(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 Al

(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 201 1 (13.04.201 1) GB

(72) Inventors; and
(71) Applicants : **FRISEN, Jonas** [SE/SE]; Bragevagen 5, SE-114 26 Stockholm (SE). **STAHL, Patrik** [SE/SE]; Fatburstrappan 18, T3: l, SE-1 18 26 Stockholm (SE). **LUNDEBERG, Joakim** [SE/SE]; Torsvikssvan 20, SE-181 34 Lidingo (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).

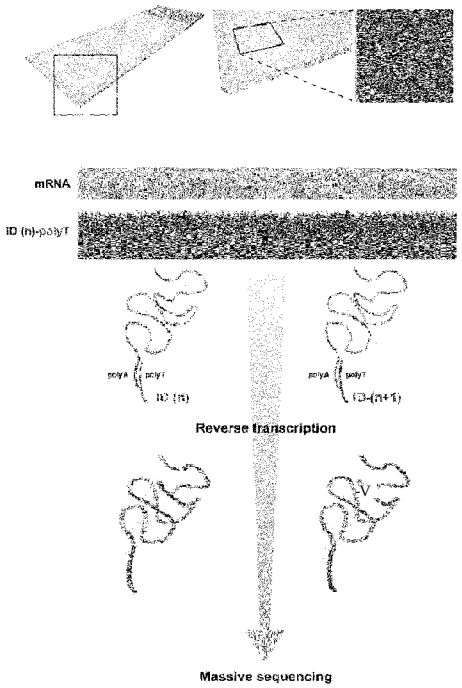
Published:

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

WO 2012/140224 A1

(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.

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: 105-1 1). 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 transcnptome or genome. Thus, the method may involve determining and/or analysing a part or subset of the transcriptome or genome, e.g. a transcnptome 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 μηι 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 1-100 μm, whereas the cross-section of plant cells typically may range from 1-10000 μm. 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 8x1 30k 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 μηι 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 0.1 μηι, 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 10 μm. In other embodiments the thickness of the tissue sample section is at least 10, 12, 13, 14, 15, 20, 30, 40 or 50 μηι. 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 μηι or more. Typically, the thickness of the tissue sample section is between 1-100 μηι, 1-50 μηι, 1-30 μηι, 1-25 μηι, 1-20 μηι, 1-15 μηι, 1-10 μηι, 2-8 μηι, 3-7 μηι or 4-6 μηι, 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. Rsal. 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 U20S 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-I D1	UUAAGTACAAATCTCGACTGCCACTCTGAACCTTCTCCTTCTCCTTCACC-TTTTTTTTTTTTTTTTTTTTTTTTVN (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	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTTTTTTTVN (SEQ ID NO: 34)	Phosphorylated	45
Capture_LP_Poly-d24T	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTTTTTTTT (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)	AGACGTGTGCTCTTCCGATCTNNNNNNNN (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)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 41)	45 (44)

Amplification handle to incorporate A handle into P handle products

A_P handle	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATCTCGACTGCCACTCTGA (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) O.lxSSC 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 lOxAmpligase 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 10μηι. 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 4 μm 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 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.

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 PlatinumOTaq DNA Polymerase from Invitrogen was used when performing the cDNA synthesis. For each reaction 25 μl 2x reaction mix (Superscript™ 111One-step RT-PCR System with PlatinumOTaq DNA Polymerase, Invitrogen), 22.5 μl H_20 and 0.5 μl 100xBSA were mixed and heated to 50°C. Superscript Ill/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 O.lxSSC 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) O.lxSSC 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 IxTris Buffered Saline (50mM Tris, 150mM NaCI, 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 MgCI2 (Roche), 200 µM dNTPs (New England Biolabs) and 0.1 U/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 (covalentlv 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); O.^g^l BSA (New England Biolabs); and 0.1 U/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 (covalentlv 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); O.^g^l 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), 0.1 µg/µl BSA (New England Biolabs); O.IU/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 0.1 U/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), O.^/ µl BSA (New England Biolabs, www.neb.com), 0.1 U/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 (5υ/µl) (Roche) was also added to both tubes, final concentration of 0.1 1U/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) O.lxSSC 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 lOxAmpligase 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) O.lxSSC 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 μηι 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 μηι 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 (20υ/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) O.lxSSC 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) O.lxSSC 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 0.1 U/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), O.^g^l BSA (New England Biolabs, www.neb.com), 0.1 U/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) O.lxSSC 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 0.1 U/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 0.1 U/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) O.lxSSC 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) (lllumina, 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) O.lxSSC 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, Wl, 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 μηι 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 10μηι. 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 (10μηι) 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 (Arraylt, 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 O.lxSSC 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 O.lxSSC at 300 rpm for 1 minute.

Probe release

The 16-well Hybridization Cassette with silicone gasket (Arraylt) 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), 0.1 U/µl USER Enzyme (NEB) and O.^g^l 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 InPELO (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 lllumina 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 μηι 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 exception 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 (ArraylT) was preheated to 37°C and attached to the Codelink slide. A volume of 50μI of cleavage mixture preheated to 37°C, consisting of 1x Endonucelase VIII Buffer (NEB), IOU/μI MutY (Trevigen), IOU/μI Endonucelase VIII (NEB), O.^/ μI 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 InPELO (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-taq 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 unigue 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 unigue 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 U20S 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 U20S 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^g 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 (Arraylt, 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 O.lxSSC 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 0.1 U/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 sees, 65°C for 30 sees, 72°C for 90 sees, 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 InPELO (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)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCCGATATGATTGCCGCTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe2 (SEQ ID NO: 51)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGAGCCGGGTTCATCTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe3 (SEQ ID NO: 52)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTGAGGCACTCTGTTGGGATTTTTTTTTTTTTTTTTTTTTTTVN
Probe4 (SEQ ID NO: 53)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGATTAGTCGCCATTCGTTTTTTTTTTTTTTTTTTTTTTTVN
Probe5 (SEQ ID NO: 54)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTACTTGAGGGTAGATGTTTTTTTTTTTTTTTTTTTTTTTTVN
Probe6 (SEQ ID NO: 55)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATGGCCAATACTGTTATCTTTTTTTTTTTTTTTTTTTTTTTVN
Probe7 (SEQ ID NO: 56)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCTACCCTGATTCGACCTTTTTTTTTTTTTTTTTTTTTTTVN
Probe8 (SEQ ID NO: 57)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCCACTTTCGCCGTAGTTTTTTTTTTTTTTTTTTTTTTTVN
Probe9 (SEQ ID NO: 58)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCAACTTTGAGCAAGATTTTTTTTTTTTTTTTTTTTTTTVN
Probe10 (SEQ ID NO: 59)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCAATTCGGAATTCCGGTTTTTTTTTTTTTTTTTTTTTTVN
Probe11 (SEQ ID NO: 60)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCCCAAGGTAATACATTTTTTTTTTTTTTTTTTTTTTTVN
Probe12 (SEQ ID NO: 61)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGCATTTCCTATTCGATTTTTTTTTTTTTTTTTTTTTTVN
Probe13 (SEQ ID NO: 62)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTTGCTAAATCTAACCGCCTTTTTTTTTTTTTTTTTTTTTTVN
Probe14 (SEQ ID NO: 63)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAATTAAATTCTGATGGTTTTTTTTTTTTTTTTTTTTTTVN
Probe15 (SEQ ID NO: 64)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCATTACATAGGTGCTAAGTTTTTTTTTTTTTTTTTTTTTTVN
Probe16 (SEQ ID NO: 65)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGACTTGCGCTCGCACTTTTTTTTTTTTTTTTTTTTTTVN
Probe17 (SEQ ID NO: 66)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTATAGTATCTCCCAAGTTCTTTTTTTTTTTTTTTTTTTTTVN
Probe18 (SEQ ID NO: 67)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGCGCCTGTAATCCGCATTTTTTTTTTTTTTTTTTTTTTVN
Probe19 (SEQ ID NO: 68)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGCCACTCTTTAGGTAGTTTTTTTTTTTTTTTTTTTTTTVN
Probe20 (SEQ ID NO: 69)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTATGCAAGTGATTGGTCTTTTTTTTTTTTTTTTTTTTTTVN
Probe21 (SEQ ID NO: 70)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCCAAGCCACGTTTATACGTTTTTTTTTTTTTTTTTTTTTTVN
Probe22 (SEQ ID NO: 71)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTACCTGATTGCTGTATAACTTTTTTTTTTTTTTTTTTTTTTVN
Probe23 (SEQ ID NO: 72)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCAGCGCATCTATCCTCTATTTTTTTTTTTTTTTTTTTTTTVN
Probe24 (SEQ ID NO: 73)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCACGCGTAGGACTAGTTTTTTTTTTTTTTTTTTTTTTVN
Probe25 (SEQ ID NO: 74)	UUUUUACACTCTTTCCCTACACGACGCTCTTCCGATCTCGACTAAGTATGTAGCGCTTTTTTTTTTTTTTTTTTTTTTVN

Frame probe
Layoutl (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 TIT TIT TTT TTT TTT TTT T

Example 9

Nimblegen 3' to 5' arrays with free 5' end Array probes

	5' to 3'
Probel (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
ProbelO (SEQ ID NO: 89)	GCGTTCAGAGTGGCAGTCGAGATCACCCGGAATTCCGAATTGGCAGATCGGAAGAGCGTAGTGTAG
Probel 1 (SEQ ID NO: 90)	GCGTTCAGAGTGGCAGTCGAGATCACATGTATTACCTTGGGCGAAGATCGGAAGAGCGTAGTGTAG
Probel2 (SEQ ID NO: 91)	GCGTTCAGAGTGGCAGTCGAGATCACCTCGAATAGGAAATGCGAAGATCGGAAGAGCGTAGTGTAG
Probel3 (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)	GTGATCTCGACTGCCACTCTGAATTTTTTTTTTTTTTTTTTTTTVN

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)	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTVN

Second PCR

A_primer (SEQ ID NO: 110)	ACACTCTTTCCCTACACGACGCTCTTCCGATCT
B_primer (SEQ ID NO: 111)	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Example 11

Template switching TemplateswitchJongB (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_Ch r2+2_RE_A (SEQ ID NO: 115) AGACGTGTGCTCTTCCGATCTCTCGCTAACAAGCAGAGAGAAC

B_A431_Chr3+7_FW_B (SEQ ID NO: 116) AGACGTGTGCTCTTCCGATCTTGAGAACAAGGGGGAAGAG

B_A431_Ch r3+7_RE_B (SEQ ID NO: 117) AGACGTGTGCTCTTCCGATCTCGGTGAAACAAGCAGGTAAC

B_NT_1_FW (SEQ ID NO: 118) AGACGTGTGCTCTTCCGATCTCATTCCCCACACTCATCACAC

B_NT_1_RE (SEQ ID NO: 119) AGACGTGTGCTCTTCCGATCTTCACACTGGAGAAAGACCC

B_NT_2_FW (SEQ ID NO: 120) AGACGTGTGCTCTTCCGATCTGGGGTTCAGAGTGATTTTTTCAG

B_NT_2_RE (SEQ ID NO: 121) AGACGTGTGCTCTTCCGATCTTCCGTTTTTCTTTCAGTGCC

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

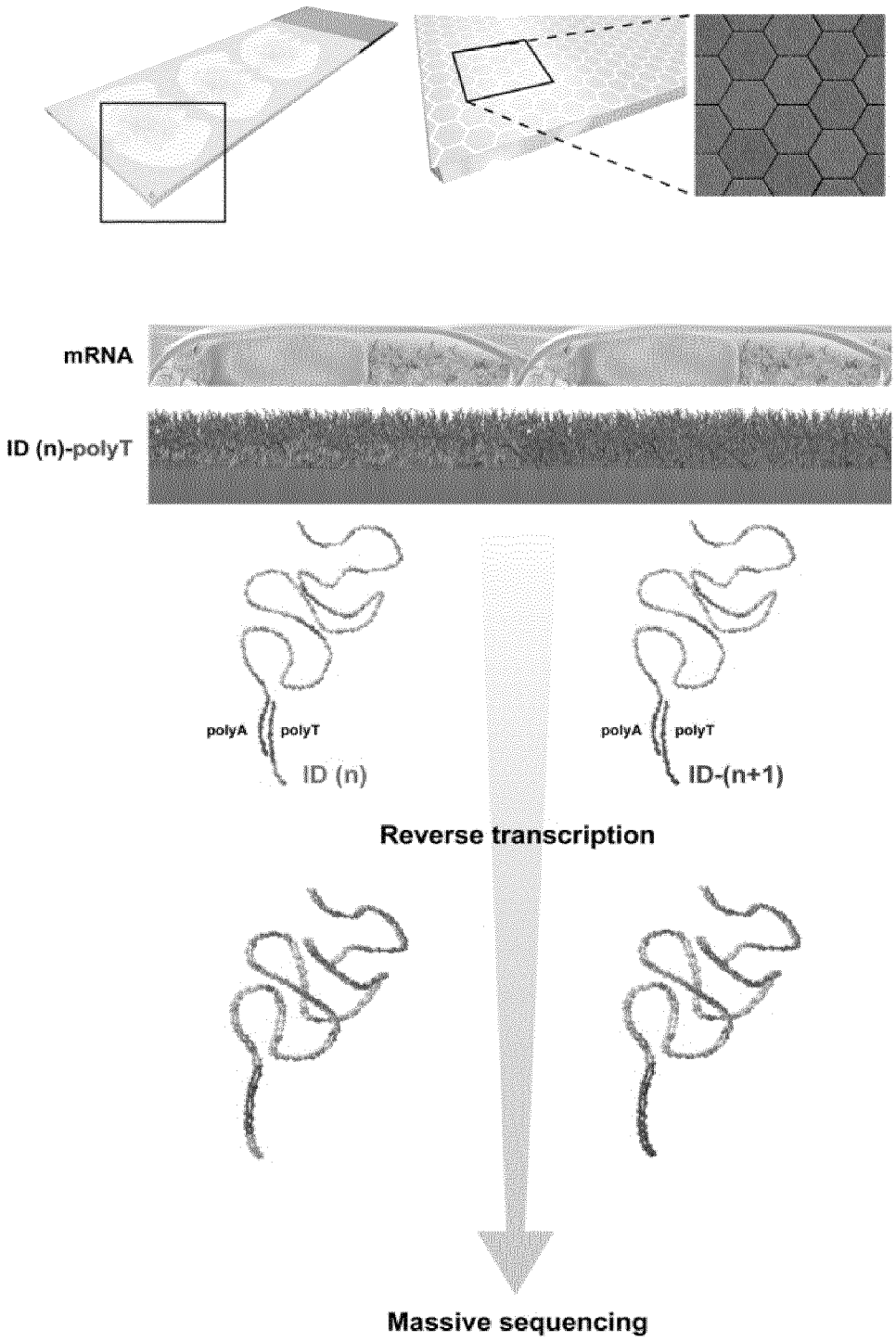

mRNA
ID (n)-polyT
polyA
polyT
ID (n)
polyA
polyT
ID-(n+1)
Reverse transcription
Massive sequencing

Figure 2

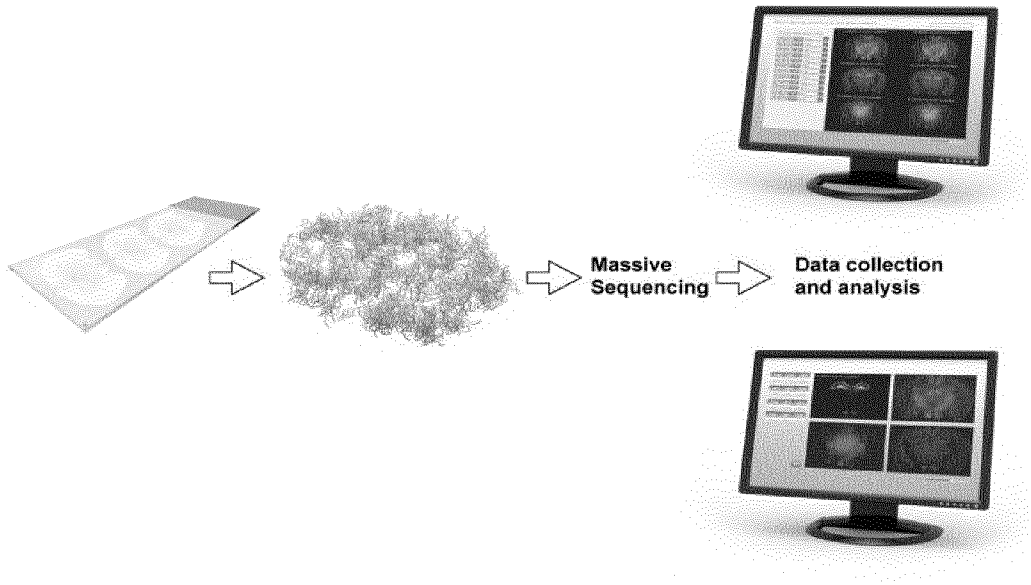


Figure 3

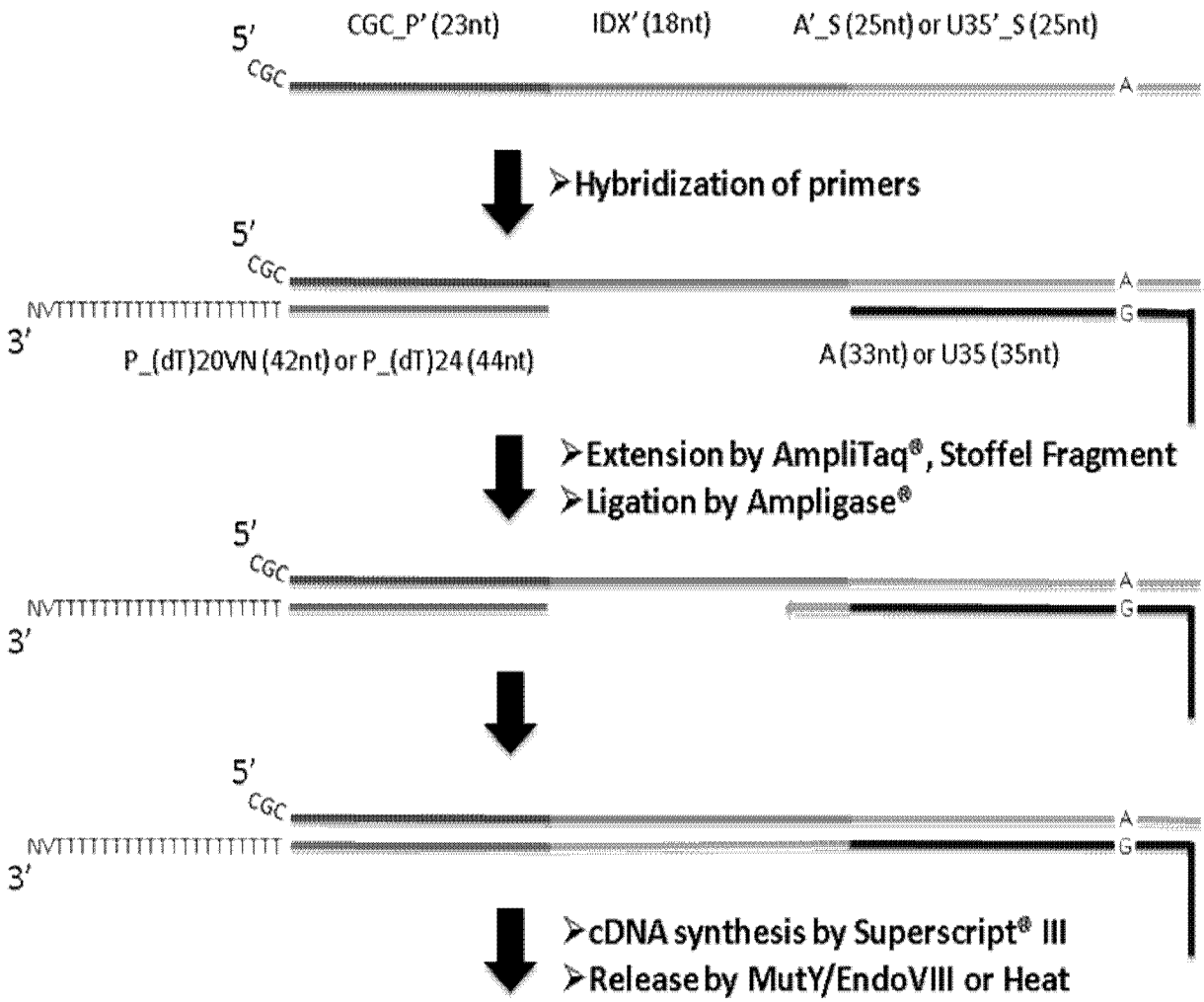

5′
CGC_P′ (23nt)
IDX′ (18nt)
A′_S (25nt) or U35′_S (25nt)
CGC
A
➢Hybridization of primers
5′
CGC
A
NVTTTTTTTTTTTTTTTTTTT
G
3′
P_(dT)20VN (42nt) or P_(dT)24 (44nt)
A (33nt) or U35 (35nt)
➢Extension by AmpliTaq®, Stoffel Fragment
➢Ligation by Ampligase®
5′
CGC
A
NVTTTTTTTTTTTTTTTTTTT
G
3′
5′
CGC
A
NVTTTTTTTTTTTTTTTTTTT
G
3′
➢cDNA synthesis by Superscript® III
➢Release by MutY/EndoVIII or Heat

Figure 4

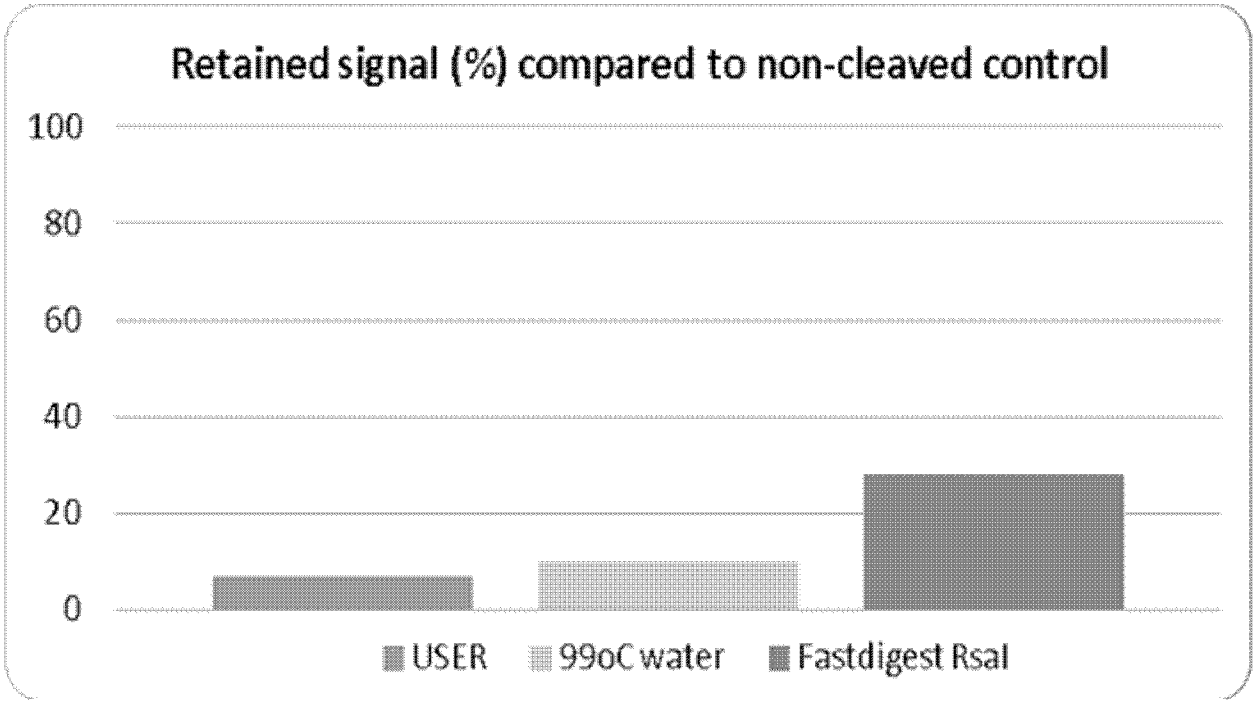

Retained signal (%) compared to non-cleaved control
100
80
60
40
20
0
USER
99oC water
Fastdigest Rsal

Figure 5

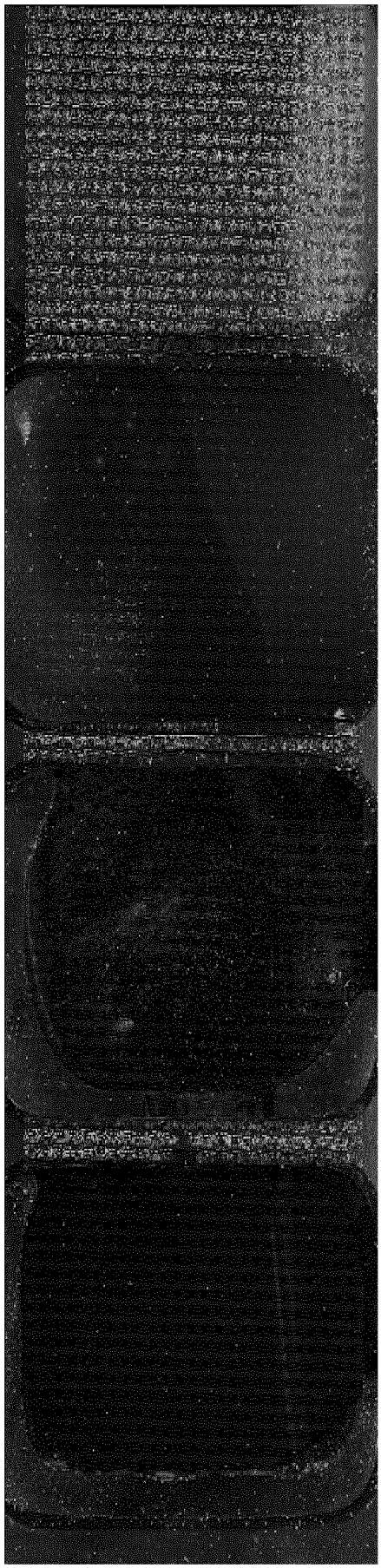


Figure 6

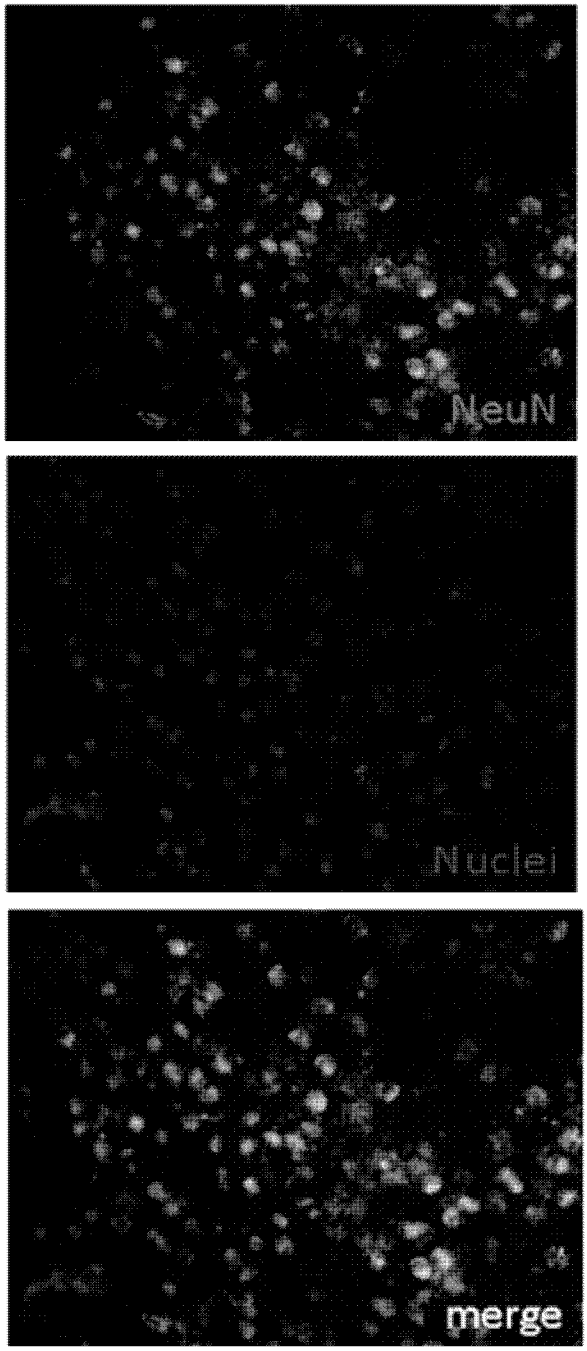


Figure 7

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

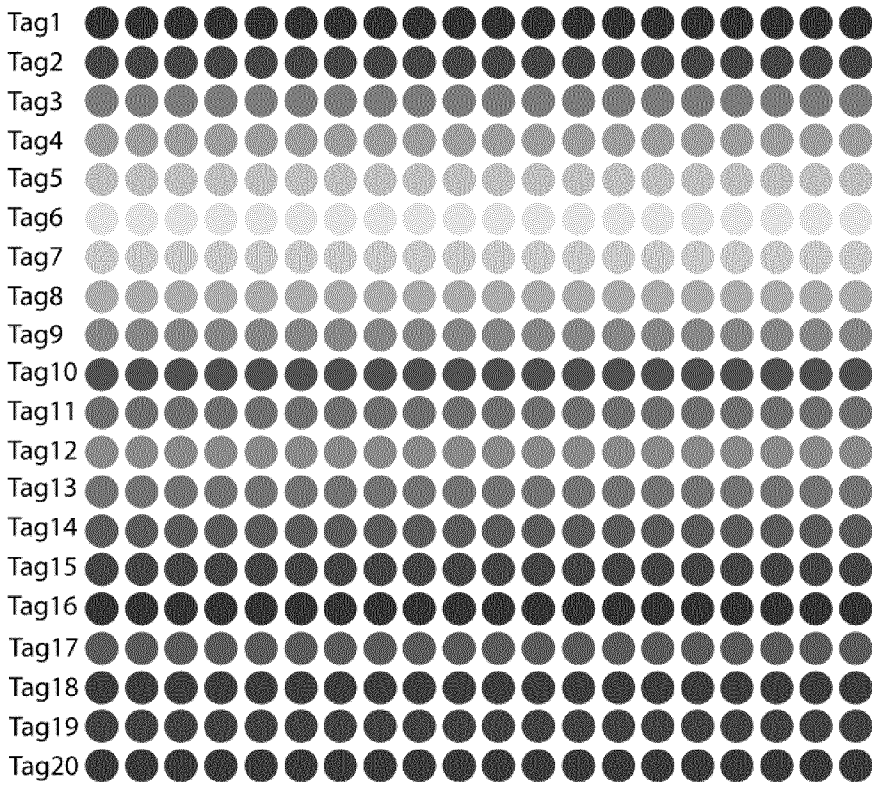


Figure 8

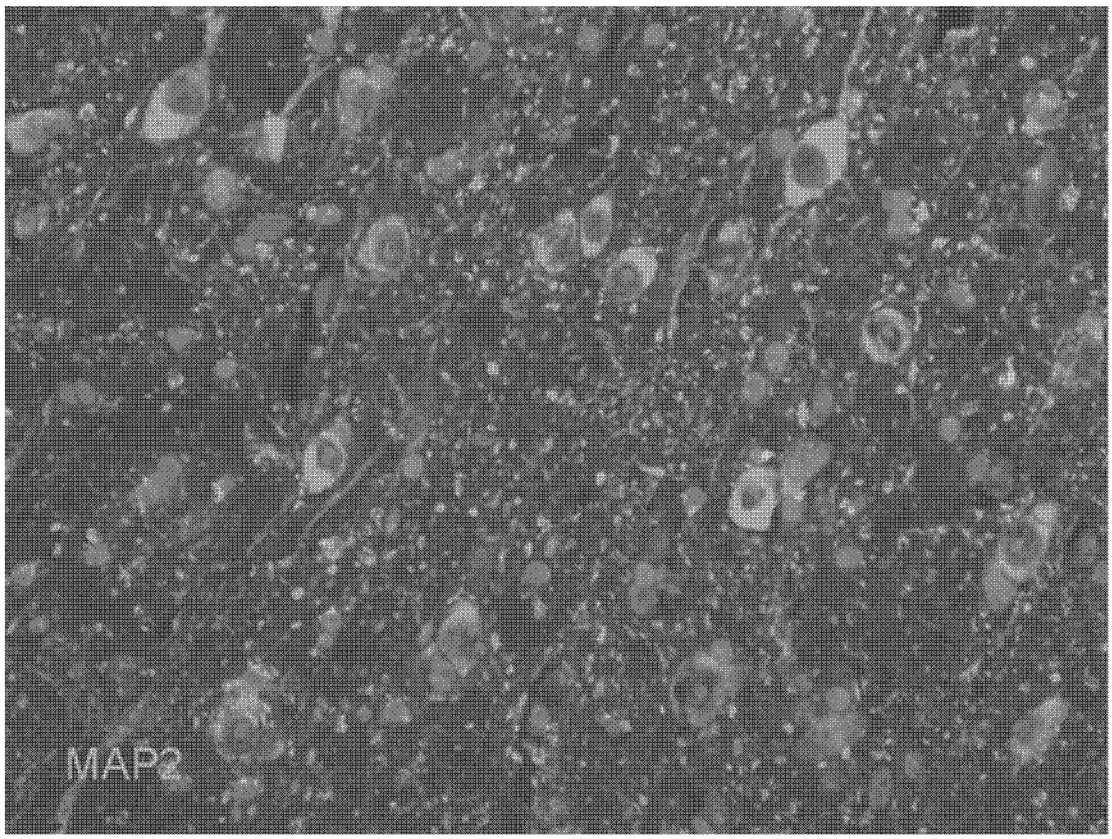


Figure 9

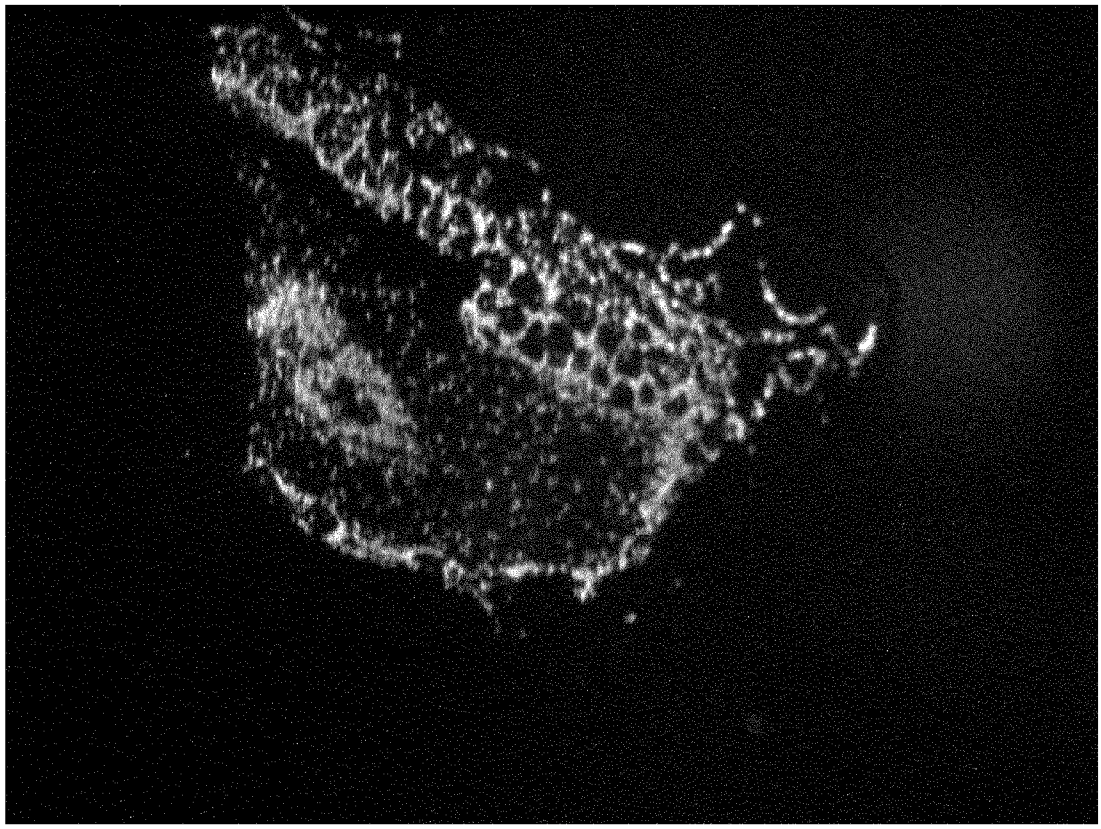


Figure 10

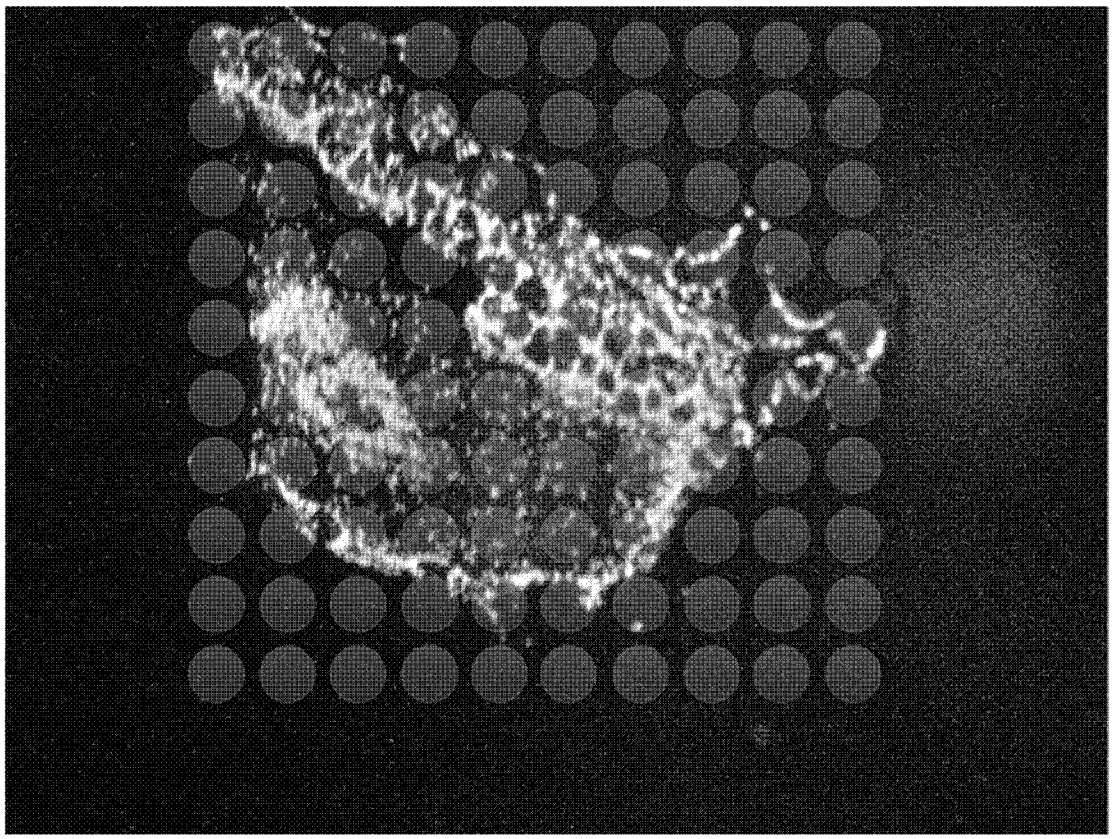


Figure 11

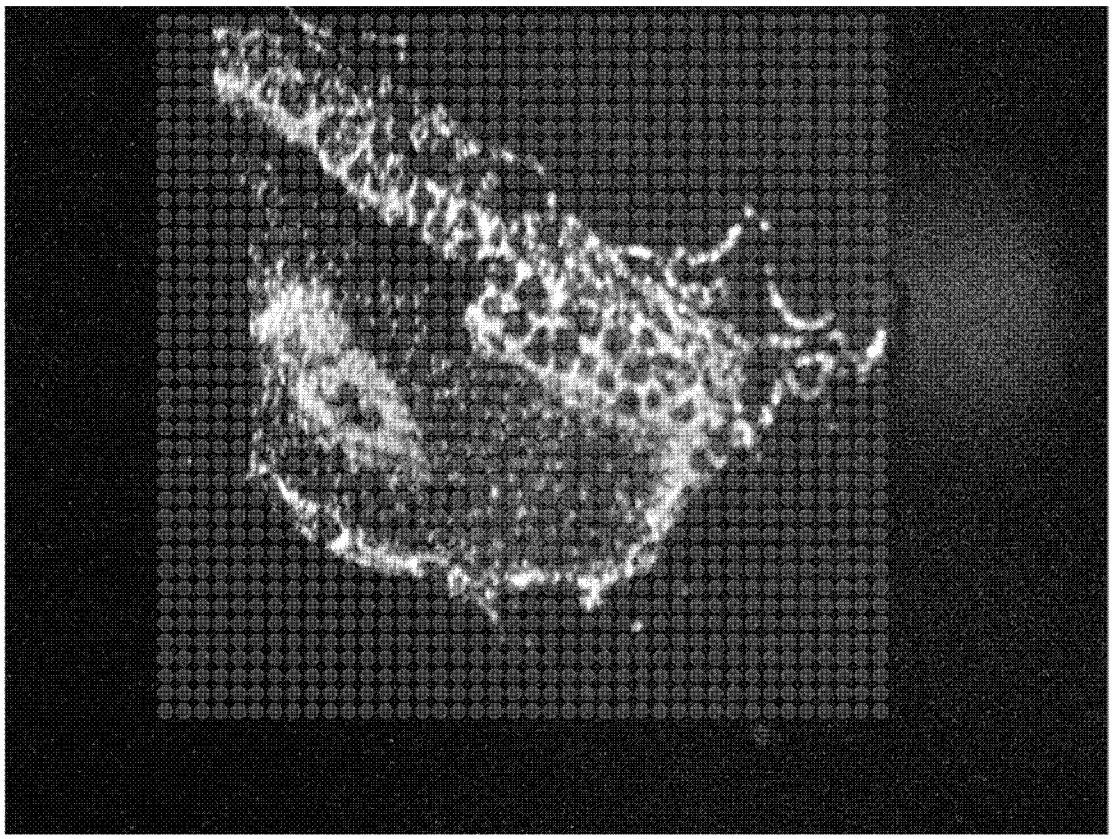


Figure 12

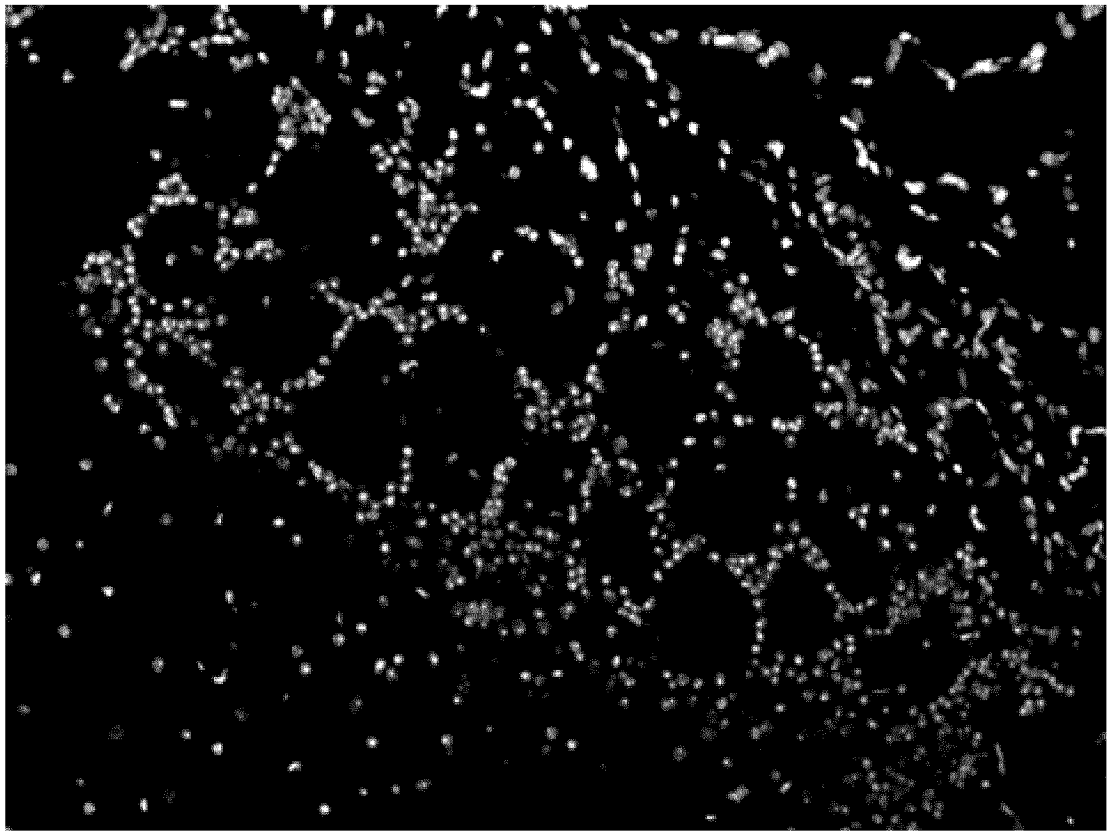
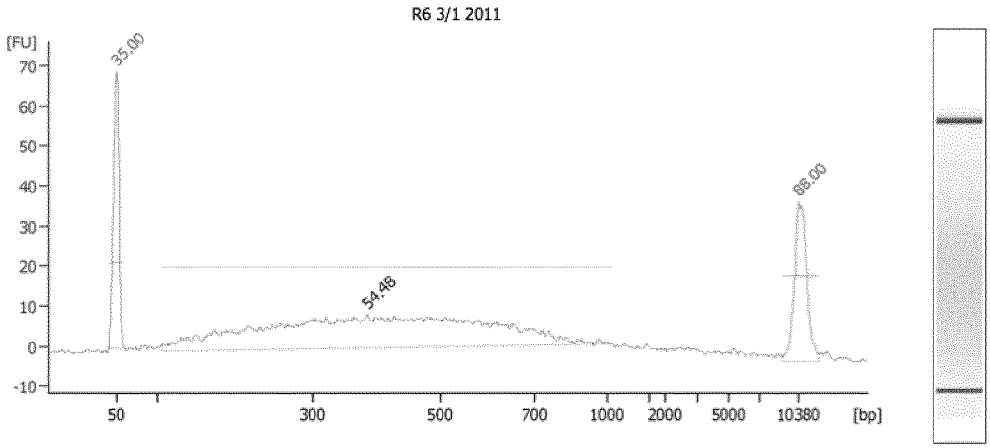


Figure 13



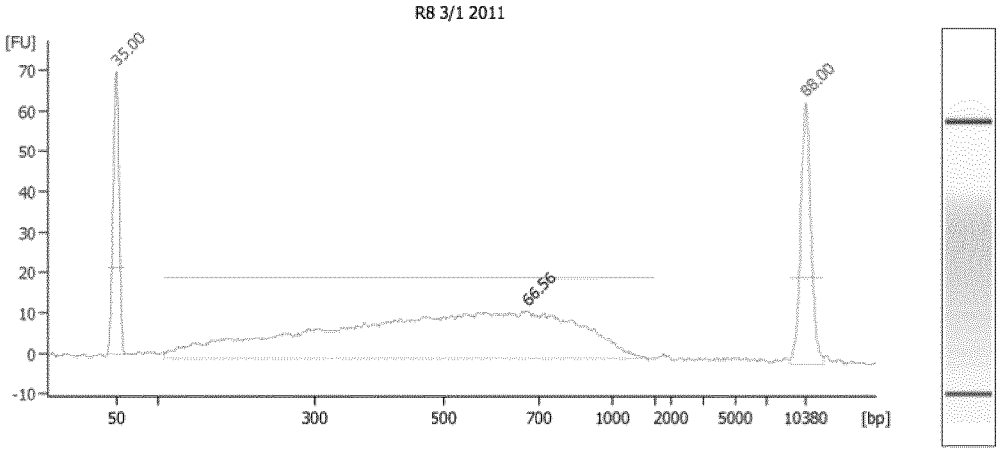
Overall Results for sample 3 : R6 3/1 2011

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

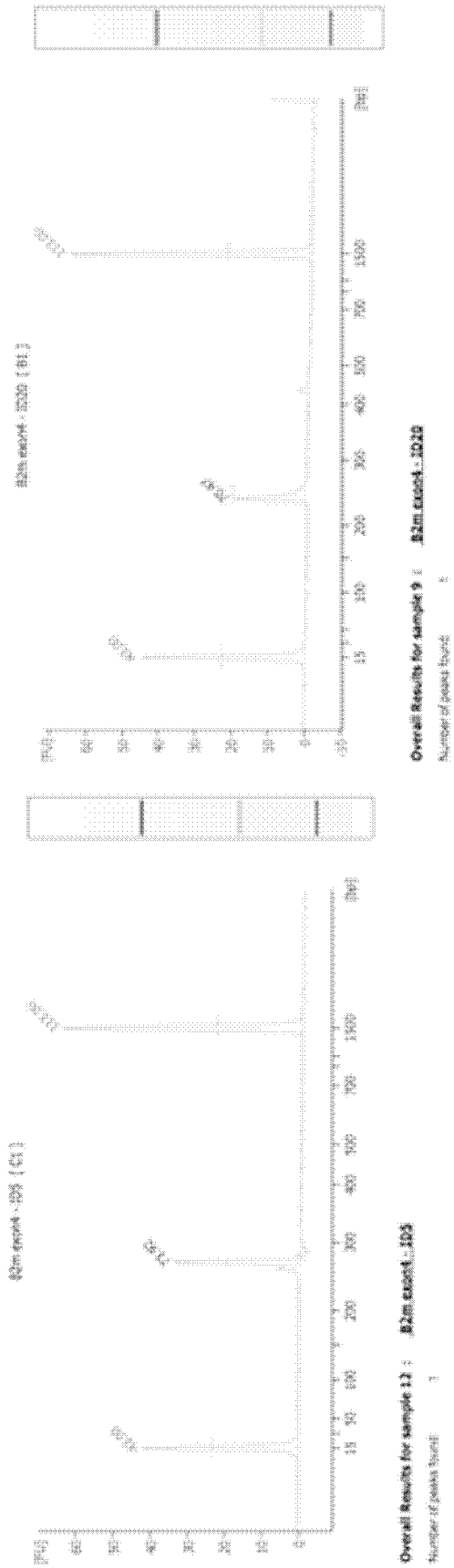

Overall Results for sample 9 :
β2m count : 2020
Number of peaks found : 4
Overall Results for sample 12 :
β2m count : 103
Number of peaks found : 1

Figure 16

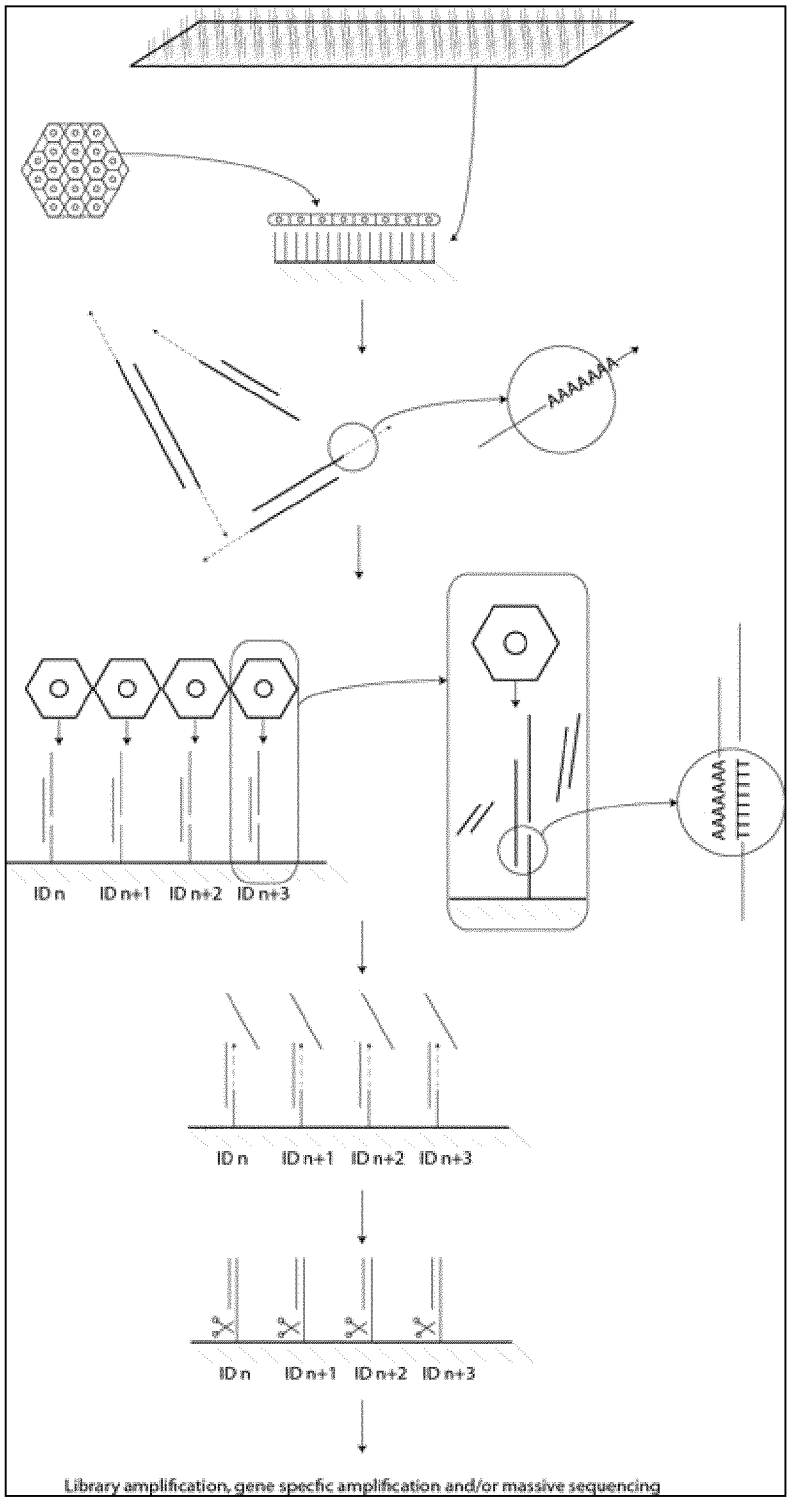
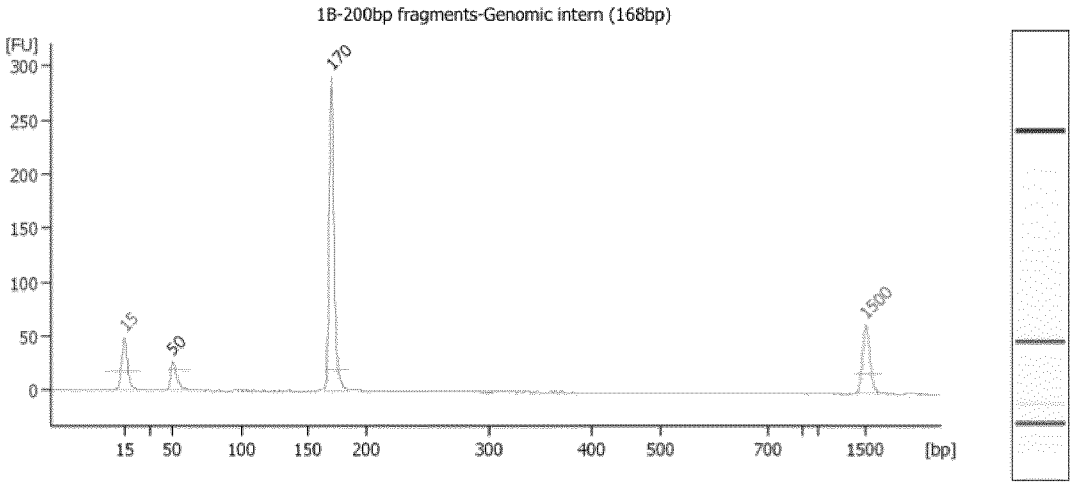

AAAAAAA
ID n
ID n+1
ID n+2
ID n+3
AAAAAAA
TTTTTTTT
ID n
ID n+1
ID n+2
ID n+3
ID n
ID n+1
ID n+2
ID n+3
Library amplification, gene specific amplification and/or massive sequencing

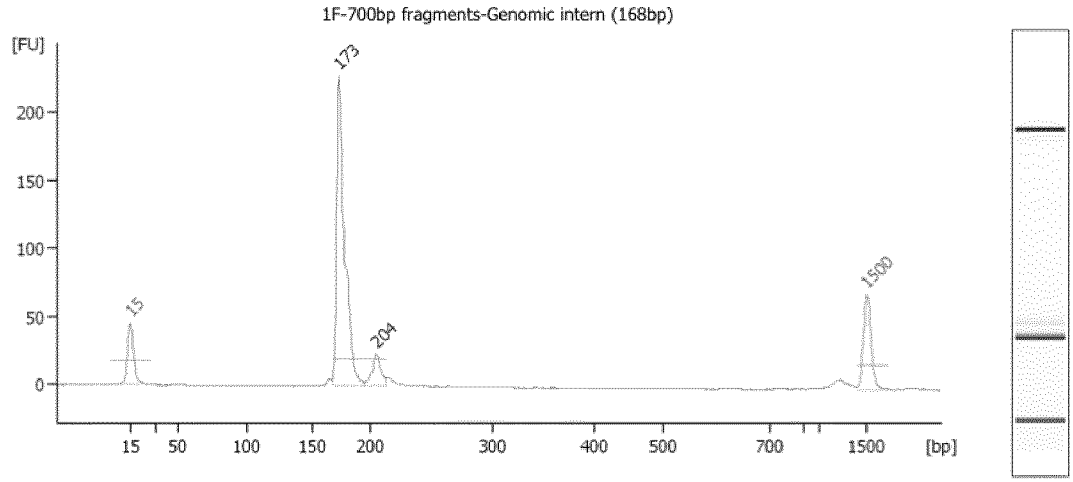
Figure 17



Overall Results for sample 3 : **1B-200bp fragments-Genomic intern (168bp)**

Number of peaks found: 2

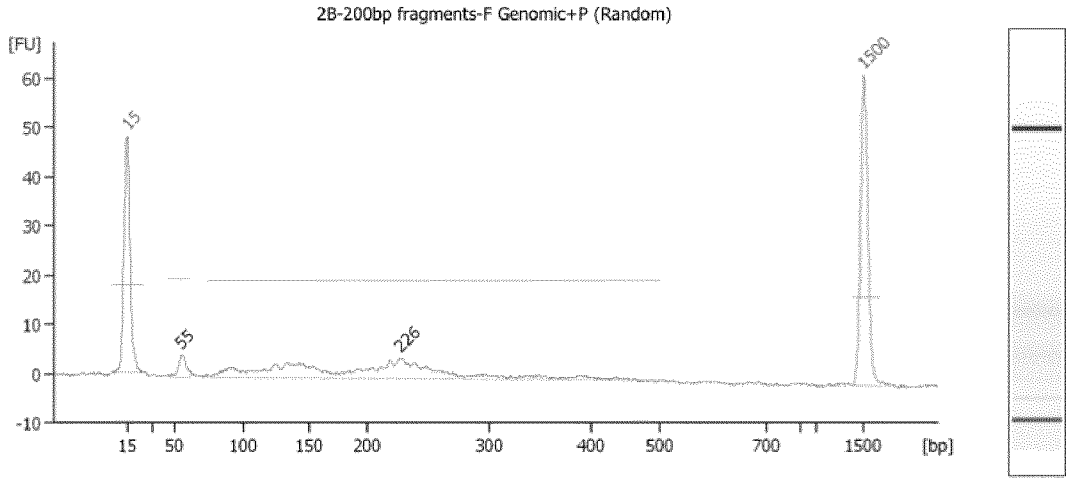
Figure 18



Overall Results for sample 7 : **1F-700bp fragments-Genomic intern (168bp)**

Number of peaks found: 2

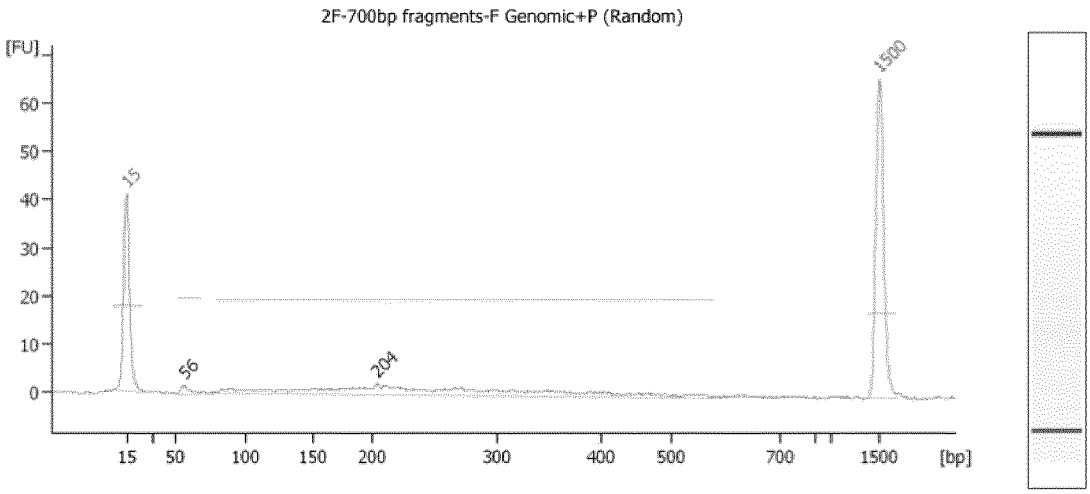
Figure 19



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

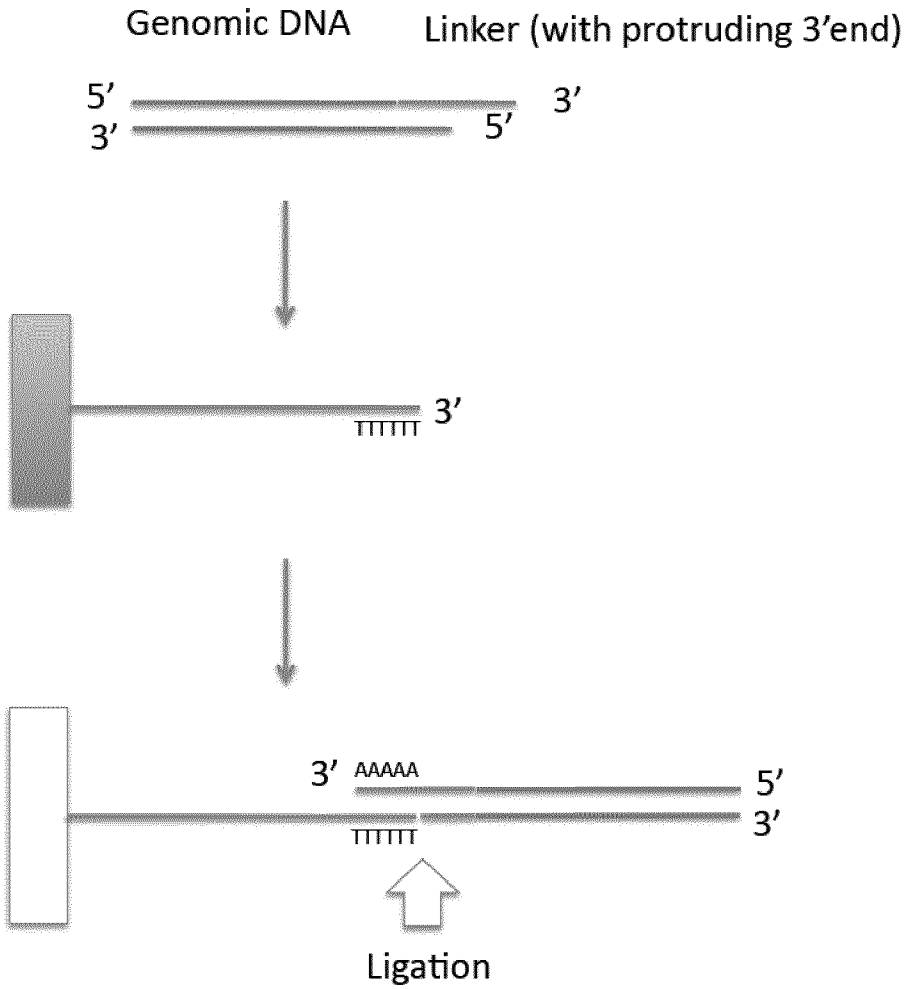

Genomic DNA
Linker (with protruding 3'end)
5'
3'
3'
5'
3'
3'
AAAAA
5'
3'
Ligation

Figure 22

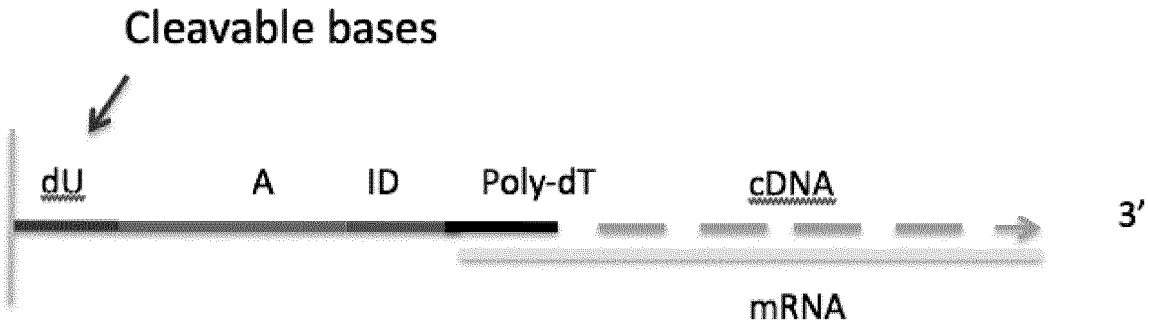


Figure 23

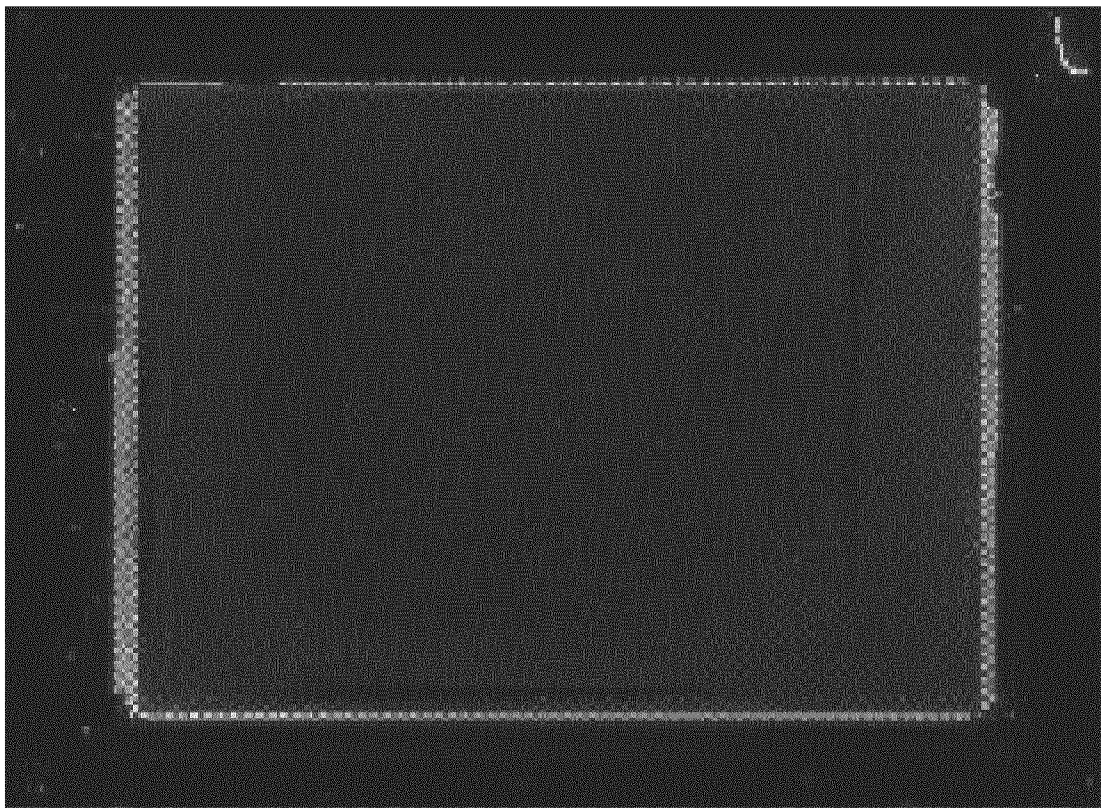


Figure 24

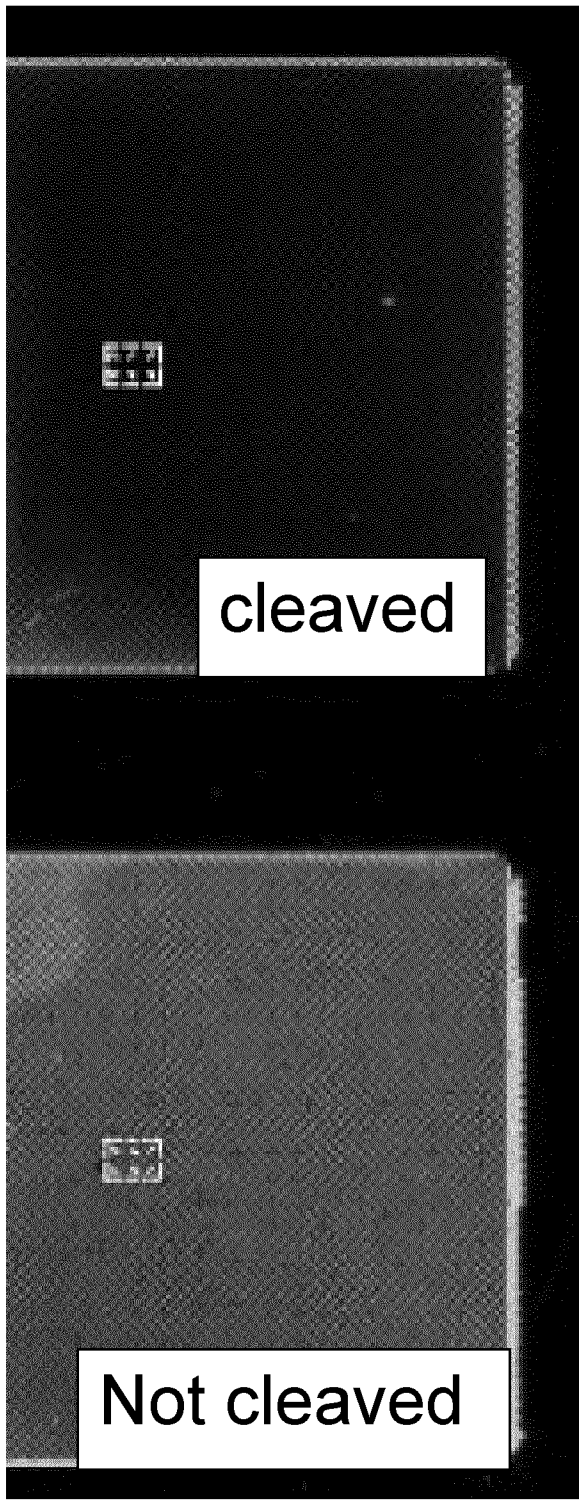


Figure 25

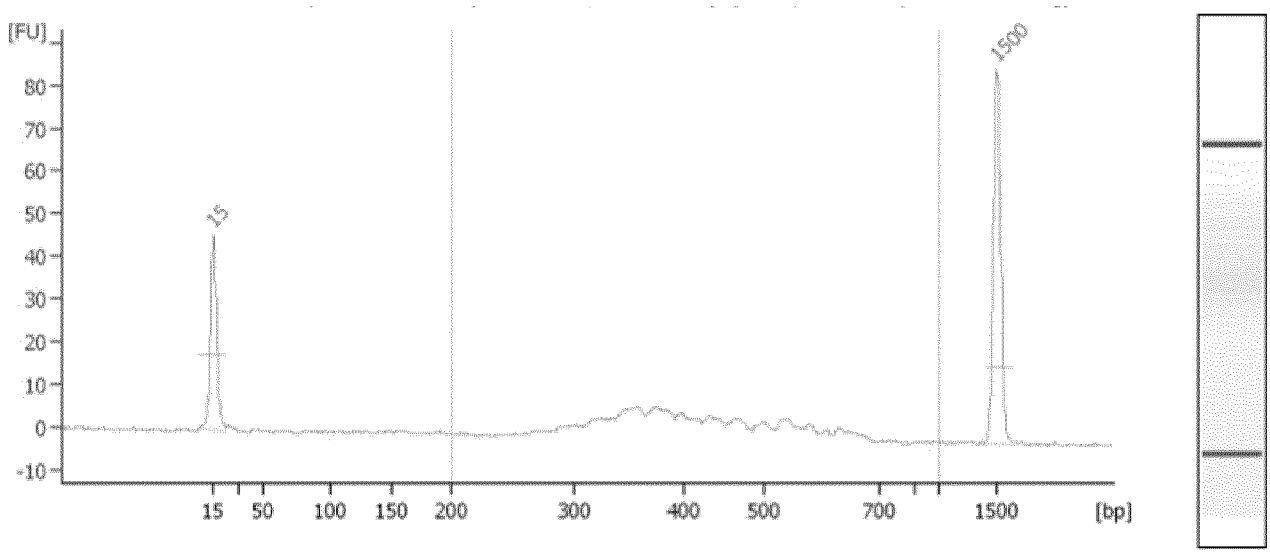

[FU]
80
70
60
50
40
30
20
10
0
-10
15
1500
15
50
100
150
200
300
400
500
700
1500
[bp]

Figure 26

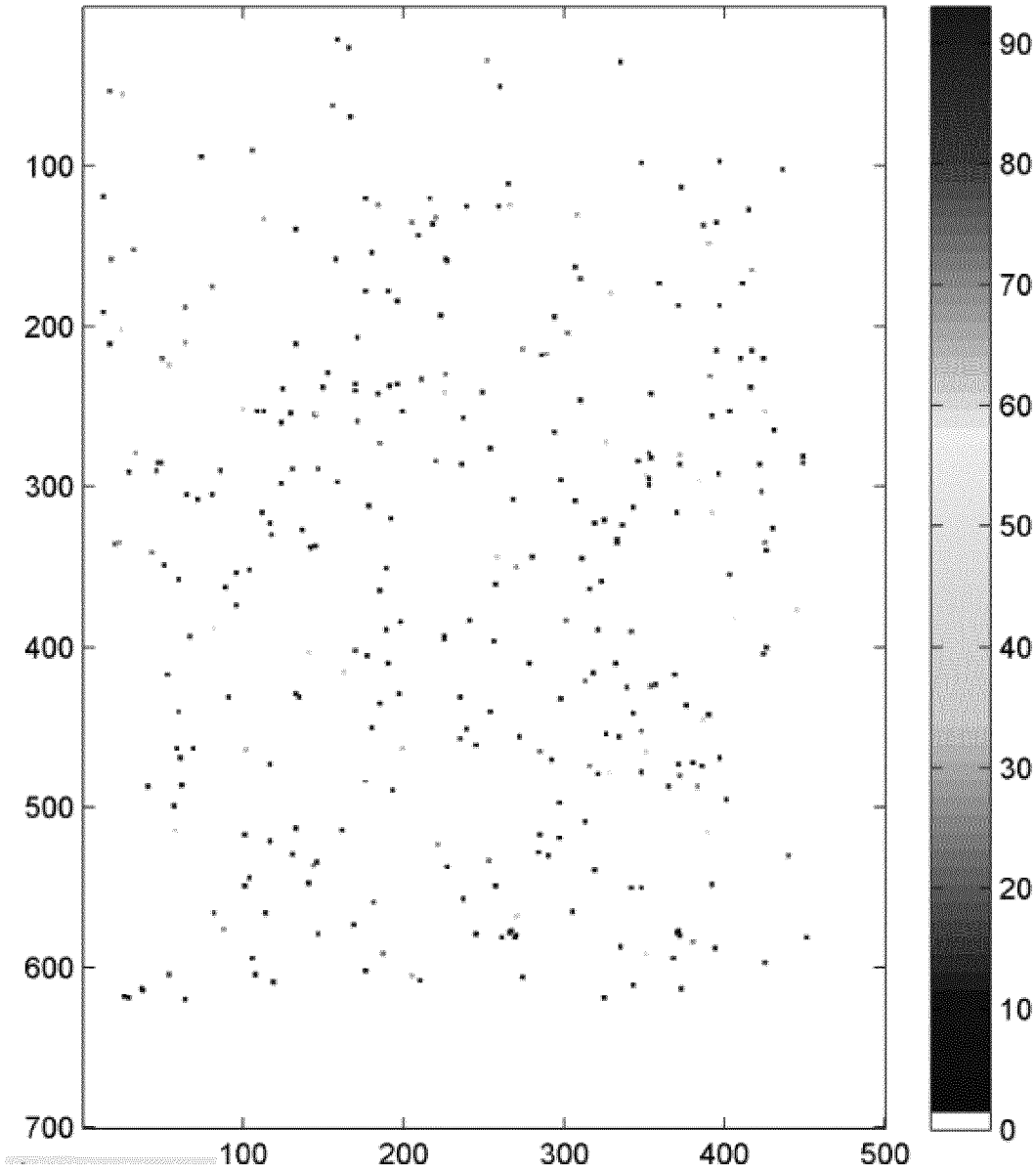

100
200
300
400
500
600
700
100
200
300
400
500
90
80
70
60
50
40
30
20
10
0

Figure 27

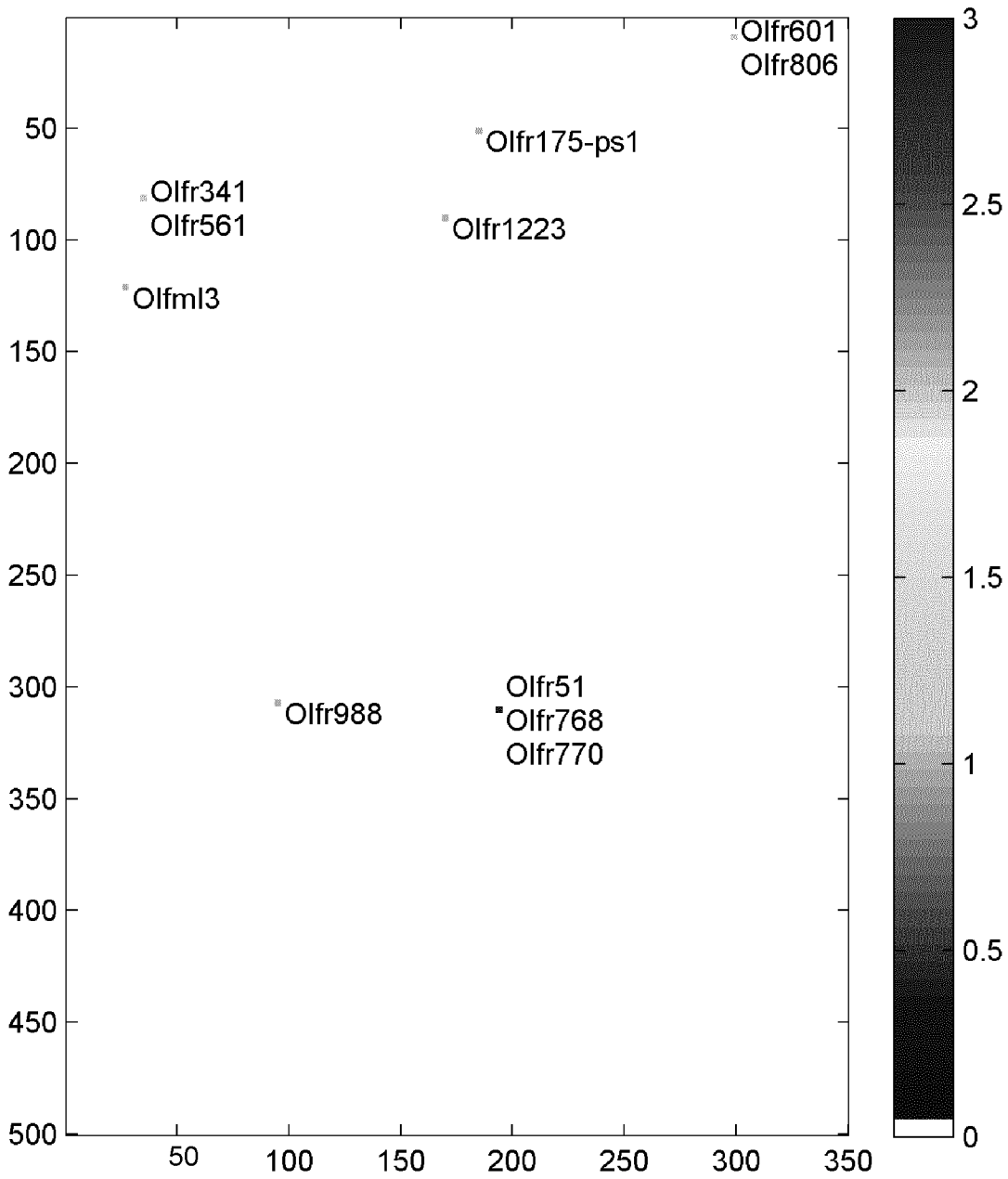

Olfr601
Olfr806
Olfr175-ps1
Olfr341
Olfr561
Olfr1223
Olfml3
Olfr988
Olfr51
Olfr768
Olfr770
50
100
150
200
250
300
350
400
450
500
50
100
150
200
250
300
350
3
2.5
2
1.5
1
0.5
0

Figure 28

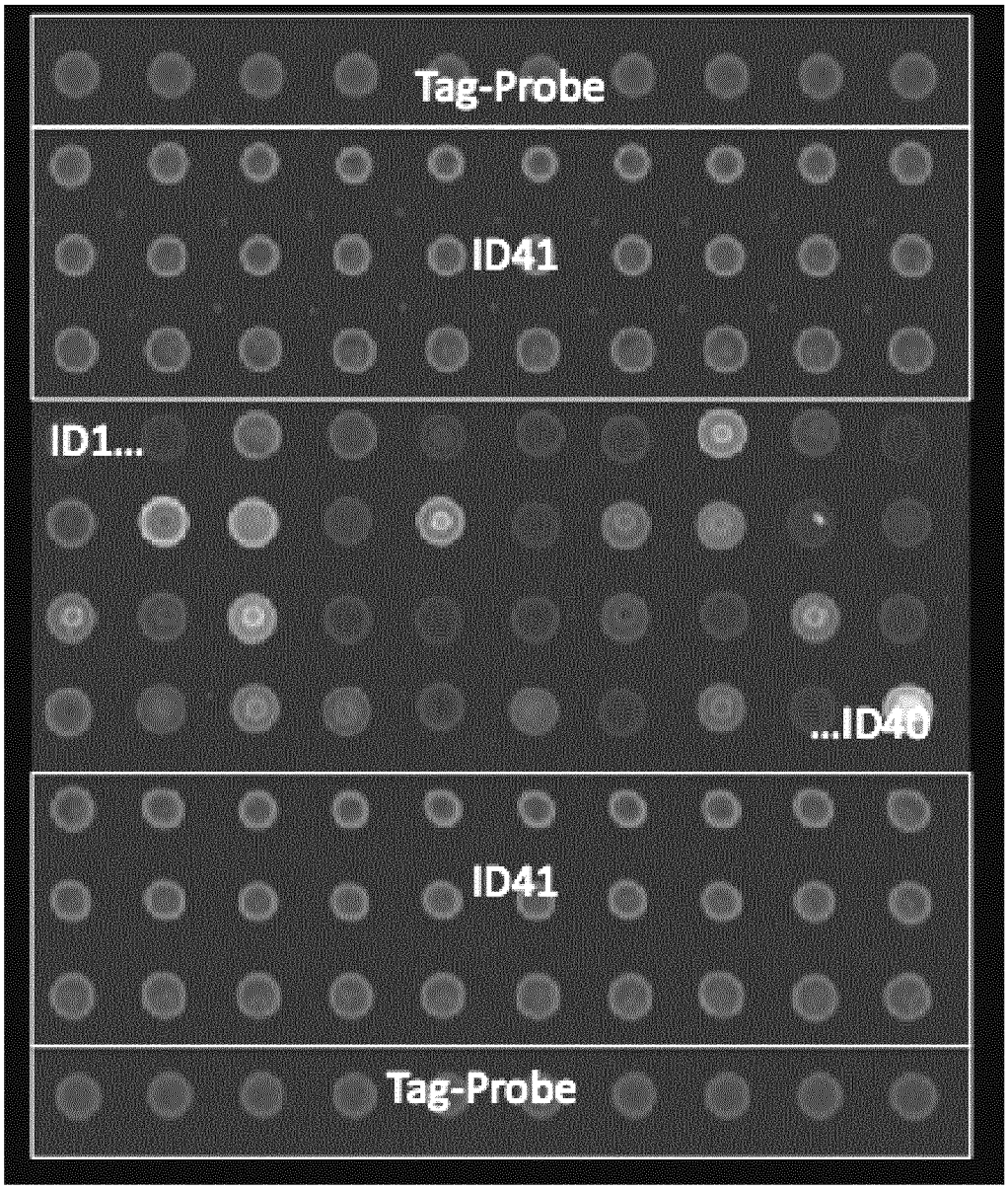


Figure 29

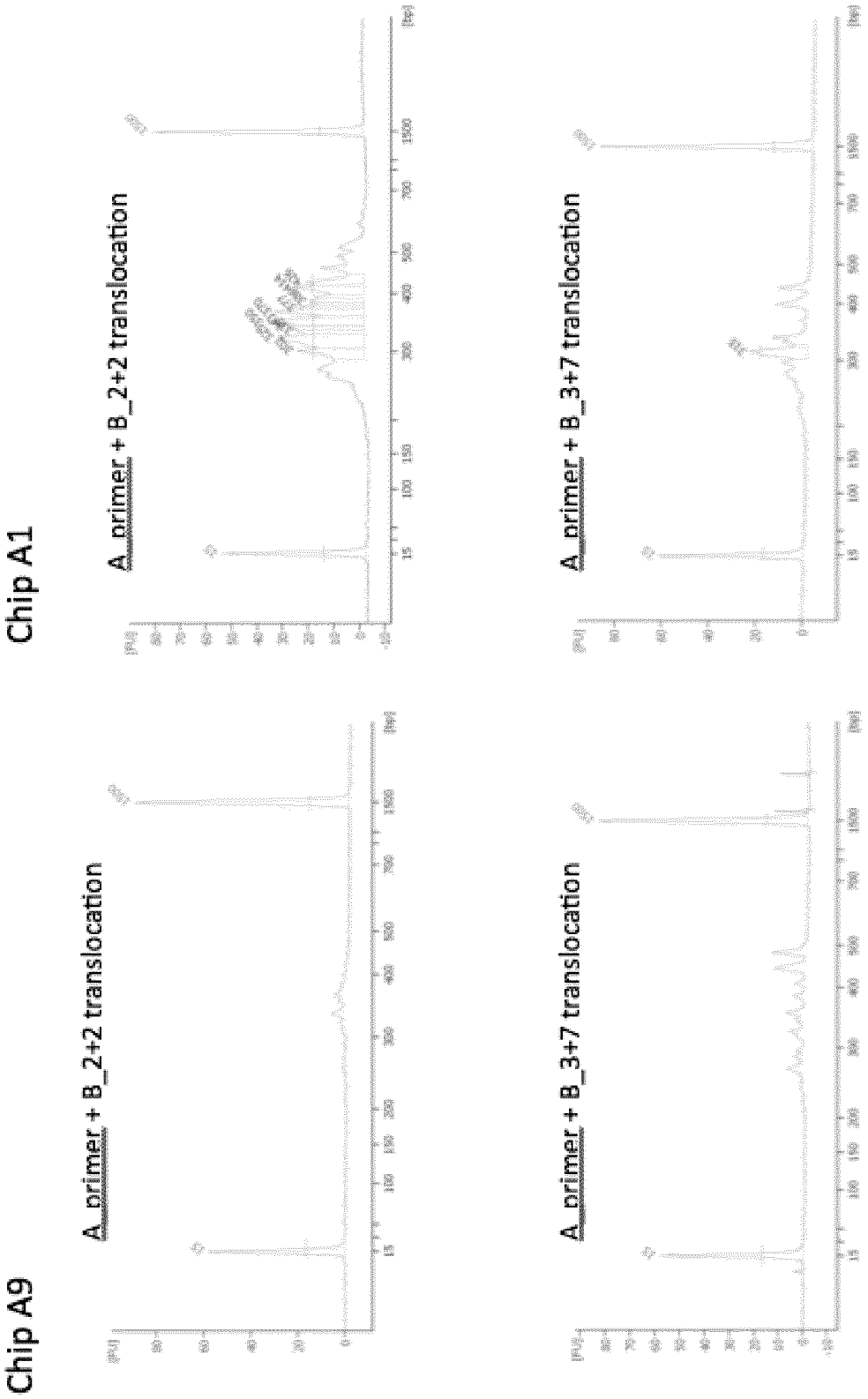

Chip A1
A_primer + B_2+2 translocation
A_primer + B_3+7 translocation
Chip A9
A_primer + B_2+2 translocation
A_primer + B_3+7 translocation

Figure 30

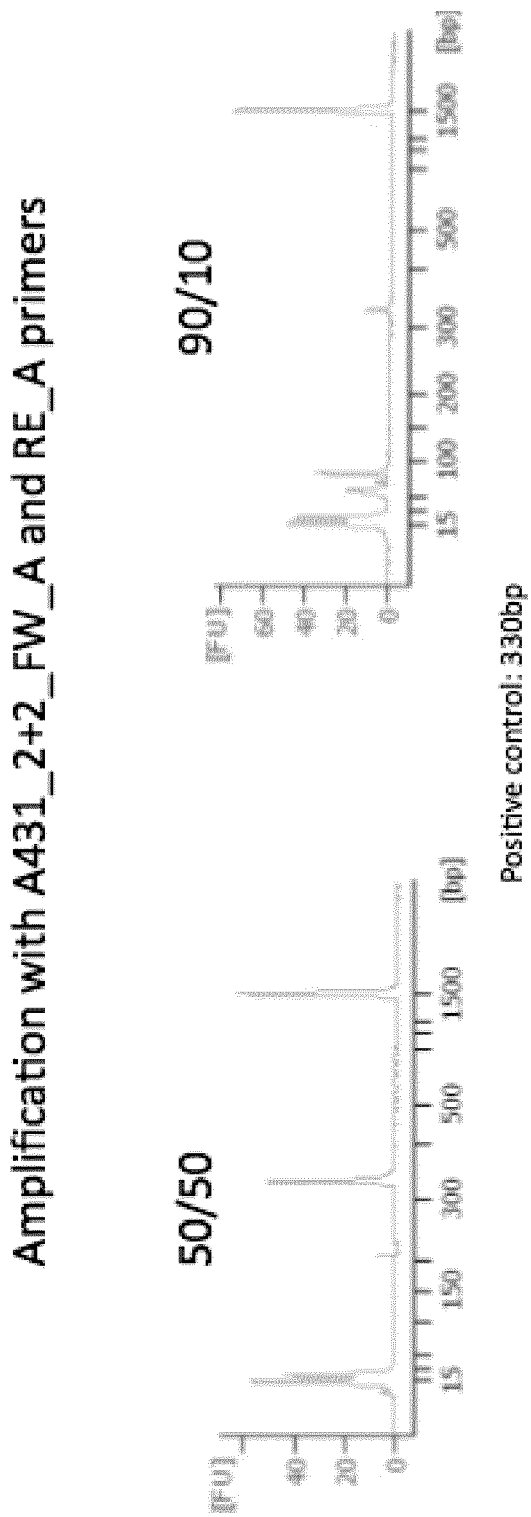
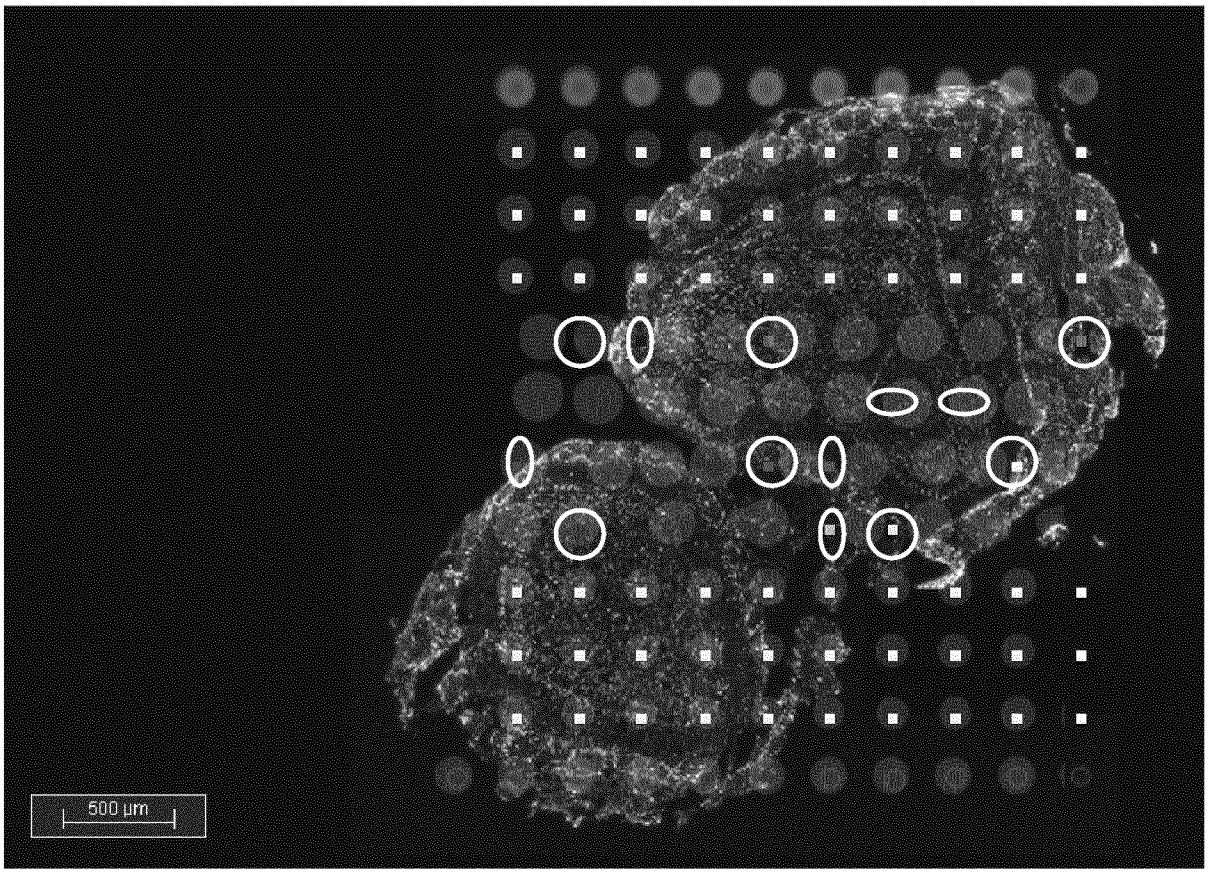

Amplification with A431_2+2_FW_A and RE_A primers
90/10
50/50
[FU]
[bp]
15
100
200
300
500
1500
150
Positive control: 330bp

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 , Retri eved from the Internet: URL: http://www.af fymetri x.com [retri eved on 2002-10-02] abstract -----	37-41
X	"GeneChi p Human Genome U133 Set" , INTERNET CITATION, 26 February 2003 (2003-02-26) , XP002232760, Retri eved from the Internet: URL: http://www.af fymetri x.com/support/tech ni cal /datasheet s/hgul33_datasheet.pdf [retri eved on 2003-02-26] abstract -----	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" documentwhich may throw doubts on priority claim(s) orwhich 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 , Wol fgang

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 genechi p high-densi ty oligonucl eoti de arrays for gene expressi on moni tori ng" , LI FE SCI ENCE NEWS, AMERSHAM LI FE SCI ENCE, US, 1 January 1998 (1998-01-01) , pages 11-14, XP002964122 , ISSN : 0969-0190 abstract	37-41
A	US 2006/211001 Al (YU JINGWEI [US] ET AL) 21 September 2006 (2006-09-21) abstract; exampl es 1-3 cl aims 1,5-6, 13-14	1-43
A	US 2003/040035 Al (SLAMON DENNIS J [US] ET AL) 27 February 2003 (2003-02-27) abstract; f i gures 1-4 cl aims 1-19	1-43
A	W0 2007/030373 A2 (ST JUDE CHI LDRENS RES HOSPITAL [US] ; CURRAN THOMAS [US] ; RICE DENNIS S) 15 March 2007 (2007-03-15) abstract; cl aims 1-7	1-43
A	EGUI LUZ C ET AL: "Mul t i t i ssue array revi ew: A chronol ogi cal descri pti on of t i ssue array techni ques , appl i cati ons 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, D0I : 10. 1016/J . PRP.2006.04.003 [retri eved on 2006-08-10] the whol e document	1-43
A	SI EVERTZ0N MARIA ET AL: "Transcri ptome analysi s i n primary neural stem cel l s usi ng a tag cDNA ampl i f i cati on method" , BMC NEUROSCI ENCE, BIOMED CENTRAL, LONDON , GB, vol . 6 , no. 1 , 15 Apri l 2005 (2005-04-15) , page 28, XP021002918, ISSN : 1471-2202 , D0I : 10. 1186/1471-2202-6-28 the whol e 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, D0I: 10. 1126/SCIENCE.270.5235.467 the whole document	1-43
A	WILLI-M0NNERAT 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	