, nucleic acids, such as capture probes, can be attached directly to the wells or to gel material in wells. Thus, the above references are illustrative of materials, compositions or apparatus that can be modified for use in the methods and compositions set forth herein.

Accordingly, a solid substrate for use in the method of the invention can include an array of beads, wherein different capture probes are attached to different beads in the array. In this embodiment, each bead can be attached to a different

capture probe and the beads can be randomly distributed on the solid support in order to effectively immobilize the different capture probes to the solid support. Optionally, the solid support can include wells having dimensions that accommodate no more than a single bead. In such a configuration, the beads may be attached to the wells due to forces resulting from the fit of the beads in the wells. It is also possible to use attachment chemistries or adhesives to hold the beads in the wells.

The probes on a solid substrate, e.g. array, may be immobilized, i.e. attached or bound, to the array preferably via the 5' or 3' end, depending on the chemical matrix of the array. Typically, for commercially available arrays, the probes are attached via a 3' linkage, thereby leaving a free 5' end. However, arrays comprising probes attached to the substrate via a 5' linkage, thereby leaving a free 3' end, are available and may be synthesized using standard techniques that are well known in the art and are described elsewhere herein.

The covalent linkage used to couple a nucleic acid probe to a solid substrate, e.g. array, may be viewed as both a direct and indirect linkage, in that the although the probe is attached by a "direct" covalent bond, there may be a chemical moiety or linker separating the "first" nucleotide of the nucleic acid probe from the, e.g. glass or silicon, substrate, i.e. an indirect linkage. For the purposes of the present invention probes that are immobilized to the substrate by a covalent bond and/or chemical linker are generally seen to be immobilized or attached directly to the substrate. As mentioned above, in some embodiments the chemical linker is cleavable, e.g. enzymatically and/or chemically, such that it forms the cleavage domain of the capture probe.

As will be described in more detail below, the capture probes of the invention may be immobilized on, or interact with, the solid substrate, e.g. array, directly or indirectly. Thus the capture probes need not bind directly to the array, but may interact indirectly, for example by binding to a molecule or entity (e.g. bead) which itself binds directly or indirectly to the solid substrate, e.g. array (e.g. the capture probe may interact with (e.g. bind or hybridize to) a binding partner for the capture probe, i.e. a surface probe, which is itself bound to the array directly or indirectly). Generally speaking, however, the capture probe will be, directly or indirectly (by one or more intermediaries), bound to, or immobilized on, the solid substrate, e.g. array.

The method of the invention may comprise probes that are immobilized via their 5' or 3' end. However, when the capture probe is immobilized directly to the solid substrate, e.g. it may be preferred that it is immobilized such that the 3' end of the capture probe is free to be extended, e.g. it is immobilized by its 5' end. In some embodiments, the capture probe may be immobilized indirectly, such that it has a free, i.e. extendible, 3' end.

By extended or extendible 3' end, it is meant that further nucleotides may be added to the most 3' nucleotide of the nucleic acid molecule, e.g. capture probe, to extend the length of the nucleic acid molecule, i.e. the standard polymerization reaction utilized to extend nucleic acid molecules, e.g. templated polymerization catalyzed by a polymerase, e.g. reverse transcriptase.

Thus, in one embodiment, the array comprises probes that are immobilized directly via their 3' end, so-called surface probes, which are defined below. Each species of surface probe comprises a region of complementarity to each species of capture probe, such that the capture probe may hybridize to the surface probe, resulting in the capture probe comprising a free extendible 3' end. However, it will be evident that, in some embodiments, the surface probes may be immobilized directly via their 5' end, such that cleavage of the cleavage domain is required to release the 3' end of the capture probe, i.e. to expose the capture domain for extension. In a preferred aspect of the invention, when the solid substrate, e.g. array, comprises surface probes, the capture probes are synthesized *in situ* on the array.

The probes may be made up of ribonucleotides and/or deoxyribonucleotides as well as synthetic nucleotide residues that are capable of participating in Watson-Crick type or analogous base pair interactions. Thus, the nucleic acid domain may be DNA or RNA or any modification thereof e.g. PNA or other derivatives containing non-nucleotide backbones. However, in the context of transcriptome analysis the capture domain of the capture probe must capable of priming a reverse transcription reaction to generate cDNA that is complementary to the captured RNA molecules. As described below in more detail, in the context of genome analysis, the capture domain of the capture probe must be capable of binding to the DNA fragments, which may comprise binding to a binding domain that has been added to the fragmented DNA. In some embodiments, the capture domain of the capture probe may prime a DNA extension (polymerase) reaction to generate DNA that is complementary to the captured DNA molecules.

In a preferred embodiment of the invention at least the capture domain of the capture probe comprises or consists of deoxyribonucleotides (dNTPs). In a particularly preferred embodiment the whole of the capture probe comprises or consists of deoxyribonucleotides.

In a preferred embodiment of the invention the capture probes are immobilized on the substrate of the array directly, i.e. by their 5' end, resulting in a free extendible 3' end.

The capture probes of the invention comprise at least three domains, a capture domain and a positional domain (or a feature identification tag or domain; the positional domain may alternatively be defined as an identification (ID) domain or tag, or as a positional tag) and a cleavage domain, as defined above. The capture probe may further comprise an amplification domain as defined further below. Where the capture probe is indirectly attached to the array surface via hybridization to a surface probe, the surface probe requires a sequence (e.g. a portion or domain) which is complementary to the capture probe. For instance, when the surface probe is immobilized on the solid substrate via its 3' end, such a complementary sequence may be complementary to a positional/identification domain and/or an amplification domain and/or cleavage domain on the capture probe. In other words the positional domain and/or amplification domain and/or cleavage domain may constitute the region or portion of the capture probe which is complementary to the surface probe. However, the capture probe may also comprise an additional domain (or region, portion or sequence) which is complementary to the surface probe.

Alternatively, when the surface probe is immobilized on the solid substrate via its 5' end, such a complementary sequence may be complementary to a cleavage domain and/or capture domain and/or positional/identification domain and/or an amplification domain the capture probe, preferably the cleavage domain. In other words the cleavage domain and/or capture domain and/or positional/identification domain and/or an amplification domain may constitute the region or portion of the capture probe which is complementary to the surface probe. However, the capture probe may also comprise an additional domain (or region, portion or sequence) which is complementary to the surface probe.

The capture domain is typically located at the 3' end of the capture probe and comprises a free 3' end that can be extended by template dependent polymerization. However, it will be evident that in some embodiments, e.g. when

the capture probe is immobilized directly or indirectly by its 3' end, the cleavage domain may be located at the 3' end and the capture domain is located directly or indirectly upstream of (i.e. at the 5' end of) the cleavage domain. The capture domain comprises a nucleotide sequence that is capable of hybridizing to nucleic acid, e.g. RNA (preferably mRNA), present in the cells of the biological specimen, e.g. tissue sample, contacted with the solid substrate, e.g. array.

Advantageously, the capture domain may be selected or designed to bind (or put more generally may be capable of binding) selectively or specifically to the particular nucleic acid, e.g. RNA, it is desired to detect or analyse. For example the capture domain may be selected or designed for the selective capture of mRNA. As is well known in the art, this may be on the basis of hybridisation to the poly-A tail of mRNA. Thus, in a preferred embodiment the capture domain comprises a poly-T oligonucleotide, i.e. a series of consecutive deoxythymidine residues linked by phosphodiester bonds, which is capable of hybridizing to the poly-A tail of mRNA. Alternatively, the capture domain may comprise nucleotides which are functionally or structurally analogous to poly-T, i.e. are capable of binding selectively to poly-A, for example a poly-U oligonucleotide or an oligonucleotide comprised of deoxythymidine analogues or combinations thereof, wherein said oligonucleotide retains the functional property of binding to poly-A. In a particularly preferred embodiment the capture domain, or more particularly the poly-T and/or -U element of the capture domain, comprises at least 10 nucleotides, preferably at least 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotides. In a further embodiment, the capture domain, or more particularly the poly-T and/or -U element of the capture domain comprises at least 25, 30 or 35 nucleotides.

Random sequences may also be used in the capture of nucleic acid, as is known in the art, e.g. random hexamers or similar sequences, and hence such random sequences may be used to form all or a part of the capture domain. For example, random sequences may be used in conjunction with poly-T (or poly-T analogue etc.) sequences. Thus, where a capture domain comprises a poly-T (or a "poly-T-like") oligonucleotide, it may also comprise a random oligonucleotide sequence. This may for example be located 5' or 3' of the poly-T sequence, e.g. at the 3' end of the capture probe, but the positioning of such a random sequence is not critical. Such a construct may facilitate the capturing of the initial part of the poly-A of mRNA. Alternatively, the capture domain may be an entirely random

sequence. Degenerate capture domains may also be used, according to principles known in the art.

The capture domain may be capable of binding selectively to a desired sub-type or subset of nucleic acid, e.g. RNA, for example a particular type of RNA such mRNA or rRNA etc. as listed above, or to a particular subset of a given type of RNA, for example, a particular mRNA species, e.g. corresponding to a particular gene or group of genes. Such a capture probe may be selected or designed based on sequence of the RNA it is desired to capture. Thus it may be a sequence-specific capture probe, specific for a particular RNA target or group of targets (target group etc). Thus, it may be based on a particular gene sequence or particular motif sequence or common/conserved sequence etc., according to principles well known in the art.

In embodiments where the capture domain comprises a sequence specific for a particular target gene or group of genes it may be desirable to synthesise the capture probes on the solid substrate directly, e.g. using ligation reactions. In particular, it may be useful to synthesise the capture domain of the capture probe on the solid substrate directly.

For instance, a solid substrate may be provided comprising capture probes immobilized by their 5' end in which the capture domain is replaced with a universal domain that is common to all of the probes on the solid substrate (such probes may be viewed as a positional domain oligonucleotide). The substrate is contacted with one or more ligation helper probes that comprise 5' to 3':

(i) a domain that is complementary to a sequence at the 5' of the capture domain; and

(ii) a domain that is complementary to a universal domain in the probes immobilized on the solid substrate,

under conditions that allow the ligation helper probe to hybridise to the probes immobilized on the solid substrate. The substrate is also contacted with one or more species of probe containing a capture domain (such probes may be viewed as a capture domain oligonucleotide), wherein the probes comprise 5' to 3':

(i) a domain that is complementary to a sequence at the 5' of the helper probe; and

(ii) a capture domain comprising a sequence specific for a particular target gene,

under conditions that allow the probes comprising the capture domain to hybridise to the helper probes.

Finally the substrate is contacted with components (e.g. a ligase enzyme and appropriate buffers, salts etc.) that facilitate the ligation of the probes containing the capture domain to the probes immobilized on the solid substrate using the ligation helper probes as a ligation template to form capture probes. Thus, if one species of probe containing the capture domain is ligated to the probes immobilized on the solid substrate, each feature of the array will be capable of capturing a single target nucleic acid from the biological specimen, e.g. to analyse the expression of a single gene. However, if more than one species of probe containing the capture domain is ligated to the probes immobilized on the solid substrate (i.e. each species of probe contains a capture domain comprising a sequence specific for a different target gene), each feature of the array will be capable of capturing multiple target nucleic acids from the biological specimen, e.g. to analyse the expression of a subset of genes. It will be evident that the ligation helper probes, probes containing the capture domain and components that facilitate the ligation of the probes may be contacted with the solid substrate separately (e.g. sequentially) or simultaneously.

Whilst a solid substrate comprising capture probes that are specific for a single gene or groups of genes may be produced by any suitable means, it will be evident that the method described above provides a convenient method of generating multiple gene specific substrates based on a single array platform. In this respect, a common or universal substrate comprising capture probes in which the capture domain is replaced with a universal domain that is common to all of the probes on the solid substrate may be manufactured and used to generate multiple substrates each specific for a particular target gene or group of target genes. Such substrates may find particular utility in diagnostic assays, e.g. for analysing the expression of genes or groups of genes that are associated with one or more diseases, e.g. cancer.

Thus, step (a) of the method described above may comprise:

(1) providing a solid substrate on which multiple species of probes are immobilized such that each species occupies a distinct position on the solid substrate and is oriented to have a free 3' end, wherein said probe is for a templated ligation reaction, wherein each species of said probe comprises a nucleic acid molecule with 5' to 3':

(i) a cleavage domain for releasing the capture probe from the surface of the solid substrate,

(ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and optionally

(iii) a universal domain;

(2) contacting said solid substrate with:

(i) a ligation helper probe as defined above;

(ii) one or more species of probe containing a capture domain as defined above; and

(iii) components that facilitate the ligation of the probes in (ii) to the probes in (1) using the helper probe in (2)(i) as a ligation template,

under suitable conditions, e.g. conditions that suitable form hybridisation of the complementary domains of the probes and the ligation of the probes in (ii) to the probes in (1) to form capture probes, which are for a primer extension reaction.

Suitable enzymes and conditions for performing ligation reactions are discussed in detail below and may be applied to the method described above.

The positional domain (feature identification domain or tag) of the capture probe is located directly or indirectly upstream, i.e. closer to the 5' end of the capture probe nucleic acid molecule, of the capture domain. Preferably the positional domain is directly adjacent to the capture domain, i.e. there is no intermediate sequence between the capture domain and the positional domain. In some embodiments the positional domain forms the 5' end of the capture probe. For instance, the positional domain may form the 5’ end of the capture probe when the cleavage domain is a cleavable cross-linker attaching the capture probe to the solid substrate. Additionally or alternatively, the positional domain may form the 5’ end of the capture probe when the capture probe is immobilized to the substrate via its 3’ end.

As discussed above, each feature (distinct position) of the solid substrate, e.g. array, comprises a spot of a species of nucleic acid probe, wherein the positional domain at each feature (or groups of adjacent features) is unique. Thus, a "species" of capture probe is defined with reference to its positional domain; a single species of capture probe will have the same positional domain. However, it is not required that each member of a species of capture probe has the same sequence in its entirety. In particular, since the capture domain may be or may comprise a random or degenerate sequence, the capture domains of individual

probes within a species may vary. Accordingly, in some embodiments where the capture domains of the capture probes are the same, each feature comprises a single probe sequence. However, in other embodiments where the capture probe varies, members of a species of probe will not have the exact same sequence, although the sequence of the positional domain of each member in the species will be the same. What is required is that each feature or site (or groups of adjacent features or sites) of the solid substrate, e.g. array, carries a capture probe of a single species (specifically each feature or position carries a capture probe which has an identical positional tag, i.e. there is a single positional domain at each feature or position or groups of adjacent features or positions). Each species has a different positional domain which identifies the species. However, each member of a species, may in some cases, as described in more detail herein, have a different capture domain, as the capture domain may be random or degenerate or may have a random or degenerate component or, as described above, capture domains specific for subsets of genes may be produced at each feature. This means that within a given feature, or position, the capture domain of the probes may differ.

Thus in some, but not necessarily in all embodiments, the nucleotide sequence of any one probe molecule immobilized at a particular feature is the same as the other probe molecules immobilized at the same feature, but the nucleotide sequence of the probes at each feature (or groups of adjacent features) is different, distinct or distinguishable from the probes immobilized at every other feature. Preferably each feature comprises a different species of probe. However, in some embodiments it may be advantageous for a group of adjacent features to comprise the same species of probe, i.e. effectively to produce a feature covering an area of the array that is greater than a single feature, e.g. to lower the resolution of the array. In other embodiments, the nucleotide sequence of the positional domain of any one probe molecule immobilized at a particular feature may be the same as the other probe molecules immobilized at the same feature but the capture domain may vary. The capture domain may nonetheless be designed to capture the same type of molecule, e.g. mRNA, in general. Thus, for instance, where the features comprise capture probes containing multiple capture domains, e.g. capture domains each specific for a particular gene, each feature is capable of capturing the same target nucleic acids.

The positional domain (or tag) of the capture probe comprises the sequence which is unique to each feature (or group of adjacent features) and acts as a

positional or spatial marker (the identification tag). In this way each region or domain of the biological specimen, e.g. tissue sample, e.g. each cell in the tissue, will be identifiable by spatial resolution across the solid substrate, e.g. array, linking the nucleic acid, e.g. RNA (e.g. the transcripts) from a certain cell to a unique positional domain sequence in the capture probe. By virtue of the positional domain a capture probe in the array may be correlated to a position in the tissue sample, for example it may be correlated to a cell in the sample. Thus, the positional domain of the capture domain may be seen as a nucleic acid tag (identification tag).

Any suitable sequence may be used as the positional domain in the capture probes of the invention. By a suitable sequence, it is meant that the positional domain should not interfere with (i.e. inhibit or distort) the interaction between the nucleic acid, e.g. RNA, of the biological specimen, e.g. tissue sample, and the capture domain of the capture probe. For example, the positional domain should be designed such that nucleic acid molecules in the tissue sample do not hybridize specifically to the positional domain. Preferably, the nucleic acid sequence of the positional domain of the capture probes has less than 80% sequence identity to the nucleic acid sequences in the biological specimen, e.g. tissue sample. Preferably, the positional domain of the capture probe has less than 70%, 60%, 50% or less than 40% sequence identity across a substantial part of the nucleic acids molecules in the biological specimen, e.g. tissue sample. Sequence identity may be determined by any appropriate method known in the art, e.g. the using BLAST alignment algorithm.

In a preferred embodiment the positional domain of each species of capture probe contains a barcode sequence. The barcode sequences may be generated using random sequence generation. The randomly generated sequences may be followed by stringent filtering by mapping to the genomes of all common reference species and with pre-set Tm intervals, GC content and a defined distance of difference to the other barcode sequences to ensure that the barcode sequences will not interfere with the capture of the nucleic acid, e.g. RNA, from the tissue sample and will be distinguishable from each other without difficulty.

In some embodiments of the invention, as described elsewhere herein, it may be advantageous to amplify the extended probes, e.g. tagged cDNA molecules, produced by the methods of the invention. Thus, in some embodiments, the capture probe comprises an amplification domain, i.e. a domain that is common to all of the capture probes and facilitates the amplification of all of the extended

probes in a single reaction. The amplification domain of the capture probe is located directly or indirectly upstream, i.e. closer to the 5' end of the capture probe nucleic acid molecule, of the positional domain. In some embodiments, e.g. when the capture probe is immobilized on the array via its 5' end, the amplification domain may be located directly or indirectly downstream, i.e. closer to the 3' end of the capture probe nucleic acid molecule, of the cleavage domain, i.e. the amplification domain may be located between the cleavage domain and the positional domain. Thus, in some embodiments the amplification domain is directly adjacent to the positional domain, i.e. there is no intermediate sequence between the positional domain and the amplification domain.

However, in some embodiments, e.g. when the capture probe is immobilized on the array via its 3' end, the amplification domain may be located indirectly upstream, i.e. closer to the 5' end of the capture probe nucleic acid molecule, of the cleavage domain. Thus, in some embodiments where the capture probe comprises an amplification domain, the domain may form the 5' end of the capture probe. For instance, the amplification domain may form the 5' end of the capture probe when the cleavage domain is a cleavable cross-linker attaching the capture probe to the solid substrate, e.g. array. Alternatively, the amplification domain may form the 5' end of the capture probe when the capture probe is immobilized via its 5' end, e.g. when the cleavage domain is located at the 3' end of the capture probe.

It will be appreciated that amplification of the extended probes may take place on the solid substrate, e.g. array, i.e. the extended probes do not need to be removed or collected (e.g. pooled) from the substrate, e.g. transferred to a separate vessel or container, prior to performing the amplification step. Nevertheless, in some embodiments it may be convenient to remove or collect (e.g. pool) the extended probes prior to amplification (or second strand synthesis, as discussed below). In this respect, the method may comprise a step of pooling the extended probes from step (c) (e.g. to form a mixture of the extended probes). The step of pooling the extended probes may be viewed as removing, collecting and/or transferring the solution comprising the extended probes from the solid substrate, e.g. array, to a separate vessel or container, e.g. a PCR tube.

The amplification domain comprises a distinct sequence to which an amplification primer may hybridize. The amplification domain of the capture probe is preferably identical for each species of capture probe. Hence a single amplification

reaction will be sufficient to amplify all of the extended probes, e.g. tagged cDNA molecules.

Any suitable sequence may be used as the amplification domain in the capture probes of the invention. By a suitable sequence, it is meant that the amplification domain should not interfere with (i.e. inhibit or distort) the interaction between the nucleic acid, e.g. RNA, of the biological specimen, e.g. tissue sample, and the capture domain of the capture probe. Furthermore, the amplification domain should comprise a sequence that is not the same or substantially the same as any sequence in the nucleic acid, e.g. RNA, of the biological specimen, e.g. tissue sample, such that the primer used in the amplification reaction can hybridized only to the amplification domain under the amplification conditions of the reaction.

For example, the amplification domain should be designed such that nucleic acid molecules in the biological specimen, e.g. tissue sample, do not hybridize specifically to the amplification domain or the complementary sequence of the amplification domain. Preferably, the nucleic acid sequence of the amplification domain of the capture probes and the complement thereof has less than 80% sequence identity to the nucleic acid sequences in the tissue sample. Preferably, the positional domain of the capture probe has less than 70%, 60%, 50% or less than 40% sequence identity across a substantial part of the nucleic acid molecules in the tissue sample. Sequence identity may be determined by any appropriate method known in the art, e.g. the using BLAST alignment algorithm.

In one representative embodiment of the invention only the positional domain of each species of capture probe is unique. Hence, the capture domains, cleavage domains and amplification domains (if present) are in one embodiment the same for every species of capture probe for any particular array to ensure that the capture of the nucleic acid, e.g. RNA, from the biological specimen, e.g. tissue sample, release of the capture probes and amplification of the extended probes is uniform across the array. However, as discussed above, in some embodiments the capture domains may differ by virtue of including random or degenerate sequences or capture domains specific for subsets of specific genes.

In embodiments where the capture probe is immobilized on the substrate of the array indirectly, e.g. via hybridisation to a surface probe, the capture probe may be synthesised on the array as described below.

For example, the surface probes may be immobilized on the substrate of the array directly by or at, e.g. their 3' end. Each species of surface probe is unique to

each feature (or group of adjacent features) (distinct position) of the array and is partly complementary to the capture probe, defined above.

Hence, in some embodiments, the surface probe comprises at its 5' end a domain (complementary capture domain) that is complementary to a part of the capture domain that does not bind to the nucleic acid, e.g. RNA, of the biological specimen, e.g. tissue sample. In other words, it comprises a domain that can hybridize to at least part of a capture domain oligonucleotide. The surface probe further comprises a domain (complementary positional domain or complementary feature identification domain) that is complementary to the positional domain of the capture probe and a domain (complementary cleavage domain) that is complementary to the cleavage domain of the capture probe. The complementary positional domain is located directly or indirectly downstream (i.e. at the 3' end) of the complementary capture domain, i.e. there may be an intermediary or linker sequence separating the complementary positional domain and the complementary capture domain. The complementary cleavage domain is located directly or indirectly downstream (i.e. at the 3' end) of the complementary positional domain, i.e. there may be an intermediary or linker sequence separating the complementary cleavage domain and the complementary positional domain. When the capture probe comprises an amplification domain, the surface probe may also contain a domain (complementary amplification domain) that is complementary to the amplification domain and located directly or indirectly downstream (i.e. at the 3' end) of the complementary positional domain, i.e. there may be an intermediary or linker sequence separating the complementary amplification domain and the complementary positional domain. However, the complementary amplification domain will be located upstream (i.e. at the 5’ end) of the complementary cleavage domain, i.e. to ensure that the extended capture probe contains the amplification domain, the positional domain and the extended sequence, i.e. the sequence at the 3’ end of the capture domain.

In embodiments where the capture probe is synthesized on the solid substrate surface, e.g. array, the surface probes of the substrate, e.g. array, may comprise a domain (e.g. complementary amplification domain) at the 3' end of the surface probe, i.e. directly or indirectly downstream of the complementary cleavage domain, which is complementary to a domain (e.g. amplification domain) of the capture probe.

In some embodiments of the invention the sequence of the surface probe shows 100% complementarity or sequence identity to the positional, cleavage and amplification (if present) domains and to the part of the capture domain that does not bind to the nucleic acid, e.g. RNA, of the biological specimen, e.g. tissue sample. In other embodiments the sequence of the surface probe may show less than 100% sequence identity to the domains of the capture probe, e.g. less than 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91% or 90%. In a particularly preferred embodiment of the invention, the complementary cleavage domain shares less than 100% sequence identity to the cleavage domain of the capture probe.

As noted above, in some embodiments of the invention, the capture probe may be synthesized or generated on the substrate of the array.

In a representative embodiment, the array comprises surface probes as defined above. Oligonucleotides that correspond to the capture domain and at least the cleavage domain of the capture probe (optionally also including the amplification domain) are contacted with the solid substrate, e.g. array, and allowed to hybridize to the complementary domains of the surface probes. Excess oligonucleotides may be removed by washing the solid substrate, e.g. array, under standard hybridization conditions. The resultant substrate, e.g. array, comprises partially single stranded probes, wherein both the 5' and 3' ends of the surface probe are double stranded and the complementary positional domain (and optionally also the complementary amplification domain) is single stranded. The solid substrate, e.g. array, may be treated with a polymerase enzyme to extend the 3' end of the cleavage domain oligonucleotide, in a template dependent manner, so as to synthesize the positional domain (and optionally the amplification domain) of the capture probe. The 3' end of the synthesized positional domain is then ligated, e.g. using a ligase enzyme, to the 5' end of the capture domain oligonucleotide to generate the capture probe. It will be understood in this regard that the 5' end of the capture domain oligonucleotide is phosphorylated to enable ligation to take place. As each species of surface probe comprises a unique complementary positional domain, each species of capture probe will comprise a unique positional domain.

The term "hybridisation" or "hybridises" as used herein refers to the formation of a duplex between nucleotide sequences which are sufficiently complementary to form duplexes via Watson-Crick base pairing. Two nucleotide sequences are "complementary" to one another when those molecules share base pair organization homology. "Complementary" nucleotide sequences will combine

with specificity to form a stable duplex under appropriate hybridization conditions. For instance, two sequences are complementary when a section of a first sequence can bind to a section of a second sequence in an anti-parallel sense wherein the 3'-end of each sequence binds to the 5'-end of the other sequence and each A, T(U), G and C of one sequence is then aligned with a T(U), A, C and G, respectively, of the other sequence. RNA sequences can also include complementary G=U or U=G base pairs. Thus, two sequences need not have perfect homology to be "complementary" under the invention. Usually two sequences are sufficiently complementary when at least about 90% (preferably at least about 95%) of the nucleotides share base pair organization over a defined length of the molecule. The domains of the capture and surface probes thus contain a region of complementarity. Furthermore the capture domain of the capture probe contains a region of complementarity for the nucleic acid, e.g. RNA (preferably mRNA), of the biological specimen, e.g. tissue sample.

In a further representative embodiment, the capture probe may be synthesised on the solid substrate, e.g. array, using polymerase extension (similarly to as described above) and a terminal transferase enzyme to add a "tail" which may constitute the capture domain. The use of terminal transferases to add nucleotide sequences to the end of an oligonucleotide is known in the art, e.g. to introduce a homopolymeric tail e.g. a poly-T tail. Accordingly, in such a synthesis an oligonucleotide that corresponds to the cleavage domain of the capture probe (a cleavage domain oligonucleotide) may be contacted with the array and allowed to hybridize to the complementary domain of the surface probes. Excess oligonucleotides may be removed by washing the array under standard hybridization conditions. The resultant array comprises partially single stranded probes, wherein the 3' ends of the surface probes are double stranded and the complementary positional domain is single stranded. The array may be treated with a polymerase enzyme to extend the 3' end of the cleavage domain oligonucleotide, in a template dependent manner, so as to synthesize the positional domain (and optionally the amplification domain) of the capture probe. The capture domain, e.g. comprising a poly-T sequence may then be introduced using a terminal transferase to add a poly-T tail to generate the capture probe.

In yet another representative example, the capture probes may be synthesised on the solid substrate, e.g. array, using amplification techniques, such as bridge amplification. For example, a first plurality of nucleic acid primers is

attached to the solid support, wherein the nucleic acid primers include a first universal primer sequence that is common to said first nucleic acid primers and a second plurality of nucleic acid primers is attached to the solid support, wherein the nucleic acid primers comprise a second universal primer sequence that is common to said second nucleic acid primers. The solid support comprising said first and second nucleic acid primers is contacted with a plurality of different probes comprising sequences that are complementary to capture domain, positional domain, cleavage domain and amplification domain (if present), a first universal primer binding sequence that hybridizes to the first universal primers on the solid support and a sequence identical to the second universal primer sequence. The probes hybridise to the first universal primers, which are extended using the nucleic acid probes as templates thereby generating a first plurality of extended nucleic acid molecules.

The first plurality of extended nucleic acid molecules comprise second universal primer binding sequences that hybridize to the second universal primers on the solid support, which are extended using the first plurality of extended nucleic acid molecules as templates thereby generating a second plurality of extended nucleic acid molecules that are identical to the probes contacted with the solid substrate. The second plurality of extended nucleic acid molecules comprise first universal primer binding sequences that hybridize to the first universal primers on the solid support and thus are able to template the extension of said first universal primers. The continued extension of the primers results in the formation of clusters on the surface of the solid substrate comprising nucleic acid molecules with sequences that are identical or complementary to the plurality of different probes contacted with the solid support.

The nucleic acid probes contacted with the solid substrate comprise a cleavage site or domain between the sequence that is identical to the second universal primer sequence and the sequence that is complementary to the capture domain. Cleavage of this cleavage site or domain results in the release of the domain comprising the second universal primer sequence. It will be evident that, in some embodiments, release of the domain comprising the second universal primer sequence may reveal the capture domain of the capture probes, and release the second plurality of extended nucleic acid molecules from the surface of the array. In some embodiments, release of the domain comprising the second universal primer sequence may reveal the positional domain or amplification domain of the capture

probes, and release the second plurality of extended nucleic acid molecules from the surface of the array.

Thus, method described above results in random clusters of capture probes attached to the solid support which can be used for capturing nucleic acids from a biological specimen using method set forth previously herein. Advantageously, the released second plurality of extended nucleic acid molecules may be removed, e.g. by washing the solid substrate, prior to performing the method of the invention.

In this respect, it will be evident that the method described above may be modified to use a plurality of probes that comprise the capture domain, positional domain and cleavage domain, rather than their complements, e.g. by including the cleavage site or domain between the first universal primer binding sequence and the capture domain. Indeed the order of the domains of the probes contacted with the solid substrate will depend on the whether the capture probe is immobilized via its 3' end or 5' end and the skilled person readily could modify the order of the domain of the probes in order to generate an solid substrate, e.g. array, comprising capture probes suitable for use in the methods of the invention. Notably, the method of synthesizing the capture probes on the arrays using bridge amplification may be particularly useful in generating solid substrates in the form of flow cells, as described above.

As used herein, the term "random" can be used to refer to the spatial arrangement of the capture probes on the solid substrate, e.g. array. In this respect, there are at least two parameters relating to the spatial arrangement of the capture probes (i.e. features) on the solid substrate, e.g. array, described herein. The first parameter relates to the spacing and relative location of features and the second relates to identity or predetermined knowledge of the particular species of molecule that is present at a particular feature. Accordingly, in some embodiments, the features of an array can be randomly spaced such that adjacent features have variable spacing between each other. Alternatively, in some embodiments, the spacing between features can be ordered, for example, forming a regular pattern such as a rectilinear grid or hexagonal grid. In another respect, the features of an array can be random with respect to the identity or predetermined knowledge of the species of capture probe that occupies each feature independent of whether spacing produces a random pattern or ordered pattern. For example, in some embodiments set forth herein a solid substrate may be contacted with a population of nucleic acids under conditions where the nucleic acids attach at sites that are

ordered with respect to their relative locations but 'randomly located' with respect to knowledge of the sequence for the nucleic acid species present at any particular site. Reference to "randomly distributing" nucleic acids at locations on a surface is intended to refer to the absence of knowledge or absence of predetermination regarding which nucleic acid will be captured at which location (regardless of whether the locations are arranged in an ordered pattern or not).

It will be evident that random arrays may be produced using any convenient means. However, in some embodiments, the random arrays are produced by amplifying capture probes on the surface of the solid substrate using primers that are attached to the surface of the solid substrate such that the location of the capture probes is not predetermined. In some embodiments, random arrays are produced using so-called bead arrays, wherein different capture probes are attached to different beads in the array, which are randomly distributed on the surface of the array such that each bead, and thus each species of capture probe, is represented only once on the solid substrate. In this respect, it will be evident that the beads may be arranged in an ordered spatial arrangement, e.g. wherein the solid substrate comprises wells in an ordered spatial arrangement having dimensions that accommodate no more than a single bead. However, such bead arrays may be viewed as random arrays insofar as there is an absence of knowledge or absence of predetermination regarding where each capture probe will be located on the solid substrate.

As the methods of the invention rely on the ability to correlate the sequence of an extended probe comprising a positional domain to a particular location on the solid substrate, it is evident that random arrays must be analysed prior to their use in the methods of the invention to determine where each species of capture probe (i.e. where each positional domain) is located on the solid substrate. This may be achieved by any suitable means known in the art, such as sequencing or using hybridisation probes.

Thus, in some embodiments, the method of the invention may comprise a step of performing a nucleic acid detection reaction on the solid substrate to determine the positional domain sequences of the randomly located probes on the solid substrate. For instance, the nucleic acid detection reaction may be a sequencing reaction to determine the positional domain sequences of the capture probes located randomly on the solid substrate. In some embodiments, the nucleic acid detection reaction may be a decoder probe hybridization reaction on the solid

substrate to determine the positional domain sequences of the capture probes located randomly on the solid substrate. These steps may be performed prior to step (b) of the method of the invention.

By way of example, combinatorial hybridization methods, such as those used for decoding of multiplex bead arrays (see e.g. US Pat. No. 8,460,865, which is incorporated herein by reference), can be used to decode a "random" array. Such methods utilize labelled nucleic acid decoder probes that are complementary to at least a portion of the positional domain sequence (e.g. barcode sequence). A hybridization reaction can be carried out using decoder probes having known labels such that the location where the labels end up on the solid substrate identifies the nucleic acid probes according to rules of nucleic acid complementarity. In some cases, pools of many different probes with distinguishable labels are used, thereby allowing a multiplex decoding operation. The number of different barcodes determined in a decoding operation can exceed the number of labels used for the decoding operation. For example, decoding can be carried out in several stages where each stage constitutes hybridization with a different pool of decoder probes. The same decoder probes can be present in different pools but the label that is present on each decoder probe can differ from pool to pool (i.e. each decoder probe is in a different "state" when in different pools). Various combinations of these states and stages can be used to expand the number of positional domains (e.g. barcodes) that can be decoded well beyond the number of distinct labels available for decoding.

Sequencing techniques, such as sequencing-by-synthesis (SBS) techniques, are a particularly useful method for determining positional domain (e.g. barcode) sequences on random arrays. By way of example, SBS can be carried out as follows. To initiate a first SBS cycle, one or more labelled nucleotides, DNA polymerase, SBS primers etc., can be contacted with one or more features on the solid substrate (e.g. feature(s) where capture probes are attached to the solid support). Those features where SBS primer extension causes a labelled nucleotide to be incorporated can be detected. Optionally, the nucleotides can include a reversible termination moiety that terminates further primer extension once a nucleotide has been added to the SBS primer. For example, a nucleotide analogue having a reversible terminator moiety can be added to a primer such that subsequent extension cannot occur until a de-blocking agent is delivered to remove the moiety. Thus, for embodiments that use reversible termination, a de-blocking

reagent can be delivered to the solid substrate (before or after detection occurs). Washes can be carried out between the various delivery steps. The cycle can then be repeated n times to extend the primer by n nucleotides, thereby detecting a sequence of length n. Exemplary SBS procedures, fluidic systems and detection platforms that can be readily adapted for use with the method of the present invention are described, for example, in WO 91/06678, WO 04/018497, WO 07/123744; US Pat. Nos. 7,057,026, 7,329,492, 7,211,414, 7,315,019 or 7,405,281; and US Pat. App. Publ. No. 2008/0108082, each of which is incorporated herein by reference.

Thus, in some embodiments, the capture probes may also comprise a domain comprising a sequencing primer binding site. In particular, the sequencing primer binding site may be located between the capture domain and positional domain of the capture probe. A domain comprising a sequencing primer binding site located between the capture domain and positional domain of the capture probe may be particularly useful for sequencing the capture probes on a random array as described above.

Thus, it will be evident from the discussion above that the capture probe may comprise one or more universal domains, i.e. domains that are common to each species of capture probe. In particular, the universal domain(s) may be primer binding sites useful in synthesising the capture probes on the solid substrate, e.g. array, and/or sequencing the capture probe on random arrays.

It will also be evident that the capture probe of the invention may comprise two different cleavage sites or domains. For instance, a first cleavage domain may be located upstream (i.e. at the 5' end) of the positional domain, e.g. when the capture probe is immobilized by its 5' end. However, in some embodiments, the first cleavage domain may be located downstream (i.e. at the 3' end) of the capture domain, e.g. when the capture probe is immobilized by its 3' end. A second cleavage domain may be located between a first universal primer binding site used in the synthesis of the capture probe on the solid substrate, e.g. array, and the capture domain or complement thereof. The cleavage sites or domains must be reactive to different cleavage reactions (e.g. contain different cleavage enzyme recognition sites) such that each one can be selectively cleaved without necessarily cleaving the other. For instance, the second cleavage site or domain can be cleaved prior to step (c) (i.e. prior to releasing and extended the capture probe), thereby separating or removing the first universal primer binding site from the other

domains of the capture probe that remain attached to a solid substrate. The first cleavage site or domain is cleaved after contacting the solid substrate with the biological specimen, i.e. simultaneously with the extension reaction.

A typical solid substrate, e.g. array, for use in the methods of the invention may contain multiple spots or "features". A feature may be defined as an area or distinct position on the array substrate at which a single species of capture probe is immobilized. Hence each feature will comprise a multiplicity of probe molecules, of the same species. It will be understood in this context that whilst it is encompassed that each capture probe of the same species may have the same sequence, this need not necessarily be the case. Each species of capture probe will have the same positional domain (i.e. each member of a species and hence each probe in a feature will be identically "tagged"), but the sequence of each member of the feature (species) may differ, because the sequence of a capture domain may differ. As described above, random or degenerate capture domains may be used or capture domains specific for a subset of genes may be used. Thus the capture probes within a feature may comprise different random or degenerate sequences or different gene specific sequences. The number and density of the features on the solid substrate, e.g. array, will determine the resolution of the array, i.e. the level of detail at which the transcriptome or genome of the tissue sample can be analysed. Hence, a higher density of features will typically increase the resolution of the array.

As discussed above, the size and number of the features on the array of the invention will depend on the nature of the tissue sample and required resolution. Thus, if it is desirable to determine a transcriptome or genome only for regions of cells within a biological specimen, e.g. tissue sample, (or the sample contains large cells) then the number and/or density of features on the array may be reduced (i.e. lower than the possible maximum number of features) and/or the size of the features may be increased (i.e. the area of each feature may be greater than the smallest possible feature), e.g. an array comprising few large features. As mentioned above, in some embodiments, adjacent features may comprise the same capture probes, i.e. capture probes containing the same positional domain, in order to effectively increase the size of a feature, i.e. to lower the resolution of the array. Alternatively, if it is desirable to determine a transcriptome or genome of individual cells within a sample, it may be necessary to use the maximum number of features possible, which would necessitate using the smallest possible feature size, e.g. an array comprising many small features.

Whilst single cell resolution may be a preferred and advantageous feature of the present invention, it is not essential to achieve this, and resolution at the cell group level is also of interest, for example to detect or distinguish a particular cell type or tissue region, e.g. normal vs. tumour cells.

In representative embodiments of the invention, an array may contain at least 2, 5, 10, 50, 100, 500, 750, 1000, 1500, 3000, 5000, 10000, 20000, 40000, 50000, 75000, 100000, 150000, 200000, 300000, 400000, 500000, 750000, 800000, 1000000, 1200000, 1500000, 1750000, 2000000, 2100000, 3000000, 3500000, 4000000, 4500000, 5000000, 6000000, 7000000, 8000000, 9000000, 10000000, 12000000 or 15000000 features. As noted above, feature size and/or average distance between features (i.e. the pitch of the features) may be decreased and this may allow greater numbers of features to be accommodated within the same or a similar area. By way of example, these features may be comprised in an area of less than about $20cm^2$, $10cm^2$, $5cm^2$, $1cm^2$, $1mm^2$, or $100\mu m^2$.

It is well-established in the art that a solid substrate may comprise more than one “array”, thereby enabling multiplex analysis of biological specimens on a single substrate. Thus, in some embodiments, the solid substrate comprises 2 or more arrays, e.g. 3, 4, 5, 6, 7, 8, 9, 10 or more arrays, e.g. 12, 15, 20, 25 or more arrays.

Thus, in some embodiments of the invention the area of each feature may be from about 0.1 μm^2, 0.5 μm^2, 1μm^2, 2 μm^2, 3 μm^2, 4 μm^2, 5 μm^2, 10 μm^2, 12 μm^2, 15 μm^2, 20 μm^2, 50 μm^2, 75 μm^2, 100 μm^2, 150 μm^2, 200 μm^2, 250 μm^2, 300 μm^2, 400 μm^2, or 500 μm^2.

As used herein, the term “pitch,” when used in reference to features of an array, is intended to refer to the centre-to-centre spacing for adjacent features. A pattern of features can be characterized in terms of average pitch. The pattern can be ordered such that the coefficient of variation around the average pitch is small or the pattern can be random in which case the coefficient of variation can be relatively large. In either case, the average pitch can be, for example, at least about 10 nm, 0.1 µm, 0.5 µm, 1 µm, 5 µm, 10 µm, 100 µm or more. Alternatively or additionally, the average pitch can be, for example, at most about 100 µm, 10 µm, 5 µm, 1 µm, 0.5 µm 0.1 µm or less. Of course, the average pitch for a particular pattern of features can be between one of the lower values and one of the upper values selected from the ranges above.

The term “biological specimen” is intended to mean one or more cell, tissue, organism or portion thereof. It will be evident that a biological specimen from any organism could be used in the methods of the invention, e.g. plant, animal or fungal. The method of the invention allows the capture of any nucleic acid, e.g. mRNA, molecules, which are present in cells that are capable of transcription and/or translation. The method of the invention is particularly suitable for isolating and analysing the transcriptome or genome of cells within a biological specimen, e.g. a tissue sample, wherein spatial resolution of the transcriptomes or genomes is desirable, e.g. where the cells are interconnected or in contact directly with adjacent cells. However, it will be apparent to a person of skill in the art that the methods of the invention may also be useful for the analysis of the transcriptome or genome of different cells or cell types within a sample even if said cells do not interact directly, e.g. a blood sample. In other words, the cells do not need to present in the context of a tissue and can be applied to the array as single cells (e.g. cells isolated from a non-fixed tissue, e.g. a blood sample). Such single cells, whilst not necessarily fixed to a certain position in a tissue, are nonetheless applied to a certain position on the array and can be individually identified. Thus, in the context of analysing cells that do not interact directly, or are not present in a tissue context, the spatial properties of the described methods may be applied to obtaining or retrieving unique or independent transcriptome or genome information from individual cells.

The biological specimen may thus be a harvested or biopsied tissue sample, or possibly a cultured sample. Representative samples include clinical samples e.g. whole blood or blood-derived products, blood cells, tissues, biopsies, or cultured tissues or cells etc. including cell suspensions. Artificial tissues may for example be prepared from cell suspension (including for example blood cells). Cells may be captured in a matrix (for example a gel matrix e.g. agar, agarose, etc) and may then be sectioned in a conventional way. Such procedures are known in the art in the context of immunohistochemistry (see e.g. Andersson *et al* 2006, J. Histochem. Cytochem. 54(12): 1413-23. Epub 2006 Sep 6).

The mode of preparation of the biological specimen, e.g. tissue sample, and how the resulting sample is handled may affect the transcriptomic or genomic analysis of the methods of the invention. Moreover, various biological specimens, e.g. tissue samples, will have different physical characteristics and it is well within the skill of a person in the art to perform the necessary manipulations to yield a biological specimen, e.g. tissue sample, for use with the methods of the invention.

However, it is evident from the disclosures herein that any method of sample preparation may be used to obtain a biological specimen, e.g. tissue sample, that is suitable for use in the methods of the invention. For instance any layer of cells with a thickness of approximately 1 cell or less may be used in the methods of the invention. In one embodiment, the thickness of the biological specimen, e.g. tissue sample, may be less than 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 or 0.1 of the cross-section of a cell. However, since as noted above, the present invention is not limited to single cell resolution and hence it is not a requirement that the biological specimen, e.g. tissue sample, has a thickness of one cell diameter or less; thicker biological specimens, e.g. tissue samples, may if desired be used. For example cryostat sections may be used, which may be e.g. 10-20 µm thick.

The biological specimen, e.g. tissue sample, may be prepared in any convenient or desired way and the invention is not restricted to any particular type of tissue preparation. Fresh, frozen, fixed or unfixed tissues may be used. Any desired convenient procedure may be used for fixing or embedding the biological specimen, e.g. tissue sample, as described and known in the art. Thus any known fixatives or embedding materials may be used.

As a first representative example of a biological specimen, e.g. tissue sample for use in the invention, a tissue may prepared by deep freezing at temperature suitable to maintain or preserve the integrity (i.e. the physical characteristics) of the tissue structure, e.g. less than -20°C and preferably less than -25, -30, -40, -50, -60, -70 or -80 °C. The frozen tissue sample may be sectioned, i.e. thinly sliced, onto the solid substrate, e.g. array, surface by any suitable means. For example, a tissue sample may be prepared using a chilled microtome, a cryostat, set at a temperature suitable to maintain both the structural integrity of the tissue sample and the chemical properties of the nucleic acids in the sample, e.g. to less than -15°C and preferably less than -20 or -25°C. Thus, the sample should be treated so as to minimize the degeneration or degradation of the nucleic acid, e.g. RNA, in the tissue. Such conditions are well-established in the art and the extent of any degradation may be monitored through nucleic acid extraction, e.g. total RNA extraction, and subsequent quality analysis at various stages of the preparation of the tissue sample.

In a second representative example, a tissue may be prepared using standard methods of formalin-fixation and paraffin-embedding (FFPE), which are well-established in the art. Following fixation of a tissue sample and embedding in a

paraffin or resin block, the tissue sample may sectioned, i.e. thinly sliced, onto the solid substrate, e.g. array. As noted above, other fixatives and/or embedding materials can be used.

It will be apparent that the biological specimen, e.g. tissue sample section, will need to be treated to remove the embedding material e.g. to deparaffinize, i.e. to remove the paraffin or resin, from the sample prior to carrying out the methods of the invention. This may be achieved by any suitable method and the removal of paraffin or resin or other material from tissue samples is well established in the art, e.g. by incubating the sample (on the surface of the solid substrate, e.g. array) in an appropriate solvent e.g. xylene, e.g. twice for 10 minutes, followed by an ethanol rinse, e.g. 99.5% ethanol for 2 minutes, 96% ethanol for 2 minutes, and 70% ethanol for 2 minutes.

It will be evident to the skilled person that the RNA in tissue sections prepared using methods of FFPE or other methods of fixing and embedding is more likely to be partially degraded than in the case of frozen tissue. However, without wishing to be bound by any particular theory, it is believed that this may be advantageous in the methods of the invention. For instance, if the RNA in the sample is partially degraded the average length of the RNA polynucleotides will be less and more randomized than a non-degraded sample. It is postulated therefore that partially degraded RNA would result in less bias in the various processing steps, described elsewhere herein, e.g. ligation of adaptors (amplification domains), amplification of the cDNA molecules and sequencing thereof.

Hence, in one embodiment of the invention the biological specimen, e.g. tissue sample, i.e. the section of the tissue sample contacted with the solid substrate, e.g. array, is prepared using FFPE or other methods of fixing and embedding. In other words the sample may be fixed, e.g. fixed and embedded. In an alternative embodiment of the invention the tissue sample is prepared by deep-freezing. In another embodiment a touch imprint of a tissue may be used, according to procedures known in the art. In other embodiments an unfixed sample may be used.

Whilst the biological specimen may be prepared for use in the methods of the invention using any suitable techniques or preparation methods, in some particularly preferred embodiments, the nucleic acid molecules of the specimen (particularly RNA molecules) are not modified prior to contact with the solid substrate (e.g. array). For instance, in some preferred embodiments the nucleic

acid molecules (e.g. RNA) of the biological specimen are not treated to enhance or facilitate their interaction with the capture probes prior to contact with the solid substrate. In particular, in some embodiments the nucleic acid molecules (e.g. RNA) of the biological specimen are not contacted with or hybridized to nucleic acids (e.g. aptamers, oligonucleotides or nucleic acid tags) that facilitate the interaction between the nucleic acid molecules of the biological sample and capture probes) prior to contacting the biological specimen with the solid substrate. For instance, in some embodiments, the biological specimen is not contacted with or hybridized to nucleic acids that function as an intermediary between the target nucleic acid of the biological specimen and the capture probes, i.e. that function to indirectly attach or link the target nucleic acid of the biological specimen to the capture probe, prior to contacting the biological specimen with the substrate (e.g. array).

The thickness of the biological specimen, e.g. tissue sample section, for use in the methods of the invention may be dependent on the method used to prepare the sample and the physical characteristics of the tissue. Thus, any suitable section thickness may be used in the methods of the invention. In representative embodiments of the invention the thickness of the tissue sample section will be at least 0.1μm, further preferably at least 0.2, 0.3, 0.4, 0.5, 0.7, 1.0, 1.5, 2, 3, 4, 5, 6, 7, 8, 9 or 10μm. In other embodiments the thickness of the tissue sample section is at least 10, 12, 13, 14, 15, 20, 30, 40 or 50μm. However, the thickness is not critical and these are representative values only. Thicker samples may be used if desired or convenient e.g. 70 or 100 μm or more. Typically, the thickness of the tissue sample section is between 1-100 μm, 1-50 μm , 1-30 μm, 1-25 μm, 1-20 μm, 1-15 μm, 1-10 μm, 2-8μm, 3-7μm or 4-6μm, but as mentioned above thicker samples may be used.

On contact of the biological specimen, e.g. tissue sample section, with the solid substrate, e.g. following removal of the embedding material, e.g. deparrafinization, the nucleic acid, e.g. RNA, molecules in the biological specimen, e.g. tissue sample, will bind to the immobilized capture probes on the array. In some embodiments it may be advantageous to facilitate the hybridization of the nucleic acid, e.g. RNA, molecules to the capture probes. Typically, facilitating the hybridization comprises modifying the conditions under which hybridization occurs. The primary conditions that can be modified are the time and temperature of the

incubation of the biological specimen, e.g. tissue sample section, on the array prior to the reverse transcription step, which is described elsewhere herein.

It will be evident that biological specimens, e.g. tissue samples, from different sources may require different treatments to allow the nucleic acids to interact with, i.e. hybridize to, the capture probes. For instance, it may be useful to permeabilize the biological specimen, e.g. tissue sample, to facilitate the interaction between the nucleic acid and the capture probes. If the tissue sample is not permeabilized sufficiently the amount of nucleic acid captured by the capture probes may be too low to enable further analysis. Conversely, if the biological specimen, e.g. tissue sample, is too permeable, the nucleic acid may diffuse away from its origin in the biological specimen, e.g. tissue sample, i.e. the nucleic acid may be captured by the capture probes may not correlate accurately with its original spatial distribution in the biological specimen, e.g. tissue sample. Hence, there must be a balance between permeabilizing the biological specimen, e.g. tissue sample, enough to obtain enable efficient interaction between the nucleic acids and the capture probes whilst maintaining the spatial resolution of the nucleic acid distribution in the biological specimen, e.g. tissue sample. The methods used to fix the biological specimen, e.g. tissue sample, may also impact on the interaction between the nucleic acid of the biological specimen, e.g. tissue sample, and the capture probes.

Thus, in some embodiments, the method further comprises a step of permeabilizing the biological specimen. This step may be performed after the step of contacting the specimen with the solid substrate. In some embodiments, the step of permeabilizing the biological specimen may take place before or contemporaneously with the step of releasing and extending the capture probes.

Suitable methods and agents for permeabilizing and/or fixing biological specimens, e.g. cells and tissues, are well known in the art and any appropriate method may be selected for use in the methods of the invention. Some proteases are particularly useful in permeabilizing cells in a biological specimen, e.g. pepsin. Particularly useful fixatives include, e.g. methanol.

As conditions for spatial tagging, e.g. localised or spatial detection, of nucleic acid molecules from a biological specimen, e.g. tissue sample, on a solid substrate vary depending on the biological specimen, e.g. tissue sample, a typical range of parameters is discussed herein. For instance, on contacting a biological specimen, e.g. tissue sample section, with the solid substrate, e.g. array, the

substrate may be incubated for at least 1 hour to allow the nucleic acid, e.g. RNA, to hybridize to the capture probes. Preferably the solid substrate, e.g. array, may be incubated for at least 2, 3, 5, 10, 12, 15, 20, 22 or 24 hours or until the tissue sample section has dried. The substrate incubation time is not critical and any convenient or desired time may be used. Typical substrate, e.g. array, incubations may be up to 72 hours. Thus, the incubation may occur at any suitable temperature, for instance at room temperature, although in a preferred embodiment the biological specimen, e.g. tissue sample section, is incubated on the substrate, e.g. array, at a temperature of at least 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36 or 37°C. Incubation temperatures of up to 55°C are commonplace in the art. In a particularly preferred embodiment the tissue sample section is allowed to dry on the array at 37°C for 24 hours. Once the biological specimen, e.g. tissue sample section, has dried the substrate, e.g. array, may be stored at room temperature before performing the combined cleavage and extension, e.g. reverse transcription, step. It will be understood that the if the biological specimen, e.g. tissue sample section, is allowed to dry on the surface of the substrate, e.g. array, it will need to be rehydrated before further manipulation of the captured nucleic acid can be achieved, e.g. the step of cleaving the capture probes from the solid substrate and simultaneously or subsequently reverse transcribing the captured RNA.

Hence, the method of the invention may comprise a further step of rehydrating the biological specimen, e.g. tissue sample, after contacting the specimen with the solid substrate, e.g. array.

In some embodiments it may be advantageous to block (e.g. mask or modify) the capture probes prior to contacting the biological specimen, e.g. tissue sample, with the substrate, e.g. array, particularly when the nucleic acid in the biological specimen, e.g. tissue sample, is subject to a process of modification prior to its capture on the array. Specifically, it may be advantageous to block or modify the free 3' end of the capture probe. In a particular embodiment, the nucleic acid in the biological specimen, e.g. tissue sample, e.g. fragmented genomic DNA, may be modified such that it can be captured by the capture probe. For instance, and as described in more detail below, an adaptor sequence (comprising a binding domain capable of binding to the capture domain of the capture probe) may be added to the end of the nucleic acid, e.g. fragmented genomic DNA. This may be achieved by, e.g. ligation of an adaptor or extension of the nucleic acid, e.g. using an enzyme to incorporate additional nucleotides at the end of the sequence, e.g. a poly-A tail. It is

necessary to block or modify the capture probes, particularly the free 3' end of the capture probe, prior to contacting the biological specimen, e.g. tissue sample, with the array to avoid modification of the capture probes, e.g. to avoid the addition of a poly-A tail to the free 3' end of the capture probes. Preferably the incorporation of a blocking domain may be incorporated into the capture probe when it is synthesised. However, the blocking domain may be incorporated to the capture probe after its synthesis.

In some embodiments the capture probes may be blocked by any suitable and reversible means that would prevent modification of the capture domains during the process of modifying the nucleic acid of the biological specimen, e.g. tissue sample, which occurs after the biological specimen, e.g. tissue sample, has been contacted with the array. In other words, the capture probes may be reversibly masked or modified such that the capture domain of the capture probe does not comprise a free 3' end, i.e. such that the 3' end is removed or modified, or made inaccessible so that the capture probe is not susceptible to the process which is used to modify the nucleic acid of the biological specimen, e.g. ligation or extension, or the additional nucleotides may be removed to reveal and/or restore the 3' end of the capture domain of the capture probe.

For example, blocking probes may be hybridised to the capture probes to mask the free 3' end of the capture domain, e.g. hairpin probes or partially double stranded probes, suitable examples of which are known in the art. The free 3' end of the capture domain may be blocked by chemical modification, e.g. addition of an azidomethyl group as a chemically reversible capping moiety, such that the capture probes do not comprise a free 3' end. Suitable alternative capping moieties are well known in the art, e.g. the terminal nucleotide of the capture domain could be a reversible terminator nucleotide, which could be included in the capture probe during or after probe synthesis.

Alternatively or additionally, the capture domain of the capture probe could be modified so as to allow the removal of any modifications of the capture probe, e.g. additional nucleotides, that occur when the nucleic acid molecules of the biological specimen, e.g. tissue sample, are modified. For instance, the capture probes may comprise an additional sequence downstream of the capture domain, i.e. 3' to capture domain, namely a blocking domain. This could be in the form of a cleavage domain as defined above, e.g. a restriction endonuclease recognition sequence or a sequence of nucleotides cleavable by specific enzyme activities, e.g.

uracil. Following the modification of the nucleic acid of the biological specimen, e.g. tissue sample, the capture probes could be subjected to an enzymatic cleavage, which would allow the removal of the blocking domain and any of the additional nucleotides that are added to the 3' end of the capture probe during the modification process. The removal of the blocking domain would reveal and/or restore the free 3' end of the capture domain of the capture probe. The blocking domain could be synthesised as part of the capture probe (e.g. as described above with respect to synthesising the probes on the solid substrate using bridge amplification) or could be added to the capture probe *in situ* (i.e. as a modification of an existing array), e.g. by ligation of the blocking domain.

The capture probes may be blocked using any combination of the blocking mechanisms described above. In this respect, it will be evident that the capture domain may be "blocked" by immobilizing the capture probes by their 3' end.

Once the nucleic acid of the biological specimen, e.g. tissue sample, e.g. fragmented genomic DNA, has been modified to enable it to hybridise to the capture domain of the capture probe, the capture probe must be unblocked, e.g. by dissociation of the blocking oligonucleotide, removal of the capping moiety and/or blocking domain.

In order to correlate the sequence analysis or transcriptome or genome information obtained from each feature of the array with the region (i.e. an area or cell) of the biological specimen, e.g. tissue sample, the biological specimen is oriented in relation to the features on the array. In other words, the tissue sample is placed on the array such that the position of a capture probe on the array may be correlated with a position in the biological specimen, e.g. tissue sample. Thus it may be identified where in the biological specimen, e.g. tissue sample, the position of each species of capture probe (or each feature of the array) corresponds. In other words, it may be identified to which location in the biological specimen, e.g. tissue sample, the position of each species of capture probe corresponds. For instance, this may be done by virtue of positional markers present on the solid substrate, e.g. array, as described below.

Conveniently, but not necessarily, the biological specimen, e.g. tissue sample, may be imaged following its contact with the array. This may be performed before or after the nucleic acid of the biological specimen, e.g. tissue sample, is processed, e.g. before or after the combined release and extension step. In some embodiments the biological specimen, e.g. tissue sample, is imaged prior to the

combined release and extension of the capture probes from the solid substrate, e.g. array. In a particularly preferred embodiment the biological specimen, e.g. tissue sample, is imaged after the combined release and extension of the capture probes from the solid substrate, e.g. array. It is not necessary to remove any residual tissue from the solid substrate, e.g. array, prior to the pooling, i.e. collecting, the released molecules (extended probes), from the substrate as the extended probes are capable of diffusing through the residue tissue. Furthermore, the step of processing the captured nucleic acid, e.g. combined release and extension of the capture probes, may act to remove some tissue from the array surface, e.g. when using tissue preparing by deep-freezing. In such a case, it may be necessary to separate residual tissue collected when pooling the extended probes from the solid substrate, if performed, prior to the step of analysing the extended probes. This may be achieved by any suitable means, e.g. centrifugation.

Thus, in some embodiments, the method may comprise a step of separating tissue from the pooled extended probes.

When the step of processing the capture nucleic acid acts to remove some tissue from the array surface, imaging of the biological specimen, e.g. tissue sample, may take place prior to the processing step, e.g. the combined release and extension step. Generally speaking, imaging may take place at any time after contacting the biological specimen, e.g. tissue sample, with the solid substrate, but before any step which degrades or removes the biological specimen, e.g. tissue sample. As noted above, this may depend on the biological specimen, e.g. tissue sample.

Advantageously, the solid substrate, e.g. array, may comprise markers to facilitate the orientation of the biological specimen, e.g. tissue sample, or the image thereof in relation to the features of the solid substrate, e.g. array. Any suitable means for marking the solid substrate, e.g. array, may be used such that they are detectable when the biological specimen, e.g. tissue sample, is imaged. For instance, a molecule, e.g. a fluorescent molecule, that generates a signal, preferably a visible signal, may be immobilized directly or indirectly on the surface of the substrate, e.g. array. Preferably, the substrate, e.g. array, comprises at least two markers in distinct positions on the surface of the array, further preferably at least 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100 markers. Conveniently several hundred or even several thousand markers may be used. The markers may be provided in a pattern, for example make up an outer edge of the

array, e.g. an entire outer row of the features of an array. Other informative patterns may be used, e.g. lines sectioning the array. This may facilitate aligning an image of the biological specimen, e.g. tissue sample, to an array, or indeed generally in correlating the features of the array to the biological specimen, e.g. tissue sample. Thus, the marker may be an immobilized molecule to which a signal giving molecule may interact to generate a signal.

In a representative example, the solid substrate, e.g. array, may comprise a marker feature, e.g. a nucleic acid probe immobilized on the substrate, e.g. array, to which a labelled nucleic acid may hybridize. For instance, the labelled nucleic acid molecule, or marker nucleic acid, may be linked or coupled to a chemical moiety capable of fluorescing when subjected to light of a specific wavelength (or range of wavelengths), i.e. excited. Such a marker nucleic acid molecule may be contacted with the substrate, e.g. array, before, contemporaneously with or after the biological specimen, e.g. tissue sample, is stained in order to visualize or image the biological specimen, e.g. tissue sample. However, the marker must be detectable when the biological specimen, e.g. tissue sample is imaged. Thus, in a preferred embodiment the marker may be detected using the same imaging conditions used to visualize the biological specimen, e.g. tissue sample.

In a particularly preferred embodiment of the invention, the substrate, e.g. array, comprises marker features to which a labelled, preferably fluorescently labelled, marker nucleic acid molecule, e.g. oligonucleotide, is hybridized.

It will be evident however, that the use of positional markers is not essential for correlating the sequence analysis or transcriptome or genome information obtained from each feature of the array with the region (i.e. an area or cell) of the biological specimen, i.e. for aligning the biological specimen with the features of the array. In this respect, it may be possible to align the topology of the biological specimen to the data generated from analysing the extended probes. For instance, a crack in a tissue sample section would correspond to the absence of extended probes from particular features on the array and this may be sufficient to align the biological specimen with the features on the array. Furthermore, in some embodiments, the features on the surface of the solid substrate are visible, particularly following staining of the biological specimen, using high resolution imaging, thereby enabling the biological specimen to be aligned with the features on the array. Thus, in some embodiments, the feature of the array may be viewed as positional markers.

The step of imaging the biological specimen, e.g. tissue sample, may use any convenient histological means known in the art, e.g. light, bright field, dark field, phase contrast, fluorescence, reflection, interference, confocal microscopy or a combination thereof. Typically the biological specimen, e.g. tissue sample, is stained prior to visualization to provide contrast between the different regions, e.g. cells, of the biological specimen, e.g. tissue sample. The type of stain used will be dependent on the type of biological specimen, e.g. tissue, and the region of the cells to be stained. Such staining protocols are known in the art. In some embodiments more than one stain may be used to visualize (image) different aspects of the biological specimen, e.g. tissue sample, e.g. different regions of the tissue sample, specific cell structures (e.g. organelles) or different cell types. In other embodiments, the biological specimen, e.g. tissue sample, may be visualized or imaged without staining the sample, e.g. if the tissue sample contains already pigments that provide sufficient contrast or if particular forms of microscopy are used.

In a preferred embodiment, the tissue sample is visualized or imaged using fluorescence microscopy.

In some embodiments, extending the capture probe comprises generating cDNA from the captured (hybridized) RNA. It will be understood that this refers to the synthesis of a complementary strand of the hybridized nucleic acid, e.g. generating cDNA based on the captured RNA template (the RNA hybridized to the capture domain of the capture probe). Thus, in an initial step of extending the capture probe, e.g. the cDNA generation, the captured (hybridized) nucleic acid, e.g. RNA, acts as a template for the extension, e.g. reverse transcription, step.

Reverse transcription concerns the step of synthesizing cDNA (complementary or copy DNA) from RNA, preferably mRNA (messenger RNA), by reverse transcriptase. Thus cDNA can be considered to be a copy of the RNA present in a cell at the time at which the biological specimen, e.g. tissue sample, was taken, i.e. it represents all or some of the genes that were expressed in said cell at the time of isolation.

The capture probe, specifically the capture domain of the capture probe, acts as a primer for producing the complementary strand of the nucleic acid hybridized to the capture probe, e.g. a primer for reverse transcription. Hence, the nucleic acid, e.g. cDNA, molecules generated by the extension reaction, e.g. reverse transcription reaction, incorporate the sequence of the capture probe, i.e.

the extension reaction, e.g. reverse transcription reaction, may be seen as a way of labelling or tagging indirectly the nucleic acid, e.g. transcripts, of the biological specimen, e.g. tissue sample, that are in contact with each feature of the array or within the vicinity of the feature such that the released capture probes are able to bind to the transcripts in a localised manner. As mentioned above, each species of capture probe comprises a positional domain (feature identification tag) that represents a unique sequence for each feature (or group of adjacent features) of the array. Thus, all of the nucleic acid, e.g. cDNA, molecules synthesized at a specific feature (or in the vicinity of a specific feature) will comprise the same nucleic acid "tag".

The nucleic acid, e.g. cDNA, molecules synthesized at each feature (or in the vicinity of a specific feature) of the array may represent the genome of, or genes expressed from, the region or area of the biological specimen, e.g. tissue sample, in contact with that feature (or in the vicinity of that feature), e.g. a tissue or cell type or group or sub-group thereof, and may further represent genes expressed under specific conditions, e.g. at a particular time, in a specific environment, at a stage of development or in response to stimulus etc. Hence, the cDNA at any single feature (i.e. captured by a specific species of capture probe) may represent the genes expressed in a single cell, or if the feature is in contact with (or in the vicinity of) the sample at a cell junction, the cDNA may represent the genes expressed in more than one cell. Similarly, if a single cell is in contact with (or in the vicinity of) multiple features, then each feature (i.e. a specific species of extended capture probe) may represent a proportion of the genes expressed in said cell. Similarly, in embodiments in which the captured nucleic acid is DNA, any single feature (i.e. a specific species of extended capture probe) may be representative of the genome of a single cell or more than one cell. Alternatively, the genome of a single cell may be represented by multiple features (i.e. multiple species of extended capture probe).

The extension of the capture probe, e.g. reverse transcription reaction, may be performed using any suitable enzymes and protocol of which many exist in the art, as described herein. However, it will be evident that it is not necessary to provide a primer for the synthesis of the first nucleic acid, e.g. cDNA, strand because the capture domain of the capture probe acts as the primer, e.g. reverse transcription primer.

Preferably, in the context of the present invention the extended probes, e.g. cDNA, are treated to comprise double stranded DNA. Treatment of the extended probes to produce double stranded DNA may be achieved in a single reaction to generate only a second DNA, e.g. cDNA, strand, i.e. to produce double stranded DNA molecules without increasing the number of double stranded DNA molecules, or in an amplification reaction to generate multiple copies of the second strand, which may be in the form of single stranded DNA (e.g. linear amplification) or double stranded DNA, e.g. cDNA (e.g. exponential amplification).

The step of second strand DNA, e.g. cDNA, synthesis may take place on the solid substrate, e.g. array, (i.e. prior to pooling the extended molecules), either as a discrete step of second strand synthesis, for example using random primers as described in more detail below, or in the initial step of an amplification reaction. Alternatively, the first strand DNA (i.e. extended probes), e.g. cDNA (the strand comprising, i.e. incorporating, the capture probe) may be pooled (e.g. collected) and second strand synthesis, whether as a discrete step or in an amplification reaction may occur subsequently, e.g. in a reaction carried out in a separate vessel or container, e.g. PCR tube.

Where second strand synthesis takes place on the solid substrate, e.g. array, the method may include an optional step of removing the captured nucleic acid, e.g. RNA, before the second strand synthesis, for example using an RNA digesting enzyme (RNase), e.g. RNase H. Procedures for this are well known and described in the art. However, this is generally not necessary, and in most cases the RNA degrades naturally. RNase H can be used if desired to increase the robustness of RNA removal.

For instance, in biological specimens, e.g. tissue samples, that comprise large amounts of RNA, the step of generating the double stranded cDNA may yield a sufficient amount of cDNA that it may be sequenced directly (following the step of pooling the extended probes). In this case, second strand cDNA synthesis may be achieved by any means known in the art and as described below. The second strand synthesis reaction may be performed on the solid substrate, i.e. prior to pooling the extended probes (e.g. transferring the extended probes to a separate vessel or container, e.g. PCR tube), or after the extended probes, e.g. cDNA, have been pooled.

In other embodiments it may be necessary to enhance, i.e. amplify, the extended probes, e.g. synthesized cDNA, to yield quantities that are sufficient for

analysis, e.g. DNA sequencing. In this embodiment, the first strand of the extended probes, e.g. cDNA molecules, which comprise also the capture probe of the features of the array, acts as a template for the amplification reaction, e.g. a polymerase chain reaction. The first reaction product of the amplification will be a second strand of DNA, e.g. cDNA, which itself will act as a template for further cycles of the amplification reaction.

In either of the above described embodiments, the second strand of DNA, e.g. cDNA, will comprise a complement of the capture probe. If the capture probe comprises an amplification domain, then this may be used for the subsequent amplification and sequence analysis of the DNA, e.g. cDNA, e.g. the amplification reaction may comprise a primer with the same sequence as the amplification domain, i.e. a primer that is complementary (i.e. hybridizes) to the complement of the amplification domain. In view of the fact that the amplification domain is upstream of the positional domain of the capture probe (in the extended probe, e.g. the first cDNA strand), the complement of the positional domain will be incorporated in the second strand of the DNA, e.g. cDNA molecules.

In embodiments where the second strand of DNA, e.g. cDNA, is generated in a single reaction, the second strand synthesis may be achieved by any suitable means. For instance, the first strand cDNA (i.e. extended probe) may be incubated with random primers, e.g. hexamer primers, and a DNA polymerase, preferably a strand displacement polymerase, e.g. klenow (exo), under conditions sufficient for templated DNA synthesis to occur. This process will yield double stranded cDNA molecules of varying lengths and is unlikely to yield full-length cDNA molecules, i.e. cDNA molecules that correspond to entire mRNA from which they were synthesized. The random primers will hybridise to the first strand cDNA molecules (i.e. extended probes) at a random position, i.e. within the sequence rather than at the end of the sequence.

If it is desirable to generate full-length DNA, e.g. cDNA, molecules, i.e. molecules that correspond to the whole of the captured nucleic acid, e.g. RNA, molecule (if the nucleic acid, e.g. RNA, was partially degraded in the tissue sample then the captured nucleic acid, e.g. RNA, molecules will not be "full-length" transcripts or the same length as the initial fragments of genomic DNA), then the 3' end of the extended probes, e.g. first stand cDNA, molecules may be modified. For example, a linker or adaptor may be ligated to the 3' end of the cDNA molecules.

This may be achieved using single stranded ligation enzymes such as T4 RNA ligase or Circligase™ (Epicentre Biotechnologies).

Alternatively, a second strand synthesis helper probe (a partially double stranded DNA molecule capable of hybridising to the 3' end of the first strand cDNA molecule), may be ligated to the 3' end of the extended probe, e.g. first strand cDNA, molecule using a double stranded ligation enzyme such as T4 DNA ligase. Other enzymes appropriate for the ligation step are known in the art and include, e.g. Tth DNA ligase, Taq DNA ligase, *Thermococcus* sp. (strain 9°N) DNA ligase (9°N™ DNA ligase, New England Biolabs), and Ampligase™ (Epicentre Biotechnologies). The second strand synthesis helper probe comprises also a specific sequence from which the second strand DNA, e.g. cDNA, synthesis may be primed using a primer that is complementary to the part of the helper probe that is ligated to the extended probe, e.g. first cDNA strand. A further alternative comprises the use of a terminal transferase active enzyme to incorporate a polynucleotide tail, e.g. a poly-A tail, at the 3' end of the extended probe, e.g. first strand of cDNA, molecules. The second strand synthesis may be primed using a poly-T primer, which may also comprise a specific amplification domain for further amplification. Other methods for generating "full-length" double stranded DNA, e.g. cDNA, molecules (or maximal length second strand synthesis) are well-established in the art.

In some embodiments, second strand synthesis may use a method of template switching, e.g. using the SMART™ technology from Clontech®. SMART (Switching Mechanism at 5' End of RNA Template) technology is well established in the art and is based that the discovery that reverse transcriptase enzymes, e.g. Superscript® II (Invitrogen), are capable of adding a few nucleotides at the 3' end of an extended cDNA molecule, i.e. to produce a DNA/RNA hybrid with a single stranded DNA overhang at the 3' end. The DNA overhang may provide a target sequence to which an oligonucleotide probe can hybridise to provide an additional template for further extension of the cDNA molecule. Advantageously, the oligonucleotide probe that hybridises to the cDNA overhang contains an amplification domain sequence, the complement of which is incorporated into the synthesised first strand cDNA product. Primers containing the amplification domain sequence, which will hybridise to the complementary amplification domain sequence incorporated into the cDNA first strand, can be added to the reaction mix to prime second strand synthesis using a suitable polymerase enzyme and the

cDNA first strand as a template. This method avoids the need to ligate adaptors to the 3' end of the cDNA first strand. Whilst template switching was originally developed for full-length mRNAs, which have a 5' cap structure, it has since been demonstrated to work equally well with truncated mRNAs without the cap structure. Thus, template switching may be used in the methods of the invention to generate full length and/or partial or truncated cDNA molecules. Thus, in a preferred embodiment of the invention, the second strand synthesis may utilise, or be achieved by, template switching. In a particularly preferred embodiment, the template switching reaction, i.e. the further extension of the cDNA first strand to incorporate the complementary amplification domain, is performed on the solid substrate (i.e. prior to pooling the extended probes). Preferably, the second strand synthesis reaction is also performed in solution on the solid substrate.

In some embodiments it may be advantageous to treat the double stranded extended probes to remove any unextended capture probes prior to amplification and/or analysis, e.g. sequence analysis. This may be achieved by any suitable means known in the art, e.g. using an enzyme to degrade the unextended probes, e.g. an exonuclease enzyme, or purification columns. Thus, in some embodiments, the method further comprises a step of treating the extended probes to removed unextended capture probes. This may be, for example, after the step of pooling the extended probes.

In embodiments where it may be necessary or advantageous to enhance, enrich or amplify the DNA, e.g. cDNA molecules, amplification domains may be incorporated in the DNA, e.g. cDNA molecules. As discussed above, a first amplification domain may be incorporated into the extended probes, e.g. the first strand of the cDNA molecules, when the capture probe comprises an amplification domain. In these embodiments, the second strand synthesis may incorporate a second amplification domain. For example, the primers used to generate the second strand cDNA, e.g. random hexamer primers, poly-T primer, the primer that is complementary to the second strand synthesis helper probe, may comprise at their 5' end an amplification domain, i.e. a nucleotide sequence to which an amplification primer may hybridize. Thus, the resultant double stranded DNA may comprise an amplification domain at or towards each 5' end of the double stranded DNA, e.g. cDNA molecules. These amplification domains may be used as targets for primers used in an amplification reaction, e.g. PCR. Alternatively, the linker or adaptor which is ligated to the 3' end of the extended probes, e.g. first strand cDNA

molecules, may comprise a second amplification domain. Similarly, a second amplification domain may be incorporated into the first strand cDNA molecules by template switching.

In embodiments where the capture probe does not comprise an amplification domain, the second strand of the cDNA molecules may be synthesised in accordance with the above description. The resultant double stranded DNA molecules may be modified to incorporate an amplification domain at the 5' end of the first DNA, e.g. cDNA strand (a first amplification domain) and, if not incorporated in the second strand DNA, e.g. cDNA synthesis step, at the 5' end of the second DNA, e.g. cDNA strand (a second amplification domain). Such amplification domains may be incorporated, e.g. by ligating double stranded adaptors to the ends of the DNA, e.g. cDNA molecules. Enzymes appropriate for the ligation step are known in the art and include, e.g. Tth DNA ligase, Taq DNA ligase, *Thermococcus* sp. (strain 9°N) DNA ligase (9°N™ DNA ligase, New England Biolabs), Ampligase™ (Epicentre Biotechnologies) and T4 DNA ligase. In a preferred embodiment the first and second amplification domains comprise different sequences.

From the above, it is therefore apparent that amplification domains may be added to the extended probes, i.e. DNA molecules, for example to the cDNA molecules, or their complements (e.g. second strand) by various methods and techniques and combinations of such techniques known in the art, e.g. by use of primers which include such a domain, ligation of adaptors, use of terminal transferase enzymes and/or by template switching methods. As is clear from the discussion herein, such domains may be added before or after the extended probes are pooled, i.e. transferred to a separate vessel or container, e.g. PCR tube.

It will be apparent from the above description that all of the DNA, e.g. cDNA molecules from a single solid substrate, e.g. array, that have been synthesized by the methods of the invention may all comprise the same first and second amplification domains. Consequently, a single amplification reaction, e.g. PCR, may be sufficient to amplify all of the DNA, e.g. cDNA molecules. Thus in a preferred embodiment, the method of the invention may comprise a step of amplifying the DNA, e.g. cDNA molecules. In one embodiment the amplification step is performed after the extended probes, e.g. cDNA molecules, have been pooled. In other embodiments amplification may be performed before the extended probes are pooled, i.e. on the solid substrate, e.g. array. It is known in the art that amplification

reactions may be carried out on solid substrates, such as arrays, by forming wells on the surface of the substrate to contain the solution or by immersing the substrate in a solution. Indeed, on-chip thermocyclers exist for carrying out such reactions. Thus, in one embodiment arrays which are known in the art as sequencing platforms or for use in any form of sequence analysis (e.g. in or by next generation sequencing technologies) may be used as the basis of the arrays of the present invention (e.g. Illumina bead arrays, flow cells etc. as discussed above)

For the synthesis of the second strand of DNA, e.g. cDNA, it is preferable to use a strand displacement polymerase (e.g. Φ29 DNA polymerase, Bst (exo^-) DNA polymerase, klenow (exo^-) DNA polymerase) if the extended probes comprise a partially double stranded nucleic acid molecule. For instance, the extended probes may be at least partially double stranded (e.g. DNA:DNA, DNA:RNA or DNA:DNA/RNA hybrid) in embodiments where the capture probe is immobilized indirectly on the substrate of the array via a surface probe. The strand displacement polymerase is necessary to ensure that the second cDNA strand synthesis incorporates the complement of the positional domain (feature identification domain) into the second DNA, e.g. cDNA strand.

As mentioned above, the method of the invention may be seen to comprise a step of pooling, e.g. collecting or recovering, the extended probes, e.g. cDNA molecules (including double stranded cDNA molecules), and/or their amplicons. As noted above, in the context of amplification on the solid substrate, the extended probes may include amplicons of the extended probes, e.g. cDNA.

The extended probes, i.e. DNA, e.g. cDNA molecules, or amplicons, that have been pooled, which may have been modified as discussed above, are analysed to investigate (e.g. determine their sequence, although as noted above actual sequence determination is not required - any method of analysing the sequence may be used). Thus, any method of nucleic acid analysis may be used. The step of sequence analysis may identify the positional domain and hence allow the analysed molecule to be localised to a position in the biological specimen, e.g. tissue sample. Similarly, the nature or identity of the analysed molecule may be determined. In this way the nucleic acid, e.g. RNA, at given position in the array, and hence in the biological specimen, e.g. tissue sample, may be determined. Hence the analysis step may include or use any method which identifies the analysed molecule (and hence the "target" molecule) and its positional domain. Generally such a method will be a sequence-specific method. For example, the

method may use sequence-specific primers or probes, particularly primers or probes specific for the positional domain and/or for a specific nucleic acid molecule to be detected or analysed e.g. a DNA molecule corresponding to a nucleic acid, e.g. RNA or cDNA molecule to be detected. Typically in such a method sequence-specific amplification primers, e.g. PCR primers may be used.

In some embodiments it may be desirable to analyse a subset or family of target related molecules, e.g. all of the sequences that encode a particular group of proteins which share sequence similarity and/or conserved domains, e.g. a family of receptors. Hence, the amplification and/or analysis methods described herein may use degenerate or gene family specific primers or probes that hybridise to a subset of the extended probes or nucleic acids derived therefrom, e.g. amplicons. In a particularly preferred embodiment, the amplification and/or analysis methods may utilise a universal primer (i.e. a primer common to all of the captured sequences) in combination with a degenerate or gene family specific primer specific for a subset of target molecules.

Thus in one embodiment, amplification-based, especially PCR-based methods of sequence analysis are used.

However, the steps of modifying and/or amplifying the extended probes, e.g. cDNA molecules, may introduce additional components into the sample, e.g. enzymes, primers, nucleotides etc. Hence, the methods of the invention may further comprise a step of purifying the sample comprising the extended probes, e.g. cDNA molecules, or amplicons thereof prior to the sequence analysis, e.g. to remove oligonucleotide primers, nucleotides, salts etc. that may interfere with the sequencing reactions. Any suitable method of purifying the extended probes, i.e. DNA, e.g. cDNA molecules, may be used.

Sequence analysis of the released DNA molecules may be direct or indirect. Thus the sequence analysis substrate (which may be viewed as the molecule which is subjected to the sequence analysis step or process) may directly be the extended probe or it may be a molecule which is derived therefrom. Thus, for example in the context of sequence analysis step which involves a sequencing reaction, the sequencing template may be the extended probe or it may be a molecule derived therefrom. For example, a first and/or second strand DNA, e.g. cDNA molecule, may be directly subjected to sequence analysis (e.g. sequencing), i.e. may directly take part in the sequence analysis reaction or process (e.g. the sequencing reaction or sequencing process, or be the molecule which is sequenced or otherwise

identified). Alternatively, the extended probe may be subjected to a step of second strand synthesis or amplification before sequence analysis (e.g. sequencing or identification by other means). The sequence analysis substrate (e.g. template) may thus be an amplicon or a second strand of an extended probe.

Both strands of a double stranded molecule may be subjected to sequence analysis (e.g. sequenced) but the invention is not limited to this and single stranded molecules (e.g. cDNA) may be analysed (e.g. sequenced). For example various sequencing technologies may be used for single molecule sequencing, e.g. the Helicos or Pacbio technologies, or nanopore sequencing technologies which are being developed. Thus, in one embodiment the extended probe, i.e. the first strand of DNA, e.g. cDNA, may be subjected to sequencing. The first strand DNA, e.g. cDNA may need to be modified at the 3' end to enable single molecule sequencing. This may be done by procedures analogous to those for handling the second DNA, e.g. cDNA, strand. Such procedures are known in the art.

In a preferred aspect of the invention the sequence analysis will identify or reveal at least a portion of the extended probe, including at least part of captured nucleic acid, e.g. RNA, sequence and the sequence of the positional domain. The sequence of the positional domain (or tag) will identify the feature (or group of adjacent features) from which the capture probe was released and hence the location or vicinity on the solid substrate at which the nucleic acid, e.g. mRNA, molecule was captured. The sequence of the captured nucleic acid, e.g. RNA, molecule may be compared with a sequence database of the organism from which the biological specimen, e.g. tissue sample, originated to determine the gene to which it corresponds. By determining which region (e.g. cell) of the biological specimen, e.g. tissue sample, was in contact with the feature, it is possible to determine which region of the biological specimen, e.g. tissue sample, was expressing said gene (or contained the gene, e.g. in the case of spatial genomics). This analysis may be achieved for all of the DNA, e.g. cDNA, molecules generated by the methods of the invention, yielding a spatial transcriptome or genome of the biological specimen, e.g. tissue sample.

By way of a representative example, sequencing data may be analysed to sort the sequences into specific species of capture probe, i.e. according to the sequence of the positional domain. This may be achieved by, e.g. using the FastX toolkit FASTQ Barcode splitter tool to sort the sequences into individual files for the respective capture probe positional domain (tag) sequences. The sequences of

each species, i.e. from each feature (or group of adjacent features), may be analyzed to determine the identity of the transcripts. For instance, the sequences may be identified using e.g. Blastn software, to compare the sequences to one or more genome databases, preferably the database for the organism from which the biological specimen, e.g. tissue sample, was obtained. The identity of the database sequence with the greatest similarity to the sequence generated by the methods of the invention will be assigned to said sequence. In general, only hits with a certainty of at least $1e^{-6}$, preferably $1e^{-7}$, $1e^{-8}$, or $1e^{-9}$ will be considered to have been successfully identified.

It will be apparent that any nucleic acid sequencing method may be utilised in the methods of the invention. However, the so-called "next generation sequencing" techniques will find particular utility in the present invention. High-throughput sequencing is particularly useful in the methods of the invention because it enables a large number of nucleic acids to be partially sequenced in a very short period of time. In view of the recent explosion in the number of fully or partially sequenced genomes, it is not essential to sequence the full length of the extended probes, e.g. cDNA molecules, to determine the gene to which each molecule corresponds. For example, the first 100 nucleotides from each end of the DNA, e.g. cDNA, molecules should be sufficient to identify both the feature from which the capture probe was released and hence the location or vicinity on the solid substrate at which the nucleic acid, e.g. mRNA, molecule was captured (i.e. its location on the solid substrate, e.g. array) and the gene expressed. The sequence reaction from the "capture probe end" of the DNA, e.g. cDNA, molecules yields the sequence of the positional domain and at least about 20 bases, preferably 30 or 40 bases of transcript or gene specific sequence data. The sequence reaction from the "non-capture probe end" may yield at least about 70 bases, preferably 80, 90, or 100 bases of transcript or gene specific sequence data.

In some embodiments it may be desirable to sequence the full-length of the extended probes generated by the methods of the invention. As mentioned above, template switching is particularly useful for generating full-length double stranded cDNA molecules and, in some embodiments, it may be desirable or advantageous to sequence the full-length of these molecules. Any suitable sequencing technique may be used to sequence the full-length DNA molecules and representative examples are presented below.

As a representative example, the sequencing reaction may be based on reversible dye-terminators, such as used in the Illumina™ technology. For example, DNA molecules are first attached to primers on, e.g. a glass or silicon slide and amplified so that local clonal colonies are formed (bridge amplification), as described above. Four types of ddNTPs are added, and non-incorporated nucleotides are washed away. Unlike pyrosequencing, the DNA can only be extended one nucleotide at a time. A camera takes images of the fluorescently labelled nucleotides then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing a next cycle. This may be repeated until the required sequence data is obtained. Using this technology, thousands of nucleic acids may be sequenced simultaneously on a single slide.

Other high-throughput sequencing techniques may be equally suitable for the methods of the invention, e.g. pyrosequencing. In this method the DNA is amplified inside water droplets in an oil solution (emulsion PCR), with each droplet containing a single DNA template attached to a single primer-coated bead that then forms a clonal colony. The sequencing machine contains many picolitre-volume wells each containing a single bead and sequencing enzymes. Pyrosequencing uses luciferase to generate light for detection of the individual nucleotides added to the nascent DNA and the combined data are used to generate sequence read-outs.

An example of a technology in development is based on the detection of hydrogen ions that are released during the polymerisation of DNA. A microwell containing a template DNA strand to be sequenced is flooded with a single type of nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide it is incorporated into the growing complementary strand. This causes the release of a hydrogen ion that triggers a hypersensitive ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence multiple nucleotides will be incorporated in a single cycle. This leads to a corresponding number of released hydrogen ions and a proportionally higher electronic signal.

Thus, it is clear that new sequencing formats are being made available, and with shorter run times as one of the main features of those platforms, it will be evident that other sequencing technologies will be useful in the methods of the invention.

An essential feature of the present invention, as described above, is a step of extending the capture probes to produce a complementary strand of the

"captured" nucleic acid molecules (i.e. the molecules hybridised to the capture domain of the capture probe), e.g. reverse transcribing the captured RNA molecules. The reverse transcription reaction is well known in the art and in representative reverse transcription reactions, the reaction mixture includes a reverse transcriptase, dNTPs and a suitable buffer. As mentioned above, in preferred embodiments, the reverse transcription reaction mixture also comprises means for releasing said capture probes from the surface of the solid substrate, e.g. a cleavage enzyme, such as the USER enzyme. The reaction mixture may comprise other components, such as RNase inhibitor(s), inhibitors of DNA-dependent DNA synthesis, e.g. actinomycin D. The primers and template are the capture domain of the capture probe and the captured RNA molecules as described above. In the subject methods, each dNTP will typically be present in an amount ranging from about 10 to 5000 μM, usually from about 20 to 1000 μM. The reaction mixture may be incubated on the solid substrate under conditions suitable for extension of the capture probes and cleavage of said probes from the solid substrate. Typically, the reaction mixture may be incubated on the solid substrate for at least 1 hour, e.g. 1-2 hours or 2, 3, 4 or more hours. However, shorter incubation times may be sufficient, e.g. about 30 minutes, such as 10-30 minutes, 15-30 minutes etc. It will be evident that an equivalent reaction may be performed to generate a complementary strand of a captured DNA molecule, using an enzyme with DNA polymerase activity. Reactions of this type are well known in the art and are described in more detail below.

In some embodiments it may be useful or advantageous to use labelled dNTPs in the extension reaction, e.g. reverse transcription reaction. The labelled dNTPs will be incorporated into the synthesized DNA molecule, e.g. cDNA molecule, thereby labelling the extended probes. In a representative embodiment, the labelled dNTP is a fluorescently labelled dNTP, e.g. Cy3-dCTP.

Thus, in some embodiments, the method further comprises a step of labelling the extended probes generated in step (c). In some embodiments, the labelling step may be contemporaneous with said extending step, i.e. by incorporation of labelled nucleotides into the extended probes. In some embodiments, the labelling step may be subsequent to the extending step, e.g. by using labelled nucleotides in the second strand synthesis and/or amplification steps, thereby incorporating labelled nucleotides into the second strand (i.e. complementary strand of the extended probes) and/or amplicons thereof.

Alternatively, labels (detectable molecules) may be incorporated into the synthesized cDNA by binding to the molecules, e.g., via intercalation.

Representative detectable molecules that may find use in such embodiments include fluorescent nucleic acid stains, such as phenanthridinium dyes, including monomers or homo- or heterodimers thereof, that give an enhanced fluorescence when complexed with nucleic acids. Examples of phenanthridinium dyes include ethidium homodimer, ethidium bromide, propidium iodide, and other alkyl-substituted phenanthridinium dyes. In another embodiment of the invention, the nucleic acid stain is or incorporates an acridine dye, or a homo- or heterodimer thereof, such as acridine orange, acridine homodimer, ethidium-acridine heterodimer, or 9-amino-6-chloro-2-methoxyacridine. In yet another embodiment of the invention, the nucleic acid stain is an indole or imidazole dye, such as Hoechst 33258, Hoechst 33342, Hoechst 34580 (BIOPROBES 34, Molecular Probes, Inc. Eugene, Oreg., (May 2000)) DAPI (4',6-diamidino-2-phenylindole) or DIPI (4',6-(diimidazolin-2-yl)-2-phenylindole). Other permitted nucleic acid stains include, but are not limited to, 7-aminoactinomycin D, hydroxystilbamidine, LDS 751, selected psoralens (furocoumarins), styryl dyes, metal complexes such as ruthenium complexes, and transition metal complexes (incorporating Tb^{3+} and Eu^{3+}, for example). In certain embodiments of the invention, the nucleic acid stain is a cyanine dye or a homo- or heterodimer of a cyanine dye that gives an enhanced fluorescence when associated with nucleic acids. Any of the dyes described in U.S. Pat. No. 4,883,867 to Lee (1989), U.S. Pat. No. 5,582,977 to Yue et al. (1996), U.S. Pat. No. 5,321,130 to Yue et al. (1994), and U.S. Pat. No. 5,410,030 to Yue et al. (1995) (all four patents incorporated by reference) may be used, including nucleic acid stains commercially available under the trademarks TOTO, BOBO, POPO, YOYO, TO-PRO, BO-PRO, PO-PRO and YO-PRO from Molecular Probes, Inc., Eugene, Oreg. Any of the dyes described in U.S. Pat. No. 5,436,134 to Haugland et al. (1995), U.S. Pat. No. 5,658,751 to Yue et al. (1997), and U.S. Pat. No. 5,863,753 to Haugland et al. (1999) (all three patents incorporated by reference) may be used, including nucleic acid stains commercially available under the trademarks SYBR Green, SYBR Gold, EvaGreen, SYTO, SYTOX, PICOGREEN, OLIGREEN, and RIBOGREEN from Molecular Probes, Inc., Eugene, Oreg. In yet other embodiments of the invention, the nucleic acid stain is a monomeric, homodimeric or heterodimeric cyanine dye that incorporates an aza- or polyazabenzazolium heterocycle, such as an azabenzoxazole, azabenzimidazole,

or azabenzothiazole, that gives an enhanced fluorescence when associated with nucleic acids, including nucleic acid stains commercially available under the trademarks SYTO, SYTOX, JOJO, JO-PRO, LOLO, LO-PRO from Molecular Probes, Inc., Eugene, Oreg. The type of nucleic acid stain may be selected based on its capacity to bind to single or double stranded nucleic acid. In embodiments where the first cDNA strand is labelled, it may be preferable to use nucleic acid stains capable of labelling single stranded nucleic acid molecules as the RNA transcript captured on the substrate and used to template cDNA synthesis may be partially or fully degraded.

Further embodiments of labelling and detecting labelled DNA molecules in the context of spatial transcriptomic and genomic methods are described in WO2014/060483, which is incorporated herein by reference.

The reverse transcriptase reaction may be carried out at any suitable temperature, which will be dependent on the properties of the enzyme. Typically, reverse transcriptase reactions are performed between 37-55°C, although temperatures outside of this range may also be appropriate. The reaction time may be as little as 1, 2, 3, 4 or 5 minutes or as much as 48 hours. Typically the reaction will be carried out for between 5-120 minutes, preferably 5-60, 5-45 or 5-30 minutes or 1-10 or 1-5 minutes according to choice. The reaction time is not critical and any desired reaction time may be used.

As indicated above, certain embodiments of the methods include an amplification step, where the copy number of generated DNA, e.g. cDNA, molecules is increased, e.g., in order to enrich the sample to obtain a better representation of the nucleic acids, e.g. transcripts, captured from the biological specimen, e.g. tissue sample. The amplification may be linear or exponential, as desired, where representative amplification protocols of interest include, but are not limited to: polymerase chain reaction (PCR); isothermal amplification, etc.

The polymerase chain reaction (PCR) is well known in the art, being described in U.S. Pat. Nos.: 4,683,202; 4,683,195; 4,800,159; 4,965,188 and 5,512,462, the disclosures of which are herein incorporated by reference. In representative PCR amplification reactions, the reaction mixture that includes the above extended probes, e.g. cDNA molecules, which may be pooled, which are combined with one or more primers that are employed in the primer extension reaction, e.g., the PCR primers that hybridize to the first and/or second amplification domains (such as forward and reverse primers employed in geometric (or

exponential) amplification or a single primer employed in a linear amplification). The oligonucleotide primers with which the extended probes, e.g. cDNA molecules, (hereinafter referred to as template DNA for convenience) is contacted will be of sufficient length to provide for hybridization to complementary template DNA under annealing conditions (described in greater detail below). The length of the primers will depend on the length of the amplification domains, but will generally be at least 10 bp in length, usually at least 15 bp in length and more usually at least 16 bp in length and may be as long as 30 bp in length or longer, where the length of the primers will generally range from 18 to 50 bp in length, usually from about 20 to 35 bp in length. The template DNA may be contacted with a single primer or a set of two primers (forward and reverse primers), depending on whether primer extension, linear or exponential amplification of the template DNA is desired.

In addition to the above components, the reaction mixture produced in the subject methods typically includes a polymerase and deoxyribonucleoside triphosphates (dNTPs). The desired polymerase activity may be provided by one or more distinct polymerase enzymes. In many embodiments, the reaction mixture includes at least a Family A polymerase, where representative Family A polymerases of interest include, but are not limited to: Thermus aquaticus polymerases, including the naturally occurring polymerase (Taq) and derivatives and homologues thereof, such as Klentaq (as described in Barnes et al, Proc. Natl. Acad. Sci USA (1994) 91:2216-2220); Thermus thermophilus polymerases, including the naturally occurring polymerase (Tth) and derivatives and homologues thereof, and the like. In certain embodiments where the amplification reaction that is carried out is a high fidelity reaction, the reaction mixture may further include a polymerase enzyme having 3'-5' exonuclease activity, e.g., as may be provided by a Family B polymerase, where Family B polymerases of interest include, but are not limited to: Thermococcus litoralis DNA polymerase (Vent) as described in Perler et al., Proc. Natl. Acad. Sci. USA (1992) 89:5577-5581; Pyrococcus species GB-D (Deep Vent); Pyrococcus furiosus DNA polymerase (Pfu) as described in Lundberg et al., Gene (1991) 108:1-6, Pyrococcus woesei (Pwo) and the like. Where the reaction mixture includes both a Family A and Family B polymerase, the Family A polymerase may be present in the reaction mixture in an amount greater than the Family B polymerase, where the difference in activity will usually be at least 10-fold, and more usually at least about 100-fold. Usually the reaction mixture will include four different types of dNTPs corresponding to the four naturally occurring bases

present, i.e. dATP, dTTP, dCTP and dGTP. In the subject methods, each dNTP will typically be present in an amount ranging from about 10 to 5000 μM, usually from about 20 to 1000 μM.

The reaction mixtures prepared in the reverse transcriptase and/or amplification steps of the subject methods may further include an aqueous buffer medium that includes a source of monovalent ions, a source of divalent cations and a buffering agent. Any convenient source of monovalent ions, such as KCl, K-acetate, NH_4-acetate, K-glutamate, NH_4Cl, ammonium sulphate, and the like may be employed. The divalent cation may be magnesium, manganese, zinc and the like, where the cation will typically be magnesium. Any convenient source of magnesium cation may be employed, including $MgCl_2$, Mg-acetate, and the like. The amount of Mg^{2+} present in the buffer may range from 0.5 to 10 mM, but will preferably range from about 3 to 6 mM, and will ideally be at about 5 mM. Representative buffering agents or salts that may be present in the buffer include Tris, Tricine, HEPES, MOPS and the like, where the amount of buffering agent will typically range from about 5 to 150 mM, usually from about 10 to 100 mM, and more usually from about 20 to 50 mM, where in certain preferred embodiments the buffering agent will be present in an amount sufficient to provide a pH ranging from about 6.0 to 9.5, where most preferred is pH 7.3 at 72 °C. Other agents which may be present in the buffer medium include chelating agents, such as EDTA, EGTA and the like.

In preparing the reverse transcriptase, DNA extension or amplification reaction mixture of the steps of the subject methods, the various constituent components may be combined in any convenient order. For example, in the amplification reaction the buffer may be combined with primer, polymerase and then template DNA, or all of the various constituent components may be combined at the same time to produce the reaction mixture.

As discussed above, a preferred embodiment of the invention the extended probes, e.g. cDNA molecules, may be modified by the addition of amplification domains to the ends of the nucleic acid molecules, which may involve a ligation reaction.

As is known in the art, ligases catalyze the formation of a phosphodiester bond between juxtaposed 3'-hydroxyl and 5'-phosphate termini of two immediately adjacent nucleic acids. Any convenient ligase may be employed, where representative ligases of interest include, but are not limited to: Temperature

sensitive and thermostable ligases. Temperature sensitive ligases include, but are not limited to, bacteriophage T4 DNA ligase, bacteriophage T7 ligase, and E. coli ligase. Thermostable ligases include, but are not limited to, Taq ligase, Tth ligase, and Pfu ligase. Thermostable ligase may be obtained from thermophilic or hyperthermophilic organisms, including but not limited to, prokaryotic, eukaryotic, or archael organisms. Certain RNA ligases may also be employed in the methods of the invention.

In this ligation step, a suitable ligase and any reagents that are necessary and/or desirable are combined with the reaction mixture and maintained under conditions sufficient for ligation of the relevant oligonucleotides to occur. Ligation reaction conditions are well known to those of skill in the art. During ligation, the reaction mixture in certain embodiments may be maintained at a temperature ranging from about 4°C to about 50°C, such as from about 20°C to about 37°C for a period of time ranging from about 5 seconds to about 16 hours, such as from about 1 minute to about 1 hour. In yet other embodiments, the reaction mixture may be maintained at a temperature ranging from about 35°C to about 45°C, such as from about 37°C to about 42°C, e.g., at or about 38°C, 39°C, 40°C or 41°C, for a period of time ranging from about 5 seconds to about 16 hours, such as from about 1 minute to about 1 hour, including from about 2 minutes to about 8 hours. In a representative embodiment, the ligation reaction mixture includes 50 mM Tris pH7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 mg/ml BSA, 0.25 units/ml RNase inhibitor, and T4 DNA ligase at 0.125 units/ml. In yet another representative embodiment, 2.125 mM magnesium ion, 0.2 units/ml RNase inhibitor; and 0.125 units/ml DNA ligase are employed. The amount of adaptor in the reaction will be dependent on the concentration of the DNA, e.g. cDNA in the sample and will generally be present at between 10-100 times the molar amount of DNA, e.g. cDNA.

By way of a representative example the method of the invention may comprise the following steps:

(a) providing a solid substrate on which multiple species of capture probes are immobilized such that each species occupies a distinct position on the solid substrate, wherein said probes are for a reverse transcription primer extension reaction (preferably where said probes are oriented on the solid substrate to have a free 3' end) and wherein each species of said capture probe comprises a nucleic acid molecule comprising:

(i) a cleavage domain for releasing the capture probe from the surface of the solid substrate,

(ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and

(iii) a capture domain;

(b) contacting said solid substrate with a biological specimen;

(c) imaging the biological specimen on the solid substrate;

(d) releasing said capture probes from the surface of the solid substrate under conditions that allow RNA of the biological specimen to hybridise to the capture domain in said capture probes and simultaneously and/or subsequently extending said capture probes using the RNA hybridised to the capture probes as extension templates to produce extended probes (cDNA molecules) thereby spatially tagging the RNA of the biological specimen;

(e) performing second strand cDNA synthesis on the extended probes; and

(f) analysing the extended probes, e.g. the sequence of the extended probes (cDNA molecules).

By way of an alternative representative example the method of the invention may comprise the following steps:

(a) providing a solid substrate on which multiple species of capture probes are immobilized such that each species occupies a distinct position on the solid substrate, wherein said probes are for a reverse transcription primer extension reaction (preferably where said probes are oriented on the solid substrate to have a free 3' end) and wherein each species of said capture probe comprises a nucleic acid molecule comprising:

(i) a cleavage domain for releasing the capture probe from the surface of the solid substrate,

(ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and

(iii) a capture domain;

(b) contacting said solid substrate with a biological specimen;

(c) optionally rehydrating the tissue sample;

(d) releasing said capture probes from the surface of the solid substrate under conditions that allow RNA of the biological specimen to hybridise to the capture domain in said capture probes and simultaneously and/or subsequently

extending said capture probes using the RNA hybridised to the capture probes as extension templates to produce extended probes (cDNA molecules) thereby spatially tagging the RNA of the biological specimen.

(e) imaging the biological specimen on the solid substrate;

(f) amplifying the extended probes; and

(g) analysing the amplified extended probes, e.g. the sequence of the amplified extended probes (cDNA molecules).

By way of yet a further representative example the method of the invention may comprise the following steps:

(a) providing a solid substrate on which multiple species of capture probes are immobilized such that each species occupies a distinct position on the solid substrate, wherein said probes are for a reverse transcription primer extension reaction (preferably where said probes are oriented on the solid substrate to have a free 3' end) and wherein each species of said capture probe comprises a nucleic acid molecule comprising:

(i) a cleavage domain for releasing the capture probe from the surface of the solid substrate,

(ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and

(iii) a capture domain;

(b) contacting said solid substrate with a biological specimen;

(c) optionally rehydrating the tissue sample;

(d) releasing said capture probes from the surface of the solid substrate under conditions that allow RNA of the biological specimen to hybridise to the capture domain in said capture probes and simultaneously and/or subsequently extending said capture probes using the RNA hybridised to the capture probes as extension templates to produce extended probes (cDNA molecules) thereby spatially tagging the RNA of the biological specimen;

(e) optionally imaging the biological specimen on the solid substrate;

(f) performing second strand cDNA synthesis on the extended probes;

(g) optionally pooling the double stranded extended probes from (f);

(h) amplifying the double stranded extended probes from (f) or (g); and

(i) analysing the amplified extended probes, e.g. the sequence of the amplified extended probes (cDNA molecules).

The present invention includes any suitable combination of the steps in the above described methods. It will be understood that the invention also encompasses variations of these methods, for example where amplification is performed on the solid substrate. Also encompassed are methods which omit the imaging step and methods that comprise a step of correlating sequence analysis information obtain from analysing the extended probes (including double stranded extended probes and/or amplicons thereof) with an image of said biological specimen.

As mentioned above, although the invention is described primarily with reference to detection or analysis of RNA, and transcriptome analysis or detection, it will be appreciated that the principles described can be applied analogously to the detection or analysis of DNA in cells and to genomic studies. Thus, more broadly viewed, the invention can be seen as being generally applicable to the detection of nucleic acids in general and in a further more particular aspect, as providing methods for the analysis or detection of DNA. Spatial information may be valuable also in a genomics context, i.e. detection and/or analysis of a DNA molecule with spatial resolution. This may be achieved by genomic tagging according to the present invention. Such spatial tagging, e.g. localised or spatial detection, methods may be useful for example in the context of studying genomic variations in different cells or regions of a tissue, for example comparing normal and diseased cells or tissues (e.g. normal vs. tumour cells or tissues) or in studying genomic changes in disease progression etc. For example, tumour tissues may comprise a heterogeneous population of cells which may differ in the genomic variants they contain (e.g. mutations and/or other genetic aberrations, for example chromosomal rearrangements, chromosomal amplifications/deletions/insertions etc.). The detection of genomic variations, or different genomic loci, in different cells in a localised way may be useful in such a context, e.g. to study the spatial distribution of genomic variations. A principal utility of such a method would be in tumour analysis. In the context of the present invention, a solid substrate, e.g. array, may be prepared which is designed, for example, to capture the genome of an entire cell on the capture probes from one feature. Different cells in the biological specimen, e.g. tissue sample, may thus be compared. Of course the invention is not limited to such a design and other variations may be possible, wherein the DNA is detected in a localised way and the position of the DNA captured by the capture probes

released from the solid substrate, e.g. array, is correlated to a position or location in the biological specimen, e.g. tissue sample.

The sequence analysis (e.g. sequencing) information obtained in the step of analysing the extended probes (including double stranded extended probes and/or amplicons thereof) may be used to obtain spatial information as to the nucleic acid in the sample. In other words the sequence analysis information may provide information as to the location of the nucleic acid in the sample. This spatial information may be derived from the nature of the sequence analysis information obtained, e.g. from a sequence determined or identified, for example it may reveal the presence of a particular nucleic acid molecule which may itself be spatially informative in the context of the tissue sample used, and/or the spatial information (e.g. spatial localisation) may be derived from the position of the tissue sample on the array, coupled with the sequence analysis information. However, as described above, spatial information may conveniently be obtained by correlating the sequence analysis data to an image of the biological specimen, e.g. tissue sample, and this represents one preferred embodiment of the invention.

The extension reaction referred to in step (c) may be defined as a polymerase-catalysed extension reaction and acts to acquire a complementary strand of the “captured nucleic acid molecule”, i.e. the nucleic acid molecule that is hybridised to the capture domain of the capture probe, i.e. by synthesising the complementary strand utilising the capture probe as a primer and the captured nucleic acid as a template. In other words it may be any primer extension reaction carried out by any polymerase enzyme. Thus, when the captured nucleic acid molecule nucleic acid is DNA, e.g. genomic DNA, the polymerase will be a DNA polymerase.

Thus, in some embodiments the invention provides the method comprises:

(a) providing a solid substrate on which multiple species of capture probes are immobilized such that each species occupies a distinct position on the solid substrate, wherein said probes are for a primer extension reaction (preferably where said probes are oriented on the solid substrate to have a free 3' end) and wherein each species of said capture probe comprises a nucleic acid molecule comprising:

(i) a cleavage domain for releasing the capture probe from the surface of the solid substrate,

(ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and

(iii) a capture domain;

(b) contacting said solid substrate with a biological specimen; and

(c) fragmenting DNA in said biological specimen tissue sample, wherein said fragmentation is carried out before, during or after contacting the solid substrate with the biological specimen in step (b);

(d) releasing said capture probes from the surface of the solid substrate under conditions that allow DNA of the biological specimen to hybridise to the capture domain in said capture probes and simultaneously and/or subsequently extending said capture probes using the DNA hybridised to the capture probes as extension templates to produce extended probes thereby spatially tagging the DNA of the biological specimen;

(e) optionally generating a complementary strand of said tagged DNA and/or optionally amplifying said tagged DNA;

(f) optionally pooling the extended probes (tagged DNA) from (d) or pooling the double stranded extended probes (tagged DNA) and/or amplicons from step (e); and

(g) analysing the extended probes (tagged DNA) from (d) or (f) and/or the double stranded extended probes (tagged DNA) and/or amplicons from step (e) or (f), e.g. the sequence of the extended probes (tagged DNA) and/or double stranded extended probes (tagged DNA) and/or amplicons .

The method may further include a step of:

(h) correlating said sequence analysis information with an image of said biological specimen, wherein the biological specimen is imaged before or after step (d), preferably before step (d).

In the context of spatial genomics, where the target nucleic acid is DNA the inclusion of imaging and image correlation steps may in some circumstances be preferred.

In embodiments in which DNA is captured, the DNA may be any DNA molecule which may occur in a cell. Thus it may be genomic, i.e. nuclear, DNA, mitochondrial DNA or plastid DNA, e.g. chloroplast DNA. In a preferred embodiment, the DNA is genomic DNA.

It will be understood that where fragmentation is carried out after the contacting in step (b), i.e. after the biological specimen is placed on the solid

substrate, e.g. array, fragmentation occurs before the DNA is hybridised to the capture domain. In other words the DNA fragments are hybridised (or more particularly, allowed to hybridise) to the capture domain in said capture probes.

As noted above, in some embodiments it is preferred that the (target) nucleic acid molecules (particularly RNA) in the biological specimen are not treated or modified to enhance or facilitate their interaction with the capture probes prior to contacting the biological specimen with the solid substrate, e.g. contacted with or hybridized to nucleic acids that function as an intermediary between the target nucleic acid of the biological specimen and the capture probes, i.e. that function to indirectly attach or link the target nucleic acid of the biological specimen to the capture probe. However, advantageously, but not necessarily, in a particular embodiment of this aspect of the invention, the DNA fragments of the biological specimen, e.g. tissue sample, may be provided with a binding domain to enable or facilitate their capture by the capture probes. Accordingly, the binding domain is capable of hybridising to the capture domain of the capture probe. Such a binding domain may thus be regarded as a complement of the capture domain (i.e. it may be viewed as a complementary capture domain), although absolute complementarity between the capture and binding domains is not required, merely that the binding domain is sufficiently complementary to allow a productive hybridisation to take place, i.e. that the DNA fragments in the biological specimen, e.g. tissue sample, are able to hybridise to the capture domain of the capture probes. Provision of such a binding domain may ensure that DNA in the sample does not bind to the capture probes until after the fragmentation step. The binding domain may be provided to the DNA fragments by procedures well known in the art, for example by ligation of adaptor or linker sequences which may contain the binding domain. For example a linker sequence with a protruding end may be used. The binding domain may be present in the single-stranded portion of such a linker, such that following ligation of the linker to the DNA fragments, the single-stranded portion containing the binding domain is available for hybridisation to the capture domain of the capture probes. Alternatively and in a preferred embodiment, the binding domain may be introduced by using a terminal transferase enzyme to introduce a polynucleotide tail, e.g. a homopolymeric tail such as a poly-A domain. This may be carried out using a procedure analogous to that described above for introducing a second amplification domain in the context of the RNA methods. Thus, in advantageous embodiments a common binding domain may be

introduced. In other words, a binding domain which is common to all the DNA fragments and which may be used to achieve the capture of the fragments on the solid substrate, e.g. array. As discussed below, in some embodiments, the common binding domain may be introduced to the DNA during (i.e. simultaneously with) the step of fragmenting the DNA, e.g. where the DNA is fragmented using a transposase (e.g. Tn5) the common binding domain may be simultaneously ligated to the DNA fragments.

Where a tailing reaction is carried out to introduce a (common) binding domain, the capture probes on the solid substrate, e.g. array, may be protected from the tailing reaction, i.e. the capture probes may be blocked or masked as described above. This may be achieved for example by hybridising a blocking oligonucleotide to the capture probe, e.g. to the protruding end (e.g. single stranded portion) of the capture probe. Where the capture domain comprises a poly-T sequence for example, such a blocking oligonucleotide may be a poly-A oligonucleotide. The blocking oligonucleotide may have a blocked 3' end (i.e. an end incapable of being extended, or tailed). The capture probes may also be protected, i.e. blocked, by chemical and/or enzymatic modifications, as described in detail above.

Thus, in some embodiments of the invention, the method may comprise a step of providing the DNA fragments with a binding domain which is capable of hybridising to the capture domain. This step may be performed after (or at the same time as) the step of fragmenting the DNA but before the step of releasing and extending the capture probes. In preferred embodiments, the step of providing the DNA fragments with a binding domain which is capable of hybridising to the capture domain is performed after the biological specimen if contacted with the substrate (e.g. array), i.e. the step is performed *in situ* on the substrate.

In the methods of DNA detection set out above, it will be evident that the step of extending the capture probe and the optional step of generating a complementary copy of the tagged DNA or of amplifying the tagged DNA, may involve or require the use of a strand displacing polymerase enzyme. For instance, the fragmented DNA hybridised to the capture probe may be partially double stranded DNA and thus the extension of the capture probe may require the displacement of one of the DNA strands. Suitable strand displacing polymerases are discussed above. In the context of producing complementary strand of the extended probe and/or amplicons of the extended probe, the use of strand

displacement polymerases is to ensure that the positional domain is copied into the complementary copy or amplicon. This will particularly be the case where the capture probe is immobilized on the solid substrate, e.g. array, by hybridisation to a surface probe.

Thus, in one embodiment, the method of the invention may be used for determining and/or analysing all of the genome of a tissue sample, e.g. the global genome of a tissue sample. However, the method is not limited to this and encompasses determining and/or analysing all or part of the genome. Thus, the method may involve determining and/or analysing a part or subset of the genome, e.g. a partial genome corresponding to a subset or group of genes or of chromosomes, e.g. a set of particular genes or chromosomes or a particular region or part of the genome, for example related to a particular disease or condition, tissue type etc. Thus, the method may be used to detect or analyse genomic sequences or genomic loci from tumour tissue as compared to normal tissue, or even within different types of cell in a tissue sample. The presence or absence, or the distribution or location of different genomic variants or loci in different cells, groups of cells, tissues or parts or types of tissue may be examined.

Viewed from another aspect, the method steps set out above can be seen as providing a method of obtaining spatial information regarding the nucleic acids, e.g. genomic sequences, variants or loci of a biological specimen, e.g. tissue sample. Put another way, the methods of the invention may be used for the labelling (or tagging) of genomes, particularly individual or spatially distributed genomes.

Alternatively viewed, the method of the invention may be seen as a method for spatial detection of DNA in a biological specimen, e.g. tissue sample, or a method for detecting DNA with spatial resolution, or for localised or spatial determination and/or analysis of DNA in a biological specimen, e.g. tissue sample. In particular, the method may be used for the localised or spatial detection or determination and/or analysis of genes or genomic sequences or genomic variants or loci (e.g. distribution of genomic variants or loci) in a biological specimen, e.g. tissue sample. The localised/spatial detection/determination/analysis means that the DNA may be localised to its native position or location within a cell or tissue in the biological specimen, e.g. tissue sample. Thus for example, the DNA may be localised to a cell or group of cells, or type of cells in the sample, or to particular regions of areas within a biological specimen, e.g. tissue sample. The native

location or position of the DNA (or in other words, the location or position of the DNA in the biological specimen, e.g. tissue sample), e.g. a genomic variant or locus, may be determined.

The step of fragmenting DNA in a tissue sample may be carried out using any desired procedure known in the art. Thus physical methods of fragmentation may be used, e.g. sonication or ultrasound treatment. Chemical methods are also known. Enzymatic methods of fragmentation may also be used, e.g. with endonucleases, for example restriction enzymes, or transposases, e.g. Tn5. Again methods and enzymes for this are well known in the art. Notably a transposase, such as Tn5, may advantageously result in the simultaneous fragmentation of DNA and ligation of adapters (e.g. comprising common binding domains) to the 5' ends of both strands of DNA fragments, which may facilitate the capture of the fragmented DNA by the capture probes, i.e. the adapters ligated to the 5' ends of the DNA may contain a domain that is capable of interacting with (i.e. hybridizing to) the capture domain of the capture probes. Fragmentation may be done before during or after preparing the biological specimen, e.g. tissue sample, for placing on a solid substrate, e.g. array, e.g. preparing a tissue section. Conveniently, fragmentation may be achieved in the step of fixing tissue. Thus for example, formalin fixation will result in fragmentation of DNA. Other fixatives may produce similar results.

In terms of the detail of preparing and using the solid substrate, e.g. array, in these aspects of the invention, it will understood that the description and detail given above in the context of RNA methods applies analogously to the more DNA detection methods set out herein. Thus, all aspects and details discussed above apply analogously. For example, the discussion of reverse transcriptase primers and reactions etc. may be applied analogously to any aspect of the extension primers, DNA polymerase reactions etc. referred to above. Likewise, references and to first and second strand cDNA synthesis may be applied analogously to the tagged DNA molecule and its complement. Methods of sequence analysis as discussed above may be used.

By way of example, the capture domain may be as described for the capture probes above. A poly-T or poly-T-containing capture domain may be used for example where the DNA fragments are provided with a binding domain comprising a poly-A sequence.

The capture probes/tagged DNA molecules (i.e. the extended probes comprising complements of captured DNA molecules) may be provided with amplification domains as described above.

The invention will be further described with reference to the following non-limiting Examples with reference to the following drawings in which:

Figure 1 shows a schematic of the experimental design described in Example 2.

Figure 2 shows a fluorescence microscopy image of an array on which cDNA has been synthesised (incorporating nucleotides labelled with Cy3) according to the methods described in Example 2. (A) shows an image following cDNA synthesis (C) and (D) or combined cDNA synthesis and release of the capture probes (E). (B) shows an image following release of the extended capture probes (C) and (D) or after a further incubation period (E).

Figure 3 shows bar charts showing: (A) the total number of unique transcripts identified following separate extension and release steps (ST1.0) and following combined extension and release steps (ST2.0); and (B) the total number of unique genes identified following separate extension and release steps (ST1.0) and following combined extension and release steps (ST2.0).

Figure 4 shows heat maps showing the number transcripts captured at each feature from mouse brain tissue sections using: (A) separate extension and release steps (ST1.0); and (B) combined extension and release steps (ST2.0).

Figure 5 shows an overlay of an image of mouse brain tissue and Penk transcripts captured at features on the array contacted with said tissue using: (A) separate extension and release steps (ST1.0); and (B) combined extension and release steps (ST2.0).

Example 1

The protocol described below is representative of the “standard” spatial transcriptomic methods in which the steps of cDNA synthesis and release of the extended probes are performed separately, known as “ST1.0”.

Preparation of in-house printed microarray with 5’ to 3’ oriented probes

The RNA-capture oligonucleotide (Table 1) was printed on glass slides to function as the capture probe. The probe was synthesized with a 5'-terminus amino linker with a C6 spacer. 1007 capture oligonucleotides with 18-mer unique barcodes, a 9-mer semi-randomized UMI and a poly-20TVN capture region were

arrayed onto the surface of Codelink Activated Slides according to the manufacturer's (Surmodics) instructions. The oligonucleotides were immobilized in 100µm features with 200µm center-to-center distance (pitch) between the spots. According to the slide manufacturer approximately 200 million oligonucleotides were immobilized in each spot. An area in the top left corner of the array was left blank for the purpose of orientation. Six 6200 x 6600 µm arrays were prepared per glass slide. After printing, surface blocking was performed according to the manufacturer's instructions.

Table 1

Probe 1

Amino-C6-UUUUUGACTCGTAATACGACTCACTATAGGGACACGACGCTCTTCCGATCT GTCCGATATGATTGCCGCWSNNWSNNVTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 3)

Collection and preparation of olfactory bulbs

Adult C57BL/6 mice (>2 months old) were euthanized and olfactory bulbs were immediately isolated and snap-frozen in Isopentane. Tissue was embedded in cold OCT before sectioning. The olfactory bulb was sectioned on a cryostat at a thickness of 10µm. Sections were mounted onto spatially barcoded arrays.

Permeabilization and reverse transcription

For each well, corresponding to each sub-array with a section, 70µl of 1x Exonuclease I Reaction Buffer (#B0293S, NEB) with 0.19µg/µl BSA (#B9000S, NEB) was added and incubated at 37°C for 30 minutes. Each well was washed with 100µl 0.1x SSC, diluted in deionized water from stock solution (#S6639, Sigma-Aldrich). Next, 70µl of 0.1% pepsin (#P7000-25G, Sigma-Aldrich) dissolved in 0.1M HCl (#318965-1000ML, Sigma-Aldrich) was added to each well and incubated at 37°C for 10 minutes. Each well was washed as previously described and 70µl of reverse transcription mix was added to each well and incubated overnight (ON).

The reverse transcription mix contained 1x First Strand Buffer (#18080-044, Invitrogen), 5mM DTT (#18080-044, Invitrogen), 500µM of dATP/dGTP/dTTP, 12.5µM of dCTP, 25µM of Cyanine 3-dCTP (#NEL576001EA, PerkinElmer), 0.19µg/µl BSA, 50ng/µl Actinomycin D (#A1410-2MG, Sigma-Aldrich), 1% DMSO (#472301-500ML, Sigma-Aldrich), 20U/µl Superscript III (#18080-04, Invitrogen) and 2U/µl RNaseOUT (#10777-019, Invitrogen). The permeabilization was carried

out for 2, 10 or 30 minutes. The cDNA synthesis was carried out for 1 hour, 3 hours, 6 hours or overnight.

To degrade and remove mouse tissue, 70µl Proteinase K (#19131, Qiagen) and PKD Buffer (pH 7.5, according to the manufacturer, #1034963, Qiagen), at a ratio of 1:7, were added to each well and incubated at 56°C for 1 hour. After the incubation, slides washed with 2x SSC with 0.1% SDS at 50°C (#71736-100ML, Sigma-Aldrich) followed by 0.2x SSC and 0.1x SSC at room temperature (RT) and finally spin-dried.

Imaging

The array was imaged at 532 nm using an Agilent microarray scanner at 100% exposure and 5µm resolution.

Release of cDNA from the array

The array was attached to an ArrayIt slide holder and 16 well mask (ArrayIt Corporation). A cleavage mixture (50µl) containing 1x Exo I buffer (New England Biolabs, Ipswich, MA, USA), 1x BSA, RNase/DNase free water, 5U of USER enzyme mix (New England Biolabs) was added to each well. The reactions were covered with a plastic sealer and incubated at 37°C for 1 hour using interval mixing of 3 seconds at 300rpm and 6 seconds rest. After the incubation, 45µl cleavage mixture was collected from each of the used wells and placed into 0.2ml PCR tubes.

Example 2

An experiment was designed to determine the effects of performing the extension and release steps simultaneously. A schematic of the experimental design is shown in Figure 1, which depicts six different treatments: C1, C2, D1, D2, E1 and E2.

C1 and C2 relate to the “standard” protocol described in Example 1 above, using the cDNA synthesis mix set out below:

a) First strand buffer	64µL
b) DTT	16µL
c) RNase & DNase free water	133µL
d) Actinmycin D	32µL
e) dNTP mix	16µL
f) Cy3 dCTP	8µL
g) BSA	3µL

h) RNase OUT 16µL

i) SuperScript® III 32µL

The tissue permeabilization step included treatment with an exonuclease buffer (C1) or exonuclease buffer comprising exonuclease I.

D1 and D2 differ from C1 and C2 only with respect to the cleavage step, wherein the cleavage mixture comprised the USER enzyme in the cDNA buffer mix. The cleavage mix is set out below:

a) First strand buffer 32µL

b) DTT 8µL

c) RNAse & DNase free water 70.5µL

d) Actinomycin D 16µL

e) dNTPs 8µL

f) BSA 1.5µL

g) USER enzyme 16µL

E1 and E2 differ from C1 and C2 insofar as the cDNA synthesis mix included the cleavage enzyme, i.e. the USER enzyme. The reaction mixture for combined, i.e. simultaneous, cDNA synthesis and probe release is set out below:

a) First strand buffer 32µL

b) DTT 8µL

c) RNase & DNase free water 50.5µL

d) Actinmycin D 16µL

e) dNTP mix 8µL

f) Cy3 dCTP 4µL

g) BSA 1.5µL

h) RNase OUT 8µL

i) SuperScript® III 16µL

j) USER enzyme 16µL

The protocols E1 and E2 did not include the tissue removal step. The E protocol is also referred to herein as ST2.0.

For each well, the release mixture was added and incubated at 37°C for 1-2 hours.

Figure 2 shows the Cy3 signal from the surface of the arrays, i.e. the fluorescent label incorporated into the extended probes, after cDNA synthesis (A) and about 3 hours after the cleavage reaction. Figure 2 demonstrates that protocol

C results in cDNA synthesis on the array surface (A) and the subsequent release of the extended probes from the array surface (B), as expected.

Protocol D shows that the cleavage enzyme (USER) is able to function efficiently in the reverse transcriptase buffer, as the extended probes are released from the surface of the array with similar efficacy to the cleavage reaction in protocol C.

Protocol E shows that the capture probes were successfully cleaved simultaneously with cDNA synthesis. In this respect, the extended probes from protocols C and E were pooled and subjected to sequence analysis to confirm that the cDNA was successfully synthesized in protocol E.

Collection of extended probe mixture

65µl of the extended probe mixture was collected from each of the used wells and placed into 0.2ml PCR tubes.

cDNA second strand synthesis for preparation of a cDNA library for sequencing

5µl of second strand mix containing 2.7x First Strand Buffer, 3.7U/µl DNA polymerase I (#18010-017, Invitrogen) and 0.18U/µl RNaseH (#18021-014, Invitrogen) was added and samples were incubated at 16°C for 2 hours, 5µl T4 DNA polymerase (#M0203S, NEB) was added and samples were incubated at 16°C for another 20 minutes. 25µl of deionized water or 80mM EDTA (#15575- 038, Invitrogen) was added and purified using Agencourt RNAClean XP beads (#A63987, Beckman Coulter) according to the manufacturer's protocol and eluted into deionized water. 5.6µl of the samples were mixed with 10.4µl In Vitro Transcription mix with a final content of 1x T7 Reaction Buffer (#AM1333, Ambion), 7.5mM of each NTP (#AM1333, Ambion), 1x T7 Enzyme Mix (#AM1333, Ambion) and 1U/µl SUPERaseIN (#AM2694, Ambion). Samples were incubated at 37°C for 14 hours.

The samples were purified using Agencourt RNAClean XP beads according to the manufacturer's protocol and eluted into 10µl deionized water. The amount and average fragment length of amplified RNA (aRNA) was determined by using the RNA 6000 Pico Kit (#5067-1513, Agilent) with a 2100 Bioanalyzer (Agilent) according to the manufacturer's protocol. The remaining sample and 2.5µl Ligation adapter was added to a final concentration of 0.71µM. Samples were heated at 70°C for 2 minutes then placed on ice before 4.5µl ligation mix was added to a final content of 1x T4 RNA Ligase Reaction Buffer (#B0216L, NEB), 20U/µl T4 RNA

Ligase2, truncated (#M0242L, NEB), 4U/µl RNase Inhibitor, Murine (#M0314L, NEB) and 0.5µM Ligation Adapter. Samples were incubated at 25°C for 1 hour. The samples were purified using Agencourt RNAClean XP beads according to the protocol previously described and mixed with 1µl RT-primer (IDT) to a final concentration of 1.7µM and 1µl dNTPs and a final concentration 0.83mM of each dNTP. Samples were heated at 65°C for 5 min and then placed on ice, 8µl reverse transcription mix was added to a final content of 1x First Strand Buffer, 0.05M DTT, 500µM of each dNTP, 1mM RT-primer, 10U/µl Superscript III and 2U/µl RNaseOUT. Samples were incubated at 50°C for 1 hour before placed on ice and the samples were subsequently purified using Agencourt RNAClean XP beads according to the protocol previously described. A total reaction volume of 10µl containing 1xKAPA HiFi HotStart ReadyMix (#KK2601, KAPA Biosystems), 1x EVA green (#31000, Biotium), 0.5µM PCR InPE1.0 (Eurofins), 0.01µM PCR InPE2.0 (Eurofins), 0.5µM PCR Index (Eurofins) and 2µl purified cDNA were amplified by qPCR with the following protocol: 98°C for 3 minutes, followed by cycling at 98□C for 20 seconds, 60°C for 30 seconds and 72°C for 30 seconds. After determining the amount of cycles needed based on the qPCR, the samples were amplified in a total reaction volume of 25µl. The libraries were purified and samples were eluted in 20µl elution buffer (#19086, Qiagen) and the average fragment length of finished libraries was determined by using the DNA 1000 Kit (#5067-1504, Agilent) with a 2100 Bioanalyzer according to the manufacturer's protocol. The concentration of the finished libraries was determined with Qubit dsDNA HS Assay Kit (#Q32854, Life Technologies) according to the manufacturer's protocol. Finished libraries were diluted to 4nM and sequenced on the Illumina NextSeq platform using paired-end sequencing according to the manufacturer's protocol. Typically, 31 or 51 bases were sequenced on read one, and 121 or 101 bases were sequenced on read two. The oligonucleotides used during library preparation were:

Ligation adapter:
rApp]AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC[ddC] (SEQ ID NO: 4)

Second reverse transcription primer:
TGACTGGAGTTCAGACGTGTGCTCTTCCGA (SEQ ID NO: 5)

PCR primer InPE1.0:
ATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCT (SEQ ID NO: 6)

PCR primer InPE2.0:
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (SEQ ID NO: 7)
PCR Index primer:
AAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTC (SEQ ID NO: 8)

The results in Figure 3 surprisingly demonstrate that the simultaneous cleavage and extension of the capture probes (protocol E, i.e. ST2.0) resulted in an increase in the total number of transcripts (i.e. extended probes) produced (Figure 3A) and also an increase in the number of unique genes detected, i.e. the simultaneous cleavage and extension of the capture probes enhanced the diversity of the nucleic acids captured from the tissue sample.

Example 3

The standard ST1.0 protocol (i.e. separate extension and cleavage reactions) and ST2.0 protocol (i.e. the simultaneous cleavage and extension of the capture probes) were conducted on mouse brain tissue using an array with capture probes containing positional domains. Figure 4 shows heat maps of the sequencing data obtained for each feature on the array using ST1.0 (A) and ST2.0 (B). Figure 5 shows an overlay of the expression data for the proenkephalin (Penk) gene on an image of the mouse brain tissue obtained using ST1.0 (A) and ST2.0 (B).

It is evident from the data in Figure 4 that the majority of the extended probes correlate directly to the part of the array in contact with the tissue sample. Thus the results in Figure 4 indicate that the released primers in the ST2.0 protocol are able to meet with, and hybridise to, transcripts under the tissue. Furthermore, the data in Figure 4 suggests that an increased number of probes are extended using the ST2.0 protocol relative to the ST1.0 protocol.

Figure 5 demonstrates that unique transcripts from the Penk gene can be spatially detected using both the ST 1.0 and ST 2.0 protocols. Both protocols show that the Penk transcripts are located in the central pear shaped structure of the tissue, demonstrating that both protocols facilitate the spatial tagging of nucleic acid molecules from a biological specimen.

Claims

1. A method for spatially tagging nucleic acids of a biological specimen comprising:

(a) providing a solid substrate on which multiple species of capture probes are immobilized such that each species occupies a distinct position on the solid substrate, wherein said probes are for a primer extension reaction and wherein each species of said capture probes comprise a nucleic acid molecule comprising:

(i) a cleavage domain for releasing the capture probe from the surface of the solid substrate,

(ii) a positional domain that corresponds to the position of the capture probe on the solid substrate, and

(iii) a capture domain;

(b) contacting said solid substrate with a biological specimen; and

(c) releasing said capture probes from the surface of the solid substrate under conditions that allow nucleic acids of the biological specimen to hybridise to the capture domain in said capture probes and simultaneously and/or subsequently extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates to produce extended probes thereby spatially tagging the nucleic acids of the biological specimen,

wherein step (c) comprises contacting said solid substrate with an aqueous reaction mixture comprising:

(i) a polymerase enzyme capable of extending said capture probes using the nucleic acid molecules hybridised to the capture probes as extension templates; and

(ii) means for releasing said capture probes from the surface of the solid substrate.

2. The method of claim 1, wherein said means for releasing said capture probes from the surface of the solid substrate comprises a cleavage enzyme capable of specifically cleaving the capture probe within the cleavage domain.

3. The method of claim 2, wherein said cleavage domain comprises a poly-U sequence and said cleavage enzyme comprises a mixture of Uracil DNA glycosylase (UDG) and an endonuclease enzyme capable of recognising

apurinic/apyrimidinic (AP) sites of dsDNA, preferably a DNA glycosylase-lyase enzyme such as Endonuclease VIII.

4. The method of claim 2, wherein said cleavage domain comprises a restriction endonuclease recognition sequence and the cleavage enzyme is a restriction endonuclease.

5. The method of any one of claims 1 to 4, wherein said polymerase enzyme is a reverse transcriptase enzyme.

6. The method of claim 5, wherein said reverse transcriptase enzyme is selected from the list consisting of M-MLV, MuLV, AMV and HIV reverse transcriptase enzymes and derivatives or mutants thereof, preferably wherein said derivatives are sequence-modified derivatives.

7. The method of claim 5 or 6, wherein the reverse transcriptase enzyme comprises a polypeptide sequence as set forth in SEQ ID NO: 2 or a polypeptide sequence that has at least 80% sequence identity thereto.

8. The method of any one of claims 5 to 7, wherein the reverse transcriptase enzyme is encoded by a nucleotide sequence comprising a sequence as set forth in SEQ ID NO: 1 or a nucleotide sequence that has at least 80% sequence identity thereto.

9. The method of any one of claims 1 to 8, wherein said method further comprises a step of synthesizing a complementary strand of said extended probes to produce double stranded extended probes.

10. The method of any one of claims 1 to 9, wherein said method further comprises a step of amplifying the extended probes and/or double stranded extended probes.

11. The method of claim 10, wherein the step of amplifying the extended probes and/or double stranded extended probes comprises a PCR.

12. The method of any one of claims 1 to 11, wherein said method further comprises a step of analysing the extended probes and/or double stranded extended probes and/or amplicons thereof.

13. The method of claim 12, wherein said step of analysing the extended probes and/or double stranded extended probes and/or amplicons thereof comprises analysing the sequences of the extended probes and/or double stranded extended probes and/or amplicons thereof.

14. The method of claim 12 or 13, wherein said method further comprises a step of correlating said analysis information, e.g. sequence analysis information, with an image of said biological specimen, wherein the biological specimen is imaged before or after step (c).

15. The method of any one of claims 1 to 14, further comprising a step of labelling the extended probes generated in step (c) and/or complementary strands and/or amplicons thereof, wherein said labelling step may be contemporaneous with, or subsequent to, said extending step.

16. The method of any one of claims 1 to 15, wherein the capture probes are immobilized on the substrate of the array directly resulting in a free extendible 3' end.

17. The method of any one of claims 1 to 16, wherein the capture probes are DNA molecules.

18. The method of any one of claims 1 to 17, wherein the capture probes further comprise an amplification domain.

19. The method of any one of claims 1 to 18, wherein the positional domain of each species of capture probe comprises a barcode sequence.

20. The method of any one of claims 1 to 19, wherein a species of capture probe comprises capture domains with different sequences.

21. The method of any one of claims 1 to 20, wherein the capture domain comprises a poly-T DNA oligonucleotide comprising at least 10 deoxythymidine residues and/or a random or degenerate oligonucleotide sequence.

22. The method of any one of claims 1 to 21, wherein the capture domain comprises a sequence specific for a particular target gene or group of genes.

23. The method of any one of claims 1 to 22, wherein the solid substrate is located within a flow cell apparatus.

24. The method of any one of claims 1 to 23, wherein the solid substrate comprises an array.

25. The method of claim 24, wherein the array is a bead array.

26. The method of any one of claims 1 to 25, wherein the capture probes are immobilized on the solid substrate in an ordered arrangement.

27. The method of any one of claims 1 to 25, wherein the capture probes are immobilized on the solid substrate in a random arrangement.

28. The method of claim 27, wherein said method further comprises a step of performing a nucleic acid detection reaction on the solid substrate to determine the positional domain sequences of the randomly located probes on the solid substrate.

29. The method of any one of claims 1 to 28, wherein the capture probes are synthesized on the solid substrate using bridge amplification.

30. The method of any one of claims 1 to 29, wherein the biological specimen is a tissue section.

31. The method of claim 30, wherein the tissue section is prepared using a fixed tissue, e.g. a formalin-fixed paraffin-embedded (FFPE) tissue, or deep-frozen tissue.

32. The method of any one of claims 1 to 31, further comprising a step of rehydrating the biological specimen after contacting the specimen with the solid substrate and prior to step (c).

33. The method of any one of claims 1 to 32, further comprising a step of permeabilizing and/or fixing the biological specimen after contacting the specimen with the solid substrate and prior to step (c).

34. The method of any one of claims 1 to 33, wherein the solid substrate comprises at least one positional marker to enable orientation of the biological specimen on the solid substrate.

35. The method of any one of claims 1 to 34, furthering comprising a step of collecting the extended probes from step (c) and/or double stranded extended probes and/or amplicons thereof.

36. The method of any one of claims 14 to 35, wherein the biological specimen is imaged using light, bright field, dark field, phase contrast, fluorescence, reflection, interference or confocal microscopy or a combination thereof.

37. The method of claim 36, wherein the biological specimen is imaged using light microscopy or fluorescence microscopy.

38. The method of any one of claims 13 to 37, wherein the sequence analysis step includes a step of sequencing, preferably wherein the sequencing step comprises a sequencing reaction based on reversible dye-terminators.

39. Use of an aqueous reaction mixture comprising:
(i) a polymerase enzyme, preferably a reverse transcriptase enzyme; and
(ii) cleavage enzyme, preferably a cleavage enzyme as defined in claim 3 or 4,
in a method for spatially tagging nucleic acids of a biological specimen.

40. The use of claim 39, wherein said method for spatially tagging nucleic acids of a biological specimen is as defined in any one of claims 1 to 38.

Figure 1

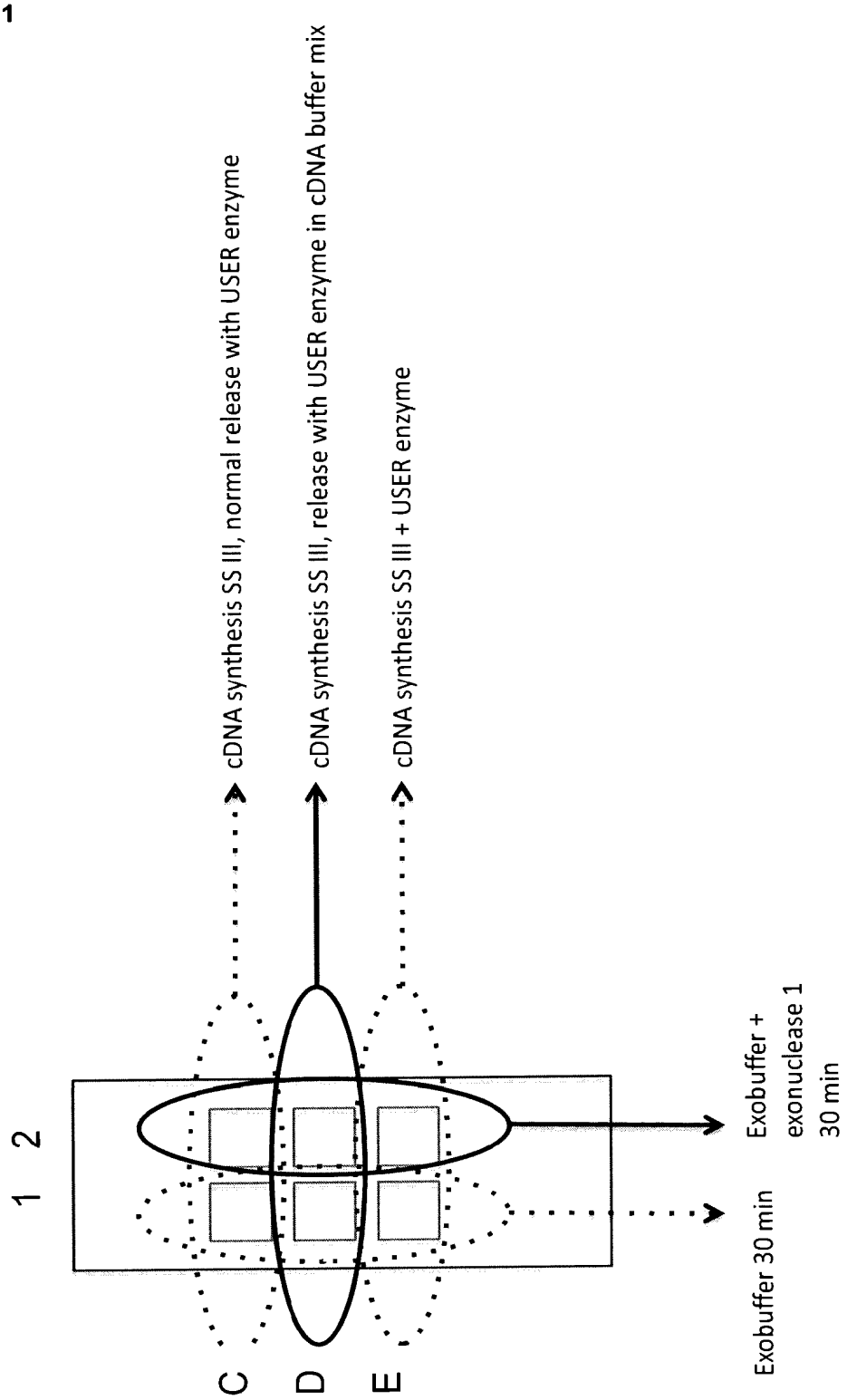

cDNA synthesis SS III, normal release with USER enzyme
cDNA synthesis SS III, release with USER enzyme in cDNA buffer mix
cDNA synthesis SS III + USER enzyme
1 2
C
D
E
Exobuffer + exonuclease 1 30 min
Exobuffer 30 min

Figure 2

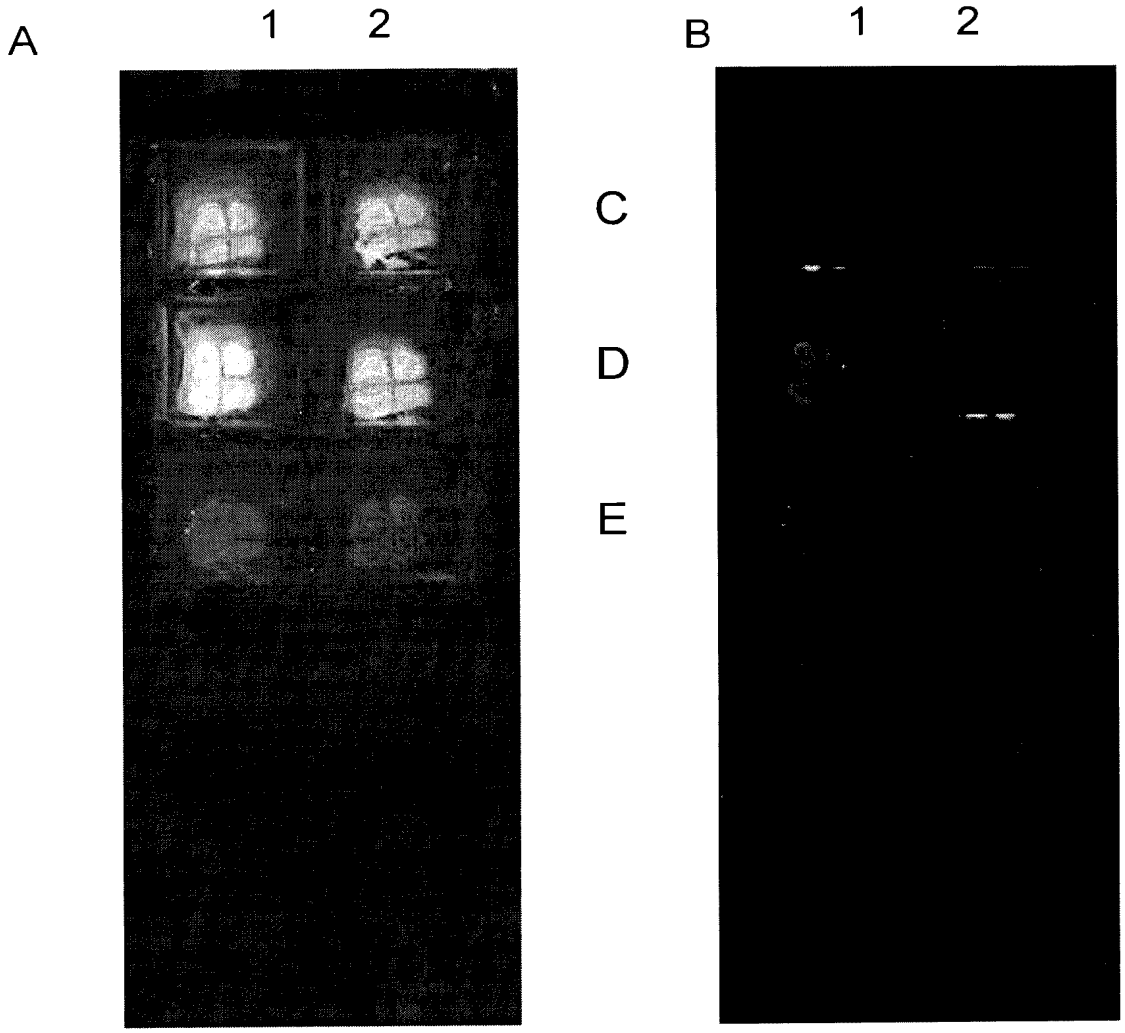


Figure 3

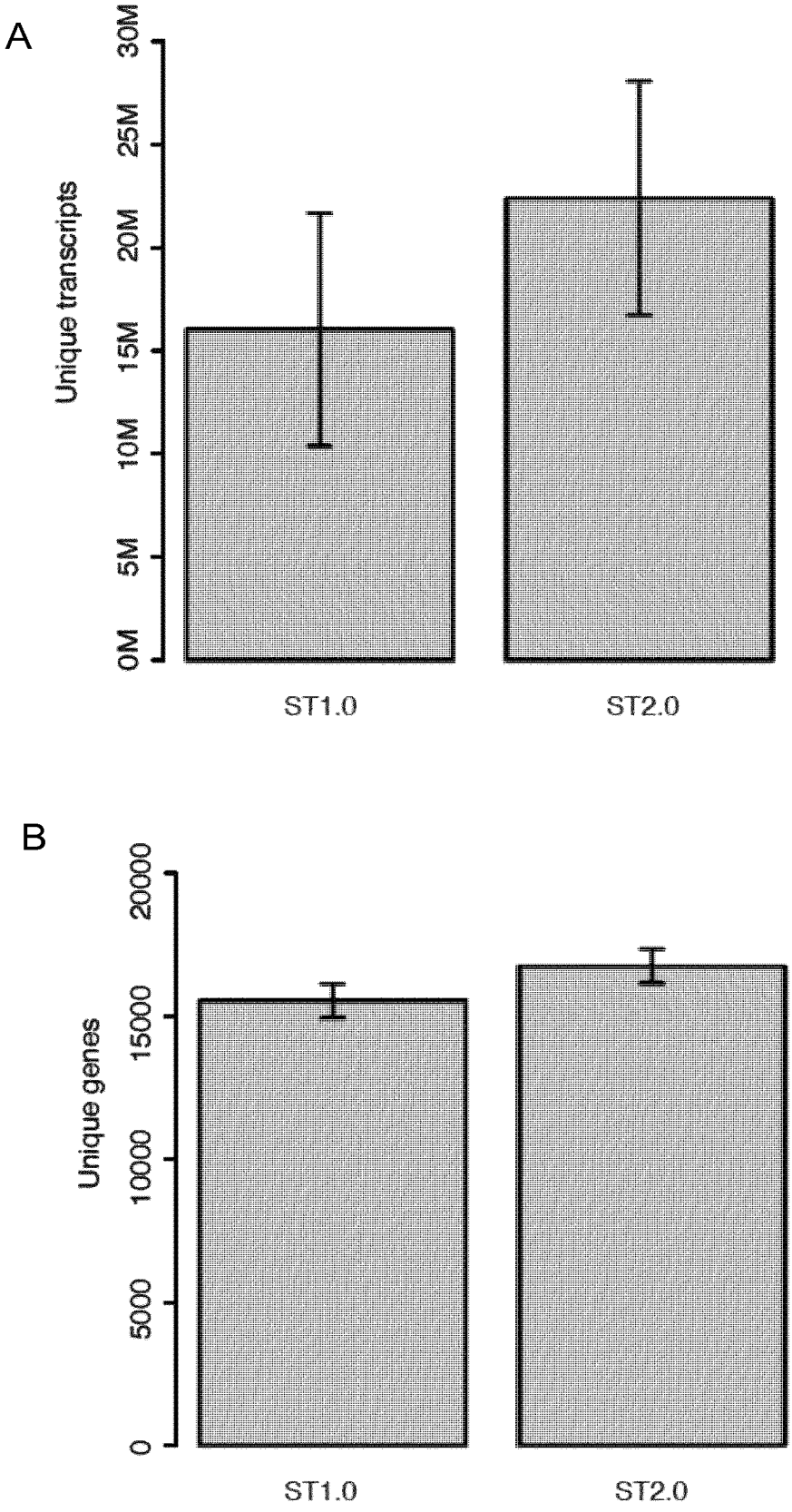
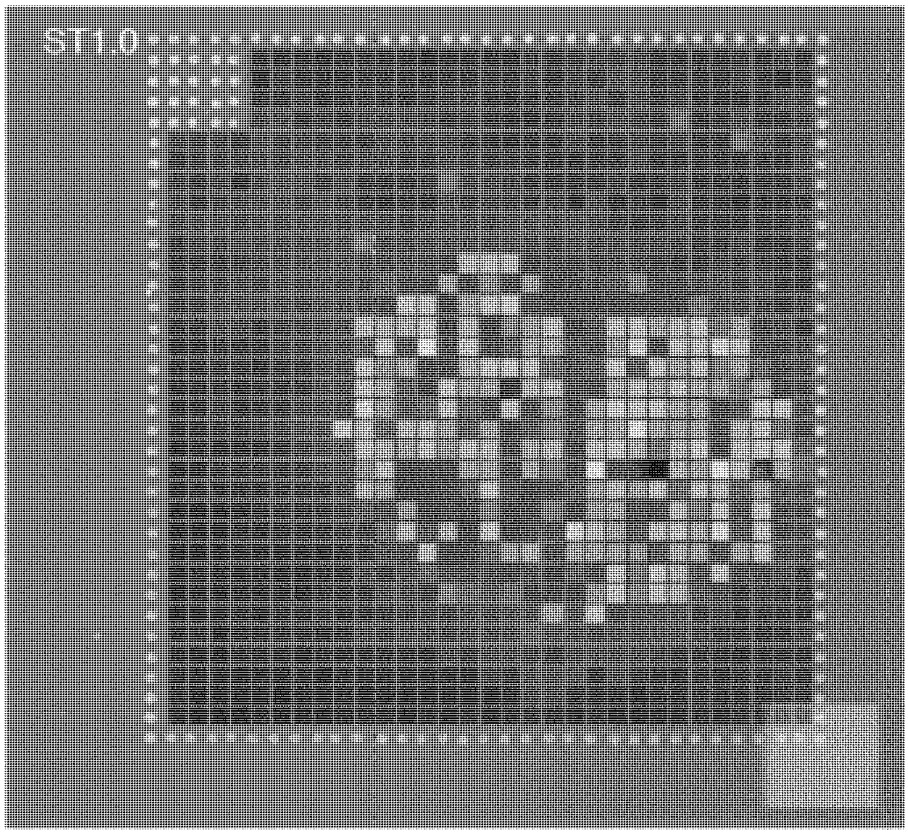

A
Unique transcripts
0M
5M
10M
15M
20M
25M
30M
ST1.0
ST2.0
B
Unique genes
0
5000
10000
15000
20000
ST1.0
ST2.0

Figure 4

A



B

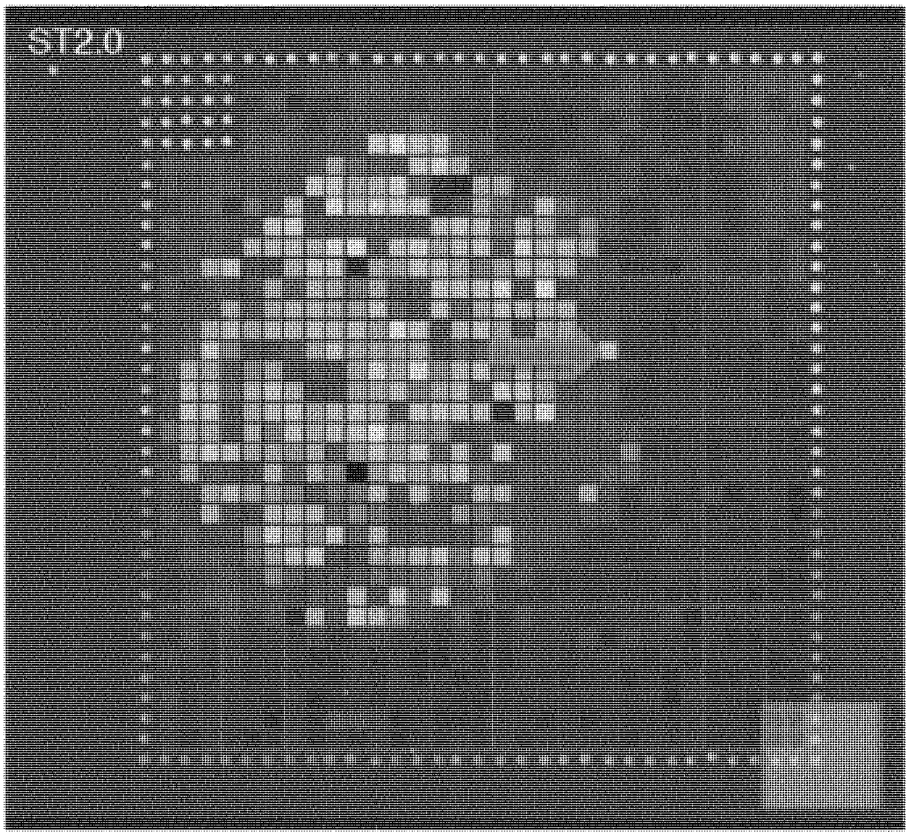
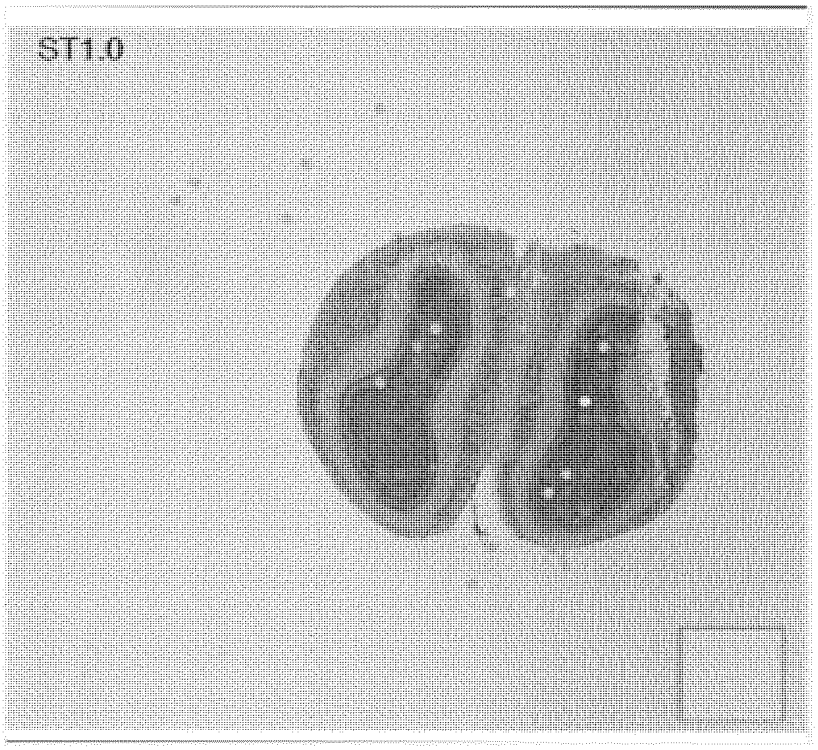
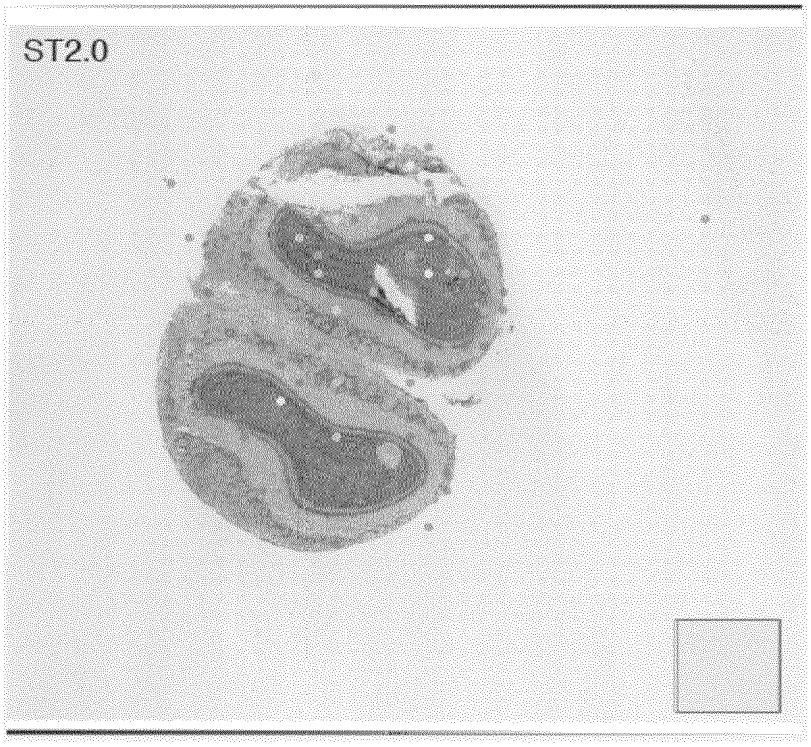


Figure 5

A



B



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/079650

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6834 C12Q1/6841 C12Q1/6874
ADD.

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

B. FIELDS SEARCHED

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

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2016/168825 A1 (CENTRILLION TECH HOLDINGS CORP; ZHOU WEI [US]) 20 October 2016 (2016-10-20)	39,40
A	paragraph [0105]; claim 5	1-38
X	WO 2012/140224 A1 (FRISEN JONAS [SE]; STAAHL PATRIK [SE]; LUNDEBERG JOAKIM [SE]) 18 October 2012 (2012-10-18)	39,40
A	page 23, paragraph 1	1-38
X	WO 2016/162309 A1 (SPATIAL TRANSCRIPTOMICS AB [SE]; ILLUMINA INC [US]) 13 October 2016 (2016-10-13)	39,40
A	page 36	1-38
X,P	WO 2017/019456 A2 (ILLUMINA INC [US]) 2 February 2017 (2017-02-02) paragraph [0343]; figures 30,44	1

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 February 2018	07/03/2018

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Aslund, Fredrik

2

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2017/079650

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2016168825 A1	20-10-2016	CN 107636169 A	26-01-2018
		EP 3283656 A1	21-02-2018
		WO 2016168825 A1	20-10-2016
WO 2012140224 A1	18-10-2012	AU 2012241730 A1	31-10-2013
		BR 112013026502 A2	29-11-2016
		CA 2832678 A1	18-10-2012
		CN 103781918 A	07-05-2014
		EP 2697391 A1	19-02-2014
		JP 5916166 B2	11-05-2016
		JP 2014513523 A	05-06-2014
		KR 20140024378 A	28-02-2014
		MX 340330 B	06-07-2016
		NZ 616407 A	24-12-2014
		RU 2013148909 A	20-05-2015
		US 2014066318 A1	06-03-2014
		WO 2012140224 A1	18-10-2012
WO 2016162309 A1	13-10-2016	CA 2982146 A1	13-10-2016
		CN 107532207 A	02-01-2018
		EP 3280820 A1	14-02-2018
		SG 11201707515S A	30-10-2017
		WO 2016162309 A1	13-10-2016
WO 2017019456 A2	02-02-2017	AU 2016298158 A1	15-02-2018
		CA 2993463 A1	02-02-2017
		WO 2017019456 A2	02-02-2017

Electronic Patent Application Fee Transmittal

Application Number:	16272487
Filing Date:	11-Feb-2019
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Filer:	Robert E. Powers/Sylvana Householder
Attorney Docket Number:	NATE-037001US 321329-2523

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Miscellaneous:				
SUBMISSION- INFORMATION DISCLOSURE STMT	1806	1	260	260
		Total in USD ($)		**260**

Electronic Acknowledgement Receipt

EFS ID:	43359054
Application Number:	16272487
International Application Number:	
Confirmation Number:	6864
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Customer Number:	141429
Filer:	Robert E. Powers/Sylvana Householder
Filer Authorized By:	Robert E. Powers
Attorney Docket Number:	NATE-037001US 321329-2523
Receipt Date:	28-JUL-2021
Filing Date:	11-FEB-2019
Time Stamp:	10:44:01
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$260
RAM confirmation Number	E20217RA44191066
Deposit Account	
Authorized User	

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	NATE-037-001US_321329-2523_IDSTransmittal.pdf	120144 824ac7ce441bcf60339cd032ce39ecb4c9fc0ce3	no	3
Warnings:					
Information:					
2	Information Disclosure Statement (IDS) Form (SB08)	NATE-037-001US_321329-2523_sb08.pdf	251543 3bf28327e9a45e4dcc16b974c8b3503e1af0af42	no	2
Warnings:					
Information:					
This is not an USPTO supplied IDS fillable form					
3	Foreign Reference	WO2009156725A1.pdf	4302554 2e56cb028e3e71cb530aa812651322f1933113fc	no	100
Warnings:					
Information:					
4	Foreign Reference	WO2010081114A2.pdf	2138252 80f665bd1e269e780ff0888c2d13e2bc9b438f76	no	46
Warnings:					
Information:					
5	Foreign Reference	WO2012140224A1.pdf	13169954 250bbb15a1ae3bba8e58b66cd2d4d4128e0a599d	no	153
Warnings:					
Information:					
6	Foreign Reference	WO2014060483A1.pdf	13084025 7266b9749e993fc5f9f9fa38534e5055bf8c3db2	no	144
Warnings:					

Information:					
7	Foreign Reference	WO2016162309A1.pdf	4702140 3452577d6159cec179ea250027aea6d5647c031c	no	70
Warnings:					
Information:					
8	Foreign Reference	WO2018091676A1.pdf	7871680 64cfb73e567a9bc0b0b1031eacd10cf50ecc426d	no	103
Warnings:					
Information:					
9	Non Patent Literature	Armani.pdf	13185876 a0d9565828b4a678920e04bcc4952af40b3000f2	no	19
Warnings:					
Information:					
10	Non Patent Literature	Bailey.pdf	10590623 c8b70b180295bd4fd597fd79b6e1987a73a9228e	no	9
Warnings:					
Information:					
11	Non Patent Literature	Cesano_Abstract1371.pdf	2321994 0ceb642d63848ad6c9d5085f2bda125a08f81b66	no	4
Warnings:					
Information:					
12	Non Patent Literature	Lee.pdf	7059648 91b256eaaaf00789fc5f2513a138ff05e349f0ee	no	5
Warnings:					
Information:					
13	Non Patent Literature	LeMaire.pdf	14229059 25cd274b74cf580f5b7498bb7acfa00c5eec6b47	no	22
Warnings:					
Information:					

14	Fee Worksheet (SB06)	fee-info.pdf	38587 d8f63818c6b25202691471ec37d447833f60c8ce	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	93066079	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	F. W. Lu

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

Via EFS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT
UNDER 37 C.F.R. §§ 1.56, 1.97, AND 1.98

In accordance with the duty of disclosure set forth in 37 C.F.R. §1.56, Applicant hereby submits the following information in conformance with 37 C.F.R. §§1.97 and 1.98. It is respectfully requested that the information be expressly considered during the prosecution of this application, and the references be made of record therein and appear among the "References Cited" on any patent to issue therefrom.

[X] Pursuant to 37 C.F.R. §1.98, a copy of each non-US patent document cited on the attached Form used in lieu of PTO/SB/08 is enclosed.

[] No copies of the foreign patent, foreign patent application, or non-patent literature publications listed on the attached Form used in lieu of PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98(d) because the publications were previously cited by or submitted to the Office in prior Application Serial No(s). ____ to which the above-identified application claims priority under 35 U.S.C. §120.

[X] No copies of any U.S. patents or U.S. patent application publications listed on the attached Form used in lieu of PTO/SB/08 are being provided pursuant to 37 C.F.R. §1.98.

[] Publication(s) listed on the attached Form used in lieu of PTO/SB/08 were cited in a foreign search or examination report corresponding to application serial no. and mailed on .

[] Enclosed is a copy of a non-English publication(s) _____ Pursuant to §609 of the M.P.E.P., Applicant submits the attached foreign search or examination report, which cites such non-English language publication(s).

[] Enclosed is a copy of a non-English publication(s) _____ English language publication _____ (copy enclosed) claims priority from this non-English publication.

[] Enclosed is an explanation of non-English publication(s) _____ for which an English translation is not available.

[] Enclosed is an English translation of non-English publication(s) _____ cited on the attached Form used in lieu of PTO/SB/08.

[] Enclosed is an English language Abstract of non-English publication(s) _____ cited on the attached Form used in lieu of PTO/SB/08.

[] Enclosed is a copy of pending patent Application No. _____.

[X] In accordance with **37 C.F.R. §1.97(c)**, this Information Disclosure Statement is filed after the period specified in 37 C.F.R. § 1.97(b), but before the mailing of any of the following: a final action under 37 C.F.R. §1.113; a notice of allowance under 37 C.F.R. §1.311; or an action that otherwise closes prosecution in this application.

In accordance with 37 C.F.R. §1.97(c) also enclosed is:

[X] Fee under 37 C.F.R. §1.17(p) in the amount of $260.00;

[] Fee under 37 C.F.R. §1.17(p) in the amount of $130.00;

[] Fee under 37 C.F.R. §1.17(p) in the amount of $65.00; or

[] Statement as specified in 37 C.F.R. §1.97(e):

[] Each item of information contained in the Information Disclosure Statement cited herein was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing date of the Information Disclosure Statement; or

[] No item of information contained in the Information Disclosure Statement submitted herewith was cited in any communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the undersigned, having made a reasonable inquiry, no item of information contained in the Information Disclosure Statement was known to any

individual designated in 37 C.F.R. §1.56(c) more than three months prior to the filing date of the Information Disclosure Statement.

In accordance with 37 C.F.R. § 1.97(g), this Information Disclosure Statement shall not be construed as to mean that a search has been made.

In accordance with 37 C.F.R. § 1.97(h), the filing of this Information Disclosure Statement shall not be construed to be an admission that the information cited in the statement is, or is considered to be material to patentability as defined by 37 C.F.R § 1.56(b).

Remarks

It is respectfully requested that the Examiner consider the above-noted information and return an initialed copy of the attached Form used in lieu of PTO/SB/08 to the undersigned.

The Examiner is invited to contact the undersigned by telephone if it is felt that a telephone interview would advance the prosecution of the present application.

The Director is hereby authorized to charge any additional fees which may be required with respect to this communication, or credit any overpayment, to Deposit Account No. 50-1283, under Order No. NATE-037/001US 321329-2523.

Dated: July 28, 2021

Respectfully submitted,
COOLEY LLP

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (617) 937-2400

By: /Robert E. Powers/
Robert E. Powers, Ph.D.
Reg. No. 77,802
Matthew Pavao, Ph.D.
Reg. No. 50,572

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/272,487	02/11/2019	Joseph M. Beechem	NATE-037001US 321329-2523	6864

141429 7590 09/28/2021

Cooley LLP/NanoString Technologies, Inc
1299 Pennsylvania Avenue NW
Suite 700
Washington, DC 20004

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
09/28/2021	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zIPPatentDocketingMailboxUS@cooley.com

Office Action Summary	Application No. 16/272,487	Applicant(s) Beechem et al.	
	Examiner FRANK W LU	Art Unit 1634	AIA (FITF) Status Yes

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTHS FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☑ Responsive to communication(s) filed on 7/26/2021.

☐ A declaration(s)/affidavit(s) under **37 CFR 1.130(b)** was/were filed on _____.

2a)☑ This action is **FINAL.** 2b) ☐ This action is non-final.

3)☐ An election was made by the applicant in response to a restriction requirement set forth during the interview on _____; the restriction requirement and election have been incorporated into this action.

4)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims*

5) ☑ Claim(s) 15-34 is/are pending in the application.

5a) Of the above claim(s) 20 and 24 is/are withdrawn from consideration.

6) ☐ Claim(s) _____ is/are allowed.

7) ☑ Claim(s) 15-19,21-23 and 25-34 is/are rejected.

8) ☐ Claim(s) _____ is/are objected to.

9) ☐ Claim(s) _____ are subject to restriction and/or election requirement

* If any claims have been determined allowable, you may be eligible to benefit from the **Patent Prosecution Highway** program at a participating intellectual property office for the corresponding application. For more information, please see http://www.uspto.gov/patents/init_events/pph/index.jsp or send an inquiry to **PPHfeedback@uspto.gov.**

Application Papers

10)☐ The specification is objected to by the Examiner.

11)☑ The drawing(s) filed on 7/26/2021 is/are: a)☑ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

Certified copies:

a)☐ All b)☐ Some** c)☐ None of the:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

** See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☑ Notice of References Cited (PTO-892)

2) ☑ Information Disclosure Statement(s) (PTO/SB/08a and/or PTO/SB/08b) Paper No(s)/Mail Date 7/28/2021.

3) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date _____.

4) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. Applicant's response to the office action filed on July 26, 2021 has been entered. The claims pending in this application are claims 15-34 wherein claims 20 and 24 have been withdrawn due to the restriction requirement mailed on September 9, 2020. Claims 15-19, 21-23 and 25-34 will be examined.

Drawings

2. New Figures 21, 22, 25, 30, and 31 and the substitute specification submitted on July 26, 2021 have been accepted by the office.

Claim Objections

3. Claim 15 is objected to because of the following informalities: (1) "the identifier oligonucleotides" in step (2) should be "the identifier oligonucleotide from each of the probes"; (2) "the released identifier oligonucleotides" in step (3) should be "the identifier oligonucleotide released from each of the probes"; and (3) "the collected identifier oligonucleotides" in step (3) should be "the collected identifier oligonucleotide released from each of the probes".

4. Claim 30 is objected to because of the following informality: "providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotides" should be "said providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotides".

Appropriate correction is required.

Claim Rejections - 35 USC § 112

5. The following is a quotation of the first paragraph of 35 U.S.C. 112(a):

> (a) IN GENERAL.—The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor or joint inventor of carrying out the invention.

The following is a quotation of the first paragraph of pre-AIA 35 U.S.C. 112:

> The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

6. Scope of Enablement

This rejection is made based on the amendment and is different from the enablement rejection in the office action mailed on January 25, 2021.

Claims 15-19, 21-23 and 25-34 are rejected under 35 U.S.C. 112(a) or 35 U.S.C. 112 (pre-AIA), first paragraph, because the specification, while being enabling for performing steps (1) to (4) of claim 15, does not reasonably provide enablement for spatially detecting the abundance of the at least one target analyte in the at least one cell in a tissue sample. The specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to use the invention commensurate in scope with these claims.

Factors to be considered in determining whether a disclosure meets the enablement requirement of 35 USC 112, first paragraph, have been described by the court in *In re Wands*, 8 USPQ2d 1400 (CA FC 1988). *Wands* states at page 1404,

"Factors to be considered in determining whether a disclosure would require undue experimentation have been summarized by the board in Ex parte Forman. They include (1) the quantity of experimentation necessary, (2) the amount of direction or guidance presented, (3) the presence or absence of working examples, (4) the nature of the invention, (5) the state of the

prior art, (6) the relative skill of those in the art, (7) the predictability or unpredictability of the art, and (8) the breadth of the claims."

The Nature of the Invention

The claims are drawn to a method for spatially detecting the abundance of at least one target analyte in at least one cell from a tissue sample. The invention is a class of invention which the CAFC has characterized as "the unpredictable arts such as chemistry and biology." *Mycogen Plant Sci., Inc. v. Monsanto Co.,* 243 F.3d 1316, 1330 (Fed. Cir. 2001).

The Breadth of The Claims

Claims 15-19, 21-23 and 25-34 encompass a method for spatially detecting the abundance of at least one target analyte in at least one cell from a tissue sample comprising: (1) contacting at least one target analyte in at least one cell in a tissue sample with at least three probes, thereby tiling the probes on the at least one target analyte, wherein each of the probes comprise a target binding domain that binds to the at least one target analyte and an identifier oligonucleotide, wherein the identifier oligonucleotide comprises: a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site, wherein the tissue sample has been treated to facilitate binding of the probes to the at least one target analyte; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotides; (3) collecting the released identifier oligonucleotides; (4) amplifying the collected identifier oligonucleotides; (5) identifying the released identifier oligonucleotides by sequencing the amplified products produced in step (4), thereby spatially detecting the abundance of the at least one target analyte in the at least one cell in a tissue sample.

Working Examples

The specification provides working examples (see pages 28-41 of US 2019/0249248 A1, which is US publication of this instant case) for: (1) Two-Ended Adapter Ligation Method for 96 Multiplexed Samples; (2) One-Ended Adapter Ligation Method for 96 Multiplexed Samples; (3) Templated-Primer Extension Method for 96 Multiplexed Samples; (4) Long Probe Hybridization Method for 96 Multiplexed Samples; (5) Short Probe Hybridization Method for 96 Multiplexed Samples; (6) Direct PCR Method for 96 Multiplexed Samples; (7) Spatially Detecting Target Analytes in a FFPE Sample; (8) Spatially Detecting Target Proteins in a Fluorescently Stained FFPE Sample; (9) Spatially Detecting Target RNAs in a FFPE Sample; (10) Spatially Detecting Target RNAs in Specific Sub-Regions of an ROI; and (11) 96-Plex Human Immuno-Oncology Panel. The specification provides no working example for spatially detecting the abundance of at least one target analyte in at least one cell from a tissue sample using the methods recited in claims 15-19, 21-23 and 25-34.

The Amount of Direction or Guidance Provided and The State of The Prior Art

Although the specification provides working examples (see pages 28-41 of US 2019/0249248 A1, which is US publication of this instant case) for: (1) Two-Ended Adapter Ligation Method for 96 Multiplexed Samples; (2) One-Ended Adapter Ligation Method for 96 Multiplexed Samples; (3) Templated-Primer Extension Method for 96 Multiplexed Samples; (4) Long Probe Hybridization Method for 96 Multiplexed Samples; (5) Short Probe Hybridization Method for 96 Multiplexed Samples; (6) Direct PCR Method for 96 Multiplexed Samples; (7) Spatially Detecting Target Analytes in a FFPE Sample; (8) Spatially Detecting Target Proteins in a Fluorescently Stained FFPE Sample; (9) Spatially Detecting Target RNAs in a FFPE Sample;

(10) Spatially Detecting Target RNAs in Specific Sub-Regions of an ROI; and (11) 96-Plex Human Immuno-Oncology Panel, the specification does not provide a guidance to show that the abundance of at least one target analyte in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23 and 25-34. Furthermore, there is no experimental condition and/or experimental data in the specification to support the claimed invention. During the process of the prior art search, the examiner has not found any prior art which is related to spatially detecting the abundance of at least one target analyte in at least one cell from a tissue sample using the methods recited in claims 15-19, 21-23 and 25-34.

Level of Skill in The Art, The Unpredictability of The Art, and The Quantity of Experimentation Necessary

While the relative skill in the art is very high (the Ph.D. degree with laboratory experience), there is no predictability whether the abundance of at least one target analyte in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23 and 25-34.

Although the specification teaches that "the term 'spatially detecting' is used in its broadest sense to refer to the identification of the presence of a specific target analyte within a specific region of interest in a sample. Spatially detecting can comprise quantifying the amount of a specific target analyte present within a specific region of interest in a sample. Spatially detecting can further comprise quantifying the relative amount of a first target analyte within a specific region of interest in a sample as compared to the amount of at least a second target analyte within a specific region of interest in a sample. Spatially detecting can also comprise quantifying the relative amount of a specific target analyte within a first region of interest in a

sample compared to the amount of the same target analyte in at least a second region of interest in the same sample or different sample" (see paragraph [0308] of US 2019/0249248 A1, which is US publication of this instant case), the definition "spatially detecting" in the specification is only limited to the identification of the presence of a specific target analyte within a specific region of interest in a sample. Since the word "abundance" means "an ample quantity: an abundant amount" (see the definition of "abundance" from a dictionary), claim 15 does not require to quantify the amount of the released identifier oligonucleotide from each of the probes and it is known that determining the abundance of a particular mRNA in a total or poly(A) RNA sample requires to determine the amount of the particular mRNA in the total or poly(A) RNA sample (see "Strategies for Detecting mRNA Northern blotting, Nuclease Protection Assays, In Situ hybridization and RT-PCR" from ThermoFisher Scientific), it is unpredictable how the abundance of at least one target analyte in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23 and 25-34 without knowing the amount of the at least one target analyte. Furthermore, the phrase "at least one target analyte" can be interpreted as two or more different target analytes, since claim 15 does not indicate how to differentiate two or more different target analytes, it is unpredictable how the abundance of at least one target analyte such as two or more different target analytes in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23 and 25-34.

Case law has established that "(t)o be enabling, the specification of a patent must teach those skilled in the art how to make and use the full scope of the claimed invention without 'undue experimentation'." *In re Wright* 990 F.2d 1557, 1561. *In re Fisher*, 427 F.2d 833, 839, 166 USPQ 18, 24 (CCPA 1970) it was determined that "[T]he scope of the claims must bear a

reasonable correlation to the scope of enablement provided by the specification to persons of ordinary skill in the art". The amount of guidance needed to enable the invention is related to the amount of knowledge in the art as well as the predictability in the art. Furthermore, the Court in *Genentech Inc. v Novo Nordisk* 42 USPQ2d 1001 held that "[I]t is the specification, not the knowledge of one skilled in the art that must supply the novel aspects of the invention in order to constitute adequate enablement".

In view of above discussion, the skilled artisan will have no way to predict the experimental results. Accordingly, it is concluded that undue experimentation is required to make the invention as it is claimed. These undue experimentation at least includes to test whether the abundance of at least one target analyte in at least one cell from a tissue sample can be spatially detected using the methods recited in claims 15-19, 21-23 and 25-34.

Conclusion

In the instant case, as discussed above, the level of unpredictability in the art is high, the specification provides one with no guidance that leads one to claimed methods. One of skill in the art cannot readily anticipate the effect of a change within the subject matter to which the claimed invention pertains. Thus given the broad claims in an art whose nature is identified as unpredictable, the unpredictability of that art, the large quantity of research required to define these unpredictable variables, the lack of guidance provided in the specification, the absence of any working example related to claimed invention and the no teaching in the prior art balanced only against the high skill level in the art, it is the position of the examiner that it would require undue experimentation for one of skill in the art to perform the method of the claim as broadly written.

Claim Rejections - 35 USC § 103

7. The following is a quotation of 35 U.S.C. 103 which forms the basis for all obviousness rejections set forth in this Office action:

> A patent for a claimed invention may not be obtained, notwithstanding that the claimed invention is not identically disclosed as set forth in section 102, if the differences between the claimed invention and the prior art are such that the claimed invention as a whole would have been obvious before the effective filing date of the claimed invention to a person having ordinary skill in the art to which the claimed invention pertains. Patentability shall not be negated by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims the examiner presumes that the subject matter of the various claims was commonly owned as of the effective filing date of the claimed invention(s) absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and effective filing dates of each claim that was not commonly owned as of the effective filing date of the later invention in order for the examiner to consider the applicability of 35 U.S.C. 102(b)(2)(C) for any potential 35 U.S.C. 102(a)(2) prior art against the later invention.

8. Claims 15-18, 22, and 26-34 are rejected under 35 U.S.C. 103 as being unpatentable over Beechem *et al.,* (US 2020/0040385 A1, priority date: July 17, 2015) in view of McCasky Feazel *et al.,* (US Patent No. 6,100,030, published on August 8, 2000).

Regarding claims 15, 22, 27-31, 33, and 34, since claim 15 does not require that a unique nucleic acid sequence which identifies the target analyte bound to the target binding domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site have specific structures or specific sequences, Beechem *et al.,* teach (1) contacting at least one target analyte in at least one cell in a tissue sample with at least three probes comprising a target binding domain and an identifier oligonucleotide (ie., the signal oligonucleotide), thereby tiling the probes on the at least one target analyte (eg., see Figure 1), wherein the identifier oligonucleotide comprises: a unique

nucleic acid sequence which identifies the target analyte bound to the target binding domain, a nucleic acid sequence comprising a unique molecular identifier, a first amplification primer binding site, and a second amplification primer binding site, wherein the tissue sample has been treated to facilitate binding of the probes to the at least one target analyte; (2) providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotides; and (3) collecting the released identifier oligonucleotides and identifying the released identifier oligonucleotides, thereby spatially detecting the abundance of at least one target analyte in at least one cell from a tissue sample as recited in claim 15, wherein the unique nucleic acid sequence which identifies the target analyte bound to the target binding domain comprises between about 5 nucleotides and about 25 nucleotides as recited in claim 22, the nucleic acid sequence comprising a unique molecular identifier comprises about 5 to about 20 nucleotides as recited in claim 27, the target binding domain comprises between about 10 nucleotides to about 70 nucleotides as recited in claim 28, each of the probes comprise a photo-cleavable linker between the identifier oligonucleotide and the target binding domain as recited in claim 29, and said providing a force to a location of the tissue sample sufficient to release the identifier oligonucleotides comprises exciting a location of the tissue sample with light of sufficient wavelength capable of cleaving the photo-cleavable linker, thereby releasing the identifier oligonucleotides as recited in claim 30, the unique nucleic acid sequence which identifies the target analyte bound to the target binding domain comprises about 12 nucleotides as recited in claim 31, the nucleic acid sequence comprising a unique molecular identifier comprises about 14 nucleotides as recited in claim 33, and the target binding domain comprises between about 35 to about 50 nucleotides as recited in claim 34 (see paragraphs [0003] to [0008], [0023], [0084], [0086], [0095], [0096], [0100] and [0106] to [0108], Figure 1 and claims 1-53).

Beechem *et al.,* do not disclose (4) amplifying the collected identifier oligonucleotides and (5) identifying the released identifier oligonucleotides by sequencing the amplified products produced in step (4) as recited in claim 15 wherein said amplifying the collected identifier oligonucleotides comprises performing PCR as recited in claim 16, said performing PCR comprises a first amplification primer and second amplification primer, wherein the first amplification primer comprises: a first flow cell binding site; a first nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotides are released; and a nucleic acid sequence complementary to the second amplification primer binding site; and wherein the second amplification primer comprises: a second flow cell binding site; a second nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotides are released; and a nucleic acid sequence complementary to the first amplification primer binding site as recited in claim 17, at least one of the flow cell binding sites comprises a flow cell adapter sequence suitable for sequencing as recited in claim 18, the at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotides are released comprises about 1 to about 20 nucleotides as recited in claim 26, and at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotides are released comprises about 8 nucleotides as recited in claim 32. However, Beechem *et al.,* teach that "[D]etecting comprises a polymerase reaction, a reverse transcriptase reaction, hybridization to an oligonucleotide microarray, mass spectrometry, hybridization to a fluorescent molecular beacon, a sequencing reaction, or nCounter® Molecular Barcodes" (see paragraph [0012]).

Regarding claims 15-18, 26, and 32, since claim 17 does not require that a first flow cell binding site, a first nucleic acid sequence which identifies the specific location of the tissue

sample from which the identifier oligonucleotides are released, a second flow cell binding site, and a second nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotides are released have specific structures or specific sequences and claim 18 does not require that a flow cell adapter has a specific structure or a specific sequence and McCasky Feazel *et al.,* teach a AFLP technique using one primer pair (see Figure 1A) and each primer in the one primer pair can be any length (see column 6), McCasky Feazel *et al.,* disclose amplifying an identifier oligonucleotides and identifying the identifier oligonucleotides by sequencing an amplified products as recited in claim 15, said performing PCR comprises a first amplification primer and second amplification primer, wherein the first amplification primer comprises: a first flow cell binding site; a first nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotides are released; and a nucleic acid sequence complementary to the second amplification primer binding site; and wherein the second amplification primer comprises: a second flow cell binding site; a second nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotides are released; and a nucleic acid sequence complementary to the first amplification primer binding site as recited in claim 17, at least one of the flow cell binding sites comprises a flow cell adapter sequence suitable for sequencing as recited in claim 18, the at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotides are released comprises about 1 to about 20 nucleotides as recited in claim 26, and at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotides are released comprises about 8 nucleotides as recited in claim 32 (see Figure 1, columns 5, 11, and 12).

Therefore, it would have been *prima facie* obvious to one having ordinary skill in the art at the time the invention was made to have performed the methods recited in claims 15-18, 26, and 32 by amplifying the collected identifier oligonucleotides and identifying the released identifier oligonucleotides by sequencing the amplified products produced in step (4) of claim 15 wherein said amplifying the collected identifier oligonucleotides comprises performing PCR, said performing PCR comprises a first amplification primer and second amplification primer, wherein the first amplification primer comprises: a first flow cell binding site; a first nucleic acid sequence which identifies the specific location of the tissue sample from which the identifier oligonucleotides are released; and a nucleic acid sequence complementary to the second amplification primer binding site; and wherein the second amplification primer comprises: a second flow cell binding site; a second nucleic acid sequence which identifies the specific location of the sample from which the identifier oligonucleotides are released; and a nucleic acid sequence complementary to the first amplification primer binding site, at least one of the flow cell binding sites comprises a flow cell adapter sequence suitable for sequencing, the at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotides are released comprises about 1 to about 20 nucleotides, and at least one of the nucleic acid sequences which identify the specific location of the tissue sample from which the identifier oligonucleotides are released comprises about 8 nucleotides in view of the prior arts of Beechem *et al.,* and McCasky Feazel *et al.*. One having ordinary skill in the art would have been motivated to do so because Beechem *et al.,* have shown that "[D]etecting comprises a polymerase reaction, a reverse transcriptase reaction, hybridization to an oligonucleotide microarray, mass spectrometry, hybridization to a fluorescent molecular beacon, a sequencing reaction, or nCounter® Molecular Barcodes" (see paragraph [0012]) while, since

McCasky Feazel *et al.,* disclose amplifying an identifier oligonucleotides and identifying the identifier oligonucleotides by sequencing amplified products as recited in claim 15 and claims 16-18 and 26 (see above and Figure 1, columns 5, 11, and 12) and the simple substitution of one kind of detection method (ie., the detection method taught by Beechem *et al.,*) from another kind of detection method (ie., the detection method taught by McCasky Feazel *et al.,* including PCR and sequencing) during the process of detecting at least one target analyte in at least one cell in a tissue sample as recited in claim 15, in the absence of convincing evidence to the contrary, would have been *prima facie* obvious to one having ordinary skill in the art at the time the invention was made since the detection method taught by Beechem *et al.,* and the detection method taught by McCasky Feazel *et al.,* including PCR and sequencing are used for the same purpose (ie., detecting at least one target analyte in at least one cell in a tissue sample). One having ordinary skill in the art at the time the invention was made would have a reasonable expectation of success to detect at least one target analyte in at least one cell in a tissue sample using the detection method taught by McCasky Feazel *et al.,* in view of the prior arts of Beechem *et al.,* and McCasky Feazel *et al.*.

Furthermore, the motivation to make the substitution cited above arises from the expectation that the prior art elements will perform their expected functions to achieve their expected results when combined for their common known purpose. Support for making the obviousness rejection comes from the M.P.E.P. at 2144.06, 2144.07 and 2144.09.

Also note that there is no invention involved in combining old elements is such a manner that these elements perform in combination the same function as set forth in the prior art without giving unobvious or unexpected results. *In re Rose* 220 F.2d. 459, 105 USPQ 237 (CCPA 1955).

Response to Arguments

In page 12, third paragraph bridging to page 14, third paragraph of applicant's remarks, applicant argues that "[A]pplicant submits that this specific combination of features is not taught or suggested by Beechem and McCasky, either alone or in combination. In addition to not being taught or suggested by the cited art, Applicant submits that the tiling of multiple probes onto a target analyte and the use of identifier oligonucleotides comprising a unique molecular identifier in the instantly claimed method unexpectedly results in superior spatial detection of the abundance of target analytes in a tissue sample. First, as explained in the instant specification at paragraph [00208], the term tiling is used to describe when more than one probe of the present disclosure is bound to a target analyte (e.g. as shown in Figure 27 of the instant specification). The specification further explains that by tiling multiple probes onto a target analyte, the target analyte will be individually detected multiple times, increasing the overall accuracy of the measurement. Finally, in the case where multiple probes are tiled onto a single target RNA, the specification explains that there may be instances where one of the probes is incorrectly detected too many times and/or another probe is detected too few times. However, because multiple probes have been tiled onto the target RNA, these outlier probes can be discarded during analysis if the signals from the other probes are detected at a similar level, thereby further increasing the accuracy of the measurement of RNA abundance. Second, as explained in the instant specification at paragraph [00319], unique molecular identifiers are short nucleic acid sequences that are used to quantify and reduce quantitative bias caused by nucleic acid amplification prior to sequencing reactions.The instantly claimed method uses the combination of tiling and unique molecular identifies to more accurately detect the spatial abundance of specific target analytes in a tissue sample, as is shown in Example 11 of the instant specification. In this example, the

instantly claimed method was used to detect the spatial abundance of 96 different target analytes in FFPE samples in 22 common human cell lines. As described in Example 11, for most of the 96 different target analytes, there was more than one probe that hybridized to the target analyte, thereby creating a tiling effect on many of the target analytes. As shown in Figures 31, 33 and 34, the use of tiling allowed for robust detection of different target analytes. Moreover, as shown in Figures 31, 33 and 34, the tiling approach allowed for the easy and rapid identification of outlier probes that can be discarded during analysis to increase the accuracy of the abundance measurement. Finally, because of the robust and accurate detection, Figures 31, 33 and 34 demonstrate that the spatial abundance of certain targets can be compared across different cell types. The ability to more accurately detect spatial abundance in tissue samples to the extent that different samples and/or different areas on a single sample can be compared is important in clinical applications directed towards ascertaining the boundaries of cancerous and non-cancerous portions of patient biopsies. Applicant submits that claim 15, as amended herein, recites a unique combination of features not disclosed by the cited art, and that these features produce a method that yields unexpectedly superior results. Accordingly, claim 15, from which all other claims subject to the rejection properly depend, is nonobvious in view of Beechem and McCasky".

The arguments have been fully considered but they are not persuasive toward the withdrawal of the rejection.

First, Beechem *et al.,* teach contacting at least one target analyte in at least one cell in a tissue sample with at least three probes, thereby tiling the probes on the at least one target analyte (see paragraphs [0023], [0084], [0086], [0095], and [0106] to [0108], and Figure 1).

Second, although applicant argues that "as explained in the instant specification at paragraph [00319], unique molecular identifiers are short nucleic acid sequences that are used to quantify and reduce quantitative bias caused by nucleic acid amplification prior to sequencing reactions', since claim 15 does not require that a unique molecular identifier has a specific structure or a specific sequence and the word "short" is relative, each of the probes taught by Beechem *et al.,* can be considered to have an unique molecular identifier that is a short nucleic acid sequence and is used to quantify and reduce quantitative bias caused by nucleic acid amplification prior to sequencing reactions.

Third, since besides amplifying an identifier oligonucleotide and identifying the identifier oligonucleotide by sequencing an amplified products in claim 15, Beechem *et al.,* teach all limitations recited in claim 15 (see above rejection under 35 U.S.C. 103). The method recited in claim 15 is expected and applicant's argument "claim 15, as amended herein, recites a unique combination of features not disclosed by the cited art, and that these features produce a method that yields unexpectedly superior results" is incorrect.

9. Claims 21 and 23 are rejected under 35 U.S.C. 103 as being unpatentable over Beechem *et al.,* in view of McCasky Feazel *et al.,* as applied to claims 15-18, 22, and 26-34 above, and further in view of Van Eijk *et al.,* (US 2003/0175729 A1, published on September 18, 2003).

The teachings of Beechem *et al.,* and McCasky Feazel *et al.,* have been summarized previously, *supra*.

Beechem *et al.,* and McCasky Feazel *et al.,* do not disclose that at least one of the flow cell binding sites comprises about 15 to about 40 nucleotides as recited in claim 21 and at least

one of the amplification primer binding sites comprises about 10 to about 40 nucleotides as recited in claim 23.

Van Eijk *et al.,* teach that conventional AFLP-adapters and -primers are about 10-30 base pairs in length (see paragraph [0161]).

Therefore, it would have been *prima facie* obvious to one having ordinary skill in the art at the time the invention was made to have performed the methods recited in claims 21 and 23 using an EcoRI primer and a Mse I primer which have about 30 base pairs in length such that at least one of the flow cell binding sites comprises about 15 nucleotides and at least one of the amplification primer binding sites comprises about 10 nucleotides in view of the prior arts of Beechem *et al.,* McCasky Feazel *et al.,* and Van Eijk *et al.*. One having ordinary skill in the art would have been motivated to do so because Van Eijk *et al.,* have suggested that conventional AFLP-adapters and -primers are about 10-30 base pairs in length (see paragraph [0161]). One having ordinary skill in the art at the time the invention was made would have a reasonable expectation of success to perform the methods recited in claims 21 and 23 using an EcoRI primer and a Mse I primer which have about 30 base pairs in length such that at least one of the flow cell binding sites comprises about 15 nucleotides and at least one of the amplification primer binding sites comprises about 10 nucleotides in view of the prior arts of Beechem *et al.,* McCasky Feazel *et al.,* and Van Eijk *et al.*.

Response to Arguments

In page 14, four and fifth paragraphs of applicant's remarks, applicant argues that "[C]laim 15, from which claims 21 and 23 properly depend, is nonobvious in view of Beechem and McCasky for at least the reasons described supra. Van Eijk fails to cure the deficiencies of

Beechem and McCasky. Accordingly, Applicant submits that the instant claims are nonobvious in view of Beechem, McCasky and Van Eijk, either alone or in combination".

The arguments have been fully considered but they are not persuasive toward the withdrawal of the rejection because Beechem *et al.,* in view of McCasky Feazel *et al.,* teach all limitations recited in claims 15-17 (see above Response to Arguments related to Rejection Item No. 8).

Conclusion

10. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

11. No claim is allowed.

12. Papers related to this application may be submitted to Group 1600 by facsimile transmission. Papers should be faxed to Group 1600 via the PTO Fax Center. The faxing of such papers must conform with the notices published in the Official Gazette, 1096 OG 30

(November 15, 1988), 1156 OG 61 (November 16, 1993), and 1157 OG 94 (December 28, 1993)(See 37 CAR § 1.6(d)). The CM Fax Center number is (571)273-8300.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Frank Lu, Ph.D., whose telephone number is (571)272-0746. The examiner can normally be reached on Monday-Friday from 9 A.M. to 5 P.M.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dave Nguyen, can be reached on (571)272-0731.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/FRANK W LU/
Primary Examiner, Art Unit 1634
September 22, 2021

Notice of References Cited	Application/Control No. 16/272,487	Applicant(s)/Patent Under Reexamination Beechem et al.	
	Examiner FRANK W LU	Art Unit 1634	Page 1 of 1

U.S. PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Name	CPC Classification	US Classification
	A					
	B					
	C					
	D					
	E					
	F					
	G					
	H					
	I					
	J					
	K					
	L					
	M					

FOREIGN PATENT DOCUMENTS

*		Document Number Country Code-Number-Kind Code	Date MM-YYYY	Country	Name	CPC Classification
	N					
	O					
	P					
	Q					
	R					
	S					
	T					

NON-PATENT DOCUMENTS

*		Include as applicable: Author, Title Date, Publisher, Edition or Volume, Pertinent Pages)
	U	The definition of “abundance” from a dictionary. Printed on 9/22/2021
	V	Strategies for Detecting mRNA Northern blotting, Nuclease Protection Assays, In Situ hybridization and RT-PCR” from ThermoFisher Scientific. Printed on 9/22/2021.
	W	
	X	

*A copy of this reference is not being furnished with this Office action. (See MPEP § 707.05(a).)
Dates in MM-YYYY format are publication dates. Classifications may be US or foreign.

Search Notes	Application/Control No.	Applicant(s)/Patent Under Reexamination
	16/272,487	Beechem et al.
	Examiner	Art Unit
	FRANK W LU	1634

CPC - Searched*		
Symbol	Date	Examiner
Update	09/22/2021	/FL/

CPC Combination Sets - Searched*		
Symbol	Date	Examiner
Update	09/22/2021	/FL/

US Classification - Searched*			
Class	Subclass	Date	Examiner
Update		09/22/2021	/FL/

* See search history printout included with this form or the SEARCH NOTES box below to determine the scope of the search.

Search Notes		
Search Notes	Date	Examiner
Update (see attached files)	09/22/2021	/FL/

Interference Search			
US Class/CPC Symbol	US Subclass/CPC Group	Date	Examiner

/FRANK W LU/ Primary Examiner, Art Unit 1634	September 22, 2021

EAST Search History

EAST Search History (Prior Art)

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	1	(abundance or amount)same probe same force same (release or releasing or cleave or cleaving or cleavage) same (amplify or amplifying or amplification) same (identify or identifying or identification) same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/09/22 23:39
L2	172	(abundance or amount)same probe same (release or releasing or cleave or cleaving or cleavage) same (amplify or amplifying or amplification) same (identify or identifying or identification) same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/09/22 23:42
L3	163	(abundance or amount)same (analyte or "nucleic acid" or DNA or RNA or polynucleotide) same probe same (release or releasing or cleave or cleaving or cleavage) same (amplify or amplifying or amplification) same (identify or identifying or identification) same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/09/22 23:43
L4	69	(abundance or amount)same (analyte or "nucleic acid" or DNA or RNA or polynucleotide) same probe same (release or releasing or cleave or cleaving or cleavage) same (amplify or amplifying or amplification) same (identify or identifying or identification) same sequencing same (detect or detecting or detection)	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/09/22 23:44
L5	66	4 AND ((C12Q2600/156 OR C12Q1/6883 OR C12Q2600/158 OR C12Q2565/107 OR C12Q1/6886 OR C12Q2600/16 OR C12Q1/6816 OR C12Q1/6881 OR C12Q2600/118 OR C12Q1/6869 OR C12Q2565/514 OR C12Q1/6827 OR C12Q2535/122 OR C12Q2600/106 OR C12Q2563/179 OR C12Q2600/172 OR C12Q2525/185 OR C12Q2537/143 OR C12Q2563/107 OR C12Q1/6818 OR C12Q1/6851 OR C12Q2525/186 OR C12Q2525/301 OR C12Q1/6804 OR C12Q1/6806 OR C12Q1/6855 OR C12Q1/686 OR C12Q2521/301 OR C12Q2600/136 OR C12Q1/6844 OR C12Q2533/107 OR C12Q1/6809 OR	US-PGPUB; USPAT; USOCR; FPRS; EPO; JPO; DERWENT; IBM_TDB	OR	ON	2021/09/22 23:48

		C12Q1/6874 OR C12Q2521/319 OR C12Q2549/101 OR C12Q2600/112 OR C12Q1/6853 OR C12Q2525/121 OR C12Q2525/131 OR C12Q2525/161 OR C12Q2527/101 OR C12Q2533/10 OR C12Q2535/131 OR C12Q2561/101 OR C12Q2563/116 OR C12Q2565/631 OR C12Q1/6837 OR C12Q1/6848 OR C12Q1/689 OR C12Q1/6895 OR C12Q2521/501 OR C12Q2525/191 OR C12Q1/68 OR C12Q2521/125 OR C12Q2523/125 OR C12Q2525/113 OR C12Q2525/179 OR C12Q2525/204 OR C12Q2531/107 OR C12Q2565/101 OR C12Q1/34 OR C12Q1/44 OR C12Q1/6813 OR C12Q1/6858 OR C12Q1/70 OR C12Q2521/101 OR C12Q2521/107 OR C12Q2521/327 OR C12Q2525/00 OR C12Q2525/155 OR C12Q2531/113 OR C12Q2533/101 OR C12Q2537/16 OR C12Q2600/154 OR C12Q2600/178 OR C12N15/1072 OR C12N9/14 OR C12N15/10 OR C12N15/1096 OR C12N15/63 OR C12N15/65 OR C12N9/22 OR C12N15/102 OR C12N15/1027 OR C12N15/1034 OR C12N15/1058 OR C12N15/1065 OR C12N15/1093 OR C12N15/111 OR C12N15/113 OR C12N15/62 OR C12N15/66 OR C12N2310/113 OR C12N2310/3231 OR C12N2310/3233 OR C12N2310/3525 OR C12N2320/30 OR C12N9/1252 OR C12N9/93 OR C12N9/96 OR Y02A90/10 OR Y02A90/26 OR C12P19/34 OR G01N2458/10 OR G01N33/53 OR G01N2333/70596 OR G01N2800/24 OR G01N2800/2871 OR G01N2800/50 OR G01N33/6845 OR G01N33/92 OR G01N2035/00287 OR G01N2333/35 OR G01N2333/96463 OR G01N2458/00 OR G01N2500/02 OR G01N2500/10 OR G01N2800/22 OR G01N33/5308 OR G01N33/6803 OR G01N33/6824 OR G01N33/6842 OR G01N33/6893 OR G01N33/86 OR C07K14/315 OR C07K14/195 OR C07K14/21 OR C07K14/212 OR C07K14/245 OR C07K14/26 OR C07K14/285 OR C07K14/31 OR C07K14/3156 OR C07K14/445 OR C07K14/82 OR C12Y309/01001 OR C12Y301/26004 OR C12Y308/01002 OR C40B40/06 OR C40B70/00 OR C40B40/00 OR C40B40/10				

		OR C40B60/14 OR G16B20/00 OR G16B25/00 OR G16B30/00 OR G16B5/00 OR A61K2039/53 OR A61K38/00 OR A61K39/00 OR A61P27/00 OR A61P27/02 OR A61P27/06 OR A61P33/06 OR A61P35/00 OR A61P35/02 OR B01J19/16 OR B01J2219/00317 OR B01J2219/00367 OR B01J2219/00378 OR B01J2219/00495 OR B01J2219/00603 OR B01J2219/0068).CPC.)				

9/22/2021 11:50:56 PM
C:\Users\flu\Documents\EAST\Workspaces\~auto.02072018.175405.bak.wsp

```
(FILE 'HOME' ENTERED AT 23:49:11 ON 22 SEP 2021)

     FILE 'MEDLINE, CAPLUS, BIOSIS, SCISEARCH, EMBASE' ENTERED AT 23:49:46
ON
     22 SEP 2021
L1       7798420 S ABUNDANCE OR AMOUNT
L2      14389347 S ANALYTE OR NUCLEIC ACID OR DNA OR RNA OR POLYNUCLEOTIDE
L3       2329406 S PROBE OR PROBES
L4       6046113 S RELEASE OR RELEASING OR CLEAVE OR CLEAVING OR CLEAVAGE
L5       1316201 S AMPLIFY OR AMPLIFYING OR AMPLIFICATION
L6      10995788 S IDENTIFY OR IDENTIFYING OR IDENTIFICATION
L7       2100957 S SEQUENCING
L8       9190353 S DETECT OR DETECTING OR DETECTION
L9            16 S L1 AND L2 AND L3 AND L4 AND L5 AND L6 AND L7 AND L8
L10      3123855 S FORCE
L11            0 S L9 AND L10
L12           10 DUP REM L9 (6 DUPLICATES REMOVED)
```

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	NATE-037/001US 321329-2523

U.S. PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Document Number Number-Kind Code[2] (if known)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages or Relevant Figures Appear
	1.	7,255,999	08-14-2007	Singh et al.	
	2.	2003/0148335	08-07-2003	Shen et al.	
	3.	2005/0131006	06-16-2005	Mukherjee et al.	
	4.	2005/0170439	08-04-2005	Chan-Hui et al.	
	5.	2005/0196786	09-08-2005	Levy	
	6.	2006/0063196	03-23-2006	Akeson et al.	
	7.	2008/0118934	05-22-2008	Gerdes et al.	
	8.	2010/0075858	03-25-2010	Davis et al.	
	9.	2011/0172115	07-14-2011	Thompson	

FOREIGN PATENT DOCUMENTS

Examiner Initials*	Cite No.[1]	Foreign Patent Document Country Code[3]-Number[4]-Kind Code[5] (*if known*)	Publication Date MM-DD-YYYY	Name of Patentee or Applicant of Cited Document	Pages, Columns, Lines, Where Relevant Passages Or Relevant Figures Appear	T[6]
	10.	WO 2009/156725 A1	12-30-2009	Trillion Genomics Limited		
	11.	WO 2010/081114 A2	07-15-2010	20/20 Genesystems, Inc.		
	12.	WO 2012/140224 A1	10-18-2012	Frisen, J.; Stahl, P.; Lundeberg, J.		
	13.	WO 2014/060483 A1	04-24-2014	Spatial Transcriptomics AB		
	14.	WO 2016/162309 A1	10-13-2016	Spatial Transcriptomics AB		
	15.	WO 2018/091676 A1	05-24-2018	Spatial Transcriptomics AB		

Examiner Signature:	/FRANK W LU/	Date Considered	09/22/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

254069444 v1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

PTO/SB/08a (07-09)
Approved for use through 11/30/2020. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Complete if Known	
Application Number	16/272,487
Filing Date	02-11-2019
First Named Inventor	Joseph M. BEECHEM
Art Unit	1634
Examiner Name	F. W. Lu
Attorney Docket Number	NATE-037/001US 321329-2523

NON PATENT LITERATURE DOCUMENTS

Examiner's Initials	Cite No.[1]	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title ofthe item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T[2]
	16.	ARMANI, M. et al., "2D-PCR: a method of mapping DNA in tissue sections," Lab Chip., 9(24):3526-3534 (2009); doi:10.1039/b910807f.	
	17.	BAILEY, R. C. et al., "DNA-Encoded Antibody Libraries: A Unified Platform for Multiplexed Cell Sorting and Detection of Genes and Proteins," J. Am. Chem. Soc., 129:1959-1967 (2007).	
	18.	CESANO, A. et al., "Abstract 1371: Spatially-resolved, multiplexed digital characterizatio nof protein distribution and abundance in FFPE tissue sections," AACR 107th Annual Meeting, April 16-20, 2016, 5 pages.	
	19.	LEE, H. et al., "Colour-barcoded magnetic microparticles for multiplexed bioassays," Nature Materials, 9:745-749 (2010).	
	20.	LEMAIRE, R. et al., "Tag-Mass: Specific Molecular Imaging of Transcriptome and Proteome by Mass Spectrometry Based on Photocleavable Tag," J Proteome Res., 6(6):2057-2067 (2007); oi:10.1021/pr0700044.	

Examiner Signature:	/FRANK W LU/	Date Considered	09/22/2021

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

[1] Applicant's unique citation designation number (optional). [2] See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. [3] Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). [4] For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. [5] Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. [6] Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

254069444 v1

ALL REFERENCES CONSIDERED EXCEPT WHERE LINED THROUGH. /F.W.L/

Docket No.: NATE-037/001US 321329-2523
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Inventor:	Joseph M. BEECHEM	Confirmation No.:	6864
Application No.:	16/272,487	Group Art Unit:	1634
Filed:	February 11, 2019	Examiner:	F. W. Lu

For: BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

COMMUNICATION IN LIEU OF RESPONSE

Applicant hereby requests a three-month extension of time to maintain the pendency of the above-identified application with regard to the Non-Final Office Action mailed September 28, 2022. Submitted herewith is a payment of $1,480 for the fee associated with the requested extension of time.

With the requested extension, and the payment of the fees, application remains pending through March 28, 2022. In lieu of responding to the Non-Final Office Action, a continuation application is filed claiming priority to the above-identified application under 35 U.S.C. § 120. The Commissioner is invited to contact the undersigned if additional information is required.

The Commissioner is hereby authorized to charge any additional fee, or credit any overpayment, to Deposit Account No. 50-1283.

Dated: March 28, 2022

Respectfully submitted,
COOLEY LLP

USPTO CUSTOMER NO. 141429

By: /Robert E. Powers/
Robert E. Powers
Reg. No. 77,802
Matthew Pavao
Reg. No. 50,572

COOLEY LLP
ATTN: IP Docketing Department
1299 Pennsylvania Avenue NW, Suite 700
Washington, DC 20004

Tel: (617) 937-2300
Fax: (202) 842-7899

266575949 v1

Electronic Patent Application Fee Transmittal

Application Number:	16272487
Filing Date:	11-Feb-2019
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Filer:	Robert E. Powers/Francesca Vandermaelen
Attorney Docket Number:	NATE-037001US 321329-2523

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Extension - 3 months with $0 paid	1253	1	1480	1480
Miscellaneous:				
			Total in USD ($)	**1480**

Electronic Acknowledgement Receipt

EFS ID:	45329481
Application Number:	16272487
International Application Number:	
Confirmation Number:	6864
Title of Invention:	BIOMOLECULAR PROBES AND METHODS OF DETECTING GENE AND PROTEIN EXPRESSION
First Named Inventor/Applicant Name:	Joseph M. Beechem
Customer Number:	141429
Filer:	Robert E. Powers/Francesca Vandermaelen
Filer Authorized By:	Robert E. Powers
Attorney Docket Number:	NATE-037001US 321329-2523
Receipt Date:	28-MAR-2022
Filing Date:	11-FEB-2019
Time Stamp:	12:23:01
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$1480
RAM confirmation Number	E20223RC23223583
Deposit Account	
Authorized User	

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Miscellaneous Incoming Letter	NATE-037_001US_Communication.pdf	95591 e5e0aafdfa698494f316ecb1d2869635e2449101	no	1
Warnings:					
Information:					
2	Fee Worksheet (SB06)	fee-info.pdf	38803 ac40907f9155e1ed18f90814fe1829a691155713	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	134394		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
16/272,487	02/11/2019	Joseph M. Beechem	NATE-037001US 321329-2523	6864

141429 7590 04/14/2022

Cooley LLP/NanoString Technologies, Inc
1299 Pennsylvania Avenue NW
Suite 700
Washington, DC 20004

EXAMINER	
LU, FRANK WEI MIN	

ART UNIT	PAPER NUMBER
1634	

NOTIFICATION DATE	DELIVERY MODE
04/14/2022	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

zIPPatentDocketingMailboxUS@cooley.com

Notice of Abandonment	Application No.	Applicant(s)
	16/272,487	Beechem et al.
	Examiner	Art Unit
	FRANK W LU	1634

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address--

This application is abandoned in view of:

1. ☑ Applicant's failure to timely file a proper reply to the Office letter mailed on 28 September 2021.
   (a) ☐ A reply was received on _____ (with a Certificate of Mailing or Transmission dated _____), which is after the expiration of the period for reply (including a total extension of time of _____ month(s)) which expired on _____.
   (b) ☐ A proposed reply was received on _____, but it does not constitute a proper reply under 37 CFR 1.113 to the final rejection. (A proper reply under 37 CFR 1.113 to a final rejection consists only of:(1) a timely filed amendment which places the application in condition for allowance; (2) a timely filed Notice of Appeal (with appeal fee); or (3) if this is utility or plant application, a timely filed Request for Continued Examination (RCE) in compliance with 37 CFR 1.114. Note that RCEs are not permitted in design applications.)
   (c) ☐ A reply was received on _____ but it does not constitute a proper reply, or a *bona fide* attempt at a proper reply, to the non-final rejection. See 37 CFR 1.85(a) and 1.111. (See explanation in box 8 below).
   (d) ☑ No reply has been received.

2. ☐ Applicant's failure to timely pay the required issue fee and publication fee, if applicable, within the statutory period of three months from the mailing date of the Notice of Allowance (PTOL-85).
   (a) ☐ The issue fee and publication fee, if applicable, was received on _____ (with a Certificate of Mailing or Transmission dated _____), which is after the expiration of the statutory period for payment of the issue fee (and publication fee) set in the Notice of Allowance (PTOL-85).
   (b) ☐ The submitted fee of $ _____ is insufficient. A balance of $ _____ is due.
   The issue fee required by 37CFR 1.18 is $ _____ . The publication fee, if required by 37 CFR 1.18(d), is $ _____ .
   (c) ☐ The issue fee and publication fee, if applicable, has not been received.

3. ☐ Applicant's failure to timely file corrected drawings as required by, and within the three-month period set in, the Notice of Allowability (PTO-37).
   (a) ☐ Proposed corrected drawings were received on _____ (with a Certificate of Mailing or Transmission dated _____), which is after the expiration of the period for reply.
   (b) ☐ No corrected drawings have been received.

4. ☐ The letter of express abandonment which is signed by the attorney or agent of record or other party authorized under 37 CFR 1.33 (b). See 37 CFR 1.138(b).

5. ☐ The letter of express abandonment which is signed by an attorney or agent (acting in a representative capacity under 37 CFR 1.34) upon the filing of a continuing application.

6. ☐ The decision by the Patent Trial and Appeal Board rendered on _____ and because the period for seeking court review of the decision has expired and there are no allowed claims.

7. ☐ The dismissal of the appeal in an application having no allowed claims. (When an appeal is dismissed in an application having no allowed claims the application stands abandoned. See MPEP 1215.04 subsections I-V for an explanation of the reasons why an appeal is dismissed.)

8. ☐ The reason(s) below:

_____

/FRANK W LU/ Primary Examiner, Art Unit 1634	April 8, 2022

Petitions to revive under 37 CFR 1.137, or requests to withdraw the holding of abandonment under 37 CFR 1.181, should be promptly filed to minimize any negative effects on patent term.